Foundations for Success

National Mathematics Advisory Panel

Reports of the Task Groups and Subcommittees

2008

U.S. Department of Education

Chapter 1: Introduction

The National Mathematics Advisory Panel

The President established the Panel via Executive Order 13398 (Appendix A), in which he also assigned responsibility to the U.S. Secretary of Education for appointment of members and for oversight of the Panel. While the presidential charge contains many explicit elements, there is a clear emphasis on the preparation of students for entry into, and success in, algebra.

Over a period of 20 months, the Panel received public testimony as a committee of the whole but worked largely in task groups and subcommittees dedicated to major components of the presidential charge. Questions like the following illustrate the scope of the Panel's inquiry:

- What is the essential content of school algebra and what do children need to know before starting to study it?
- What is known from research about how children learn mathematics?
- What is known about the effectiveness of instructional practices and materials?
- How can we best recruit, prepare, and retain effective teachers of mathematics?
- How can we make assessments of mathematical knowledge more accurate and more useful?
- What do practicing teachers of algebra say about the preparation of students whom they receive into their classrooms and about other relevant matters?
- What are the appropriate standards of evidence for the Panel to use in drawing conclusions from the research base?

Each of five task groups carried out a detailed analysis of the available evidence in a major area of the Panel's responsibility: Conceptual Knowledge and Skills, Learning Processes, Instructional Practices, Teachers and Teacher Education, and Assessment. Each of three subcommittees was charged with completion of a particular advisory function for the Panel: Standards of Evidence, Instructional Materials, and the Panel-commissioned National Survey of Algebra Teachers. Each task group and subcommittee produced a report, all of which are compiled here in this document.

The Panel took consistent note of the President's emphasis on "the best available scientific evidence" and set a high bar for admitting research results into consideration. In essence, the Panel required the work to have been carried out in a way that manifested rigor and could support generalization at the level of significance to policy. One of the subcommittee reports covers global considerations relating to standards of evidence, while individual task group reports amplify the standards in the particular context of each task group's work. In all, the Panel reviewed more than 16,000 research publications and policy reports and received public testimony from 110 individuals, of whom 69 appeared before the Panel on their own and 41 others were invited on the basis of expertise to cover particular topics. In addition, the Panel reviewed written commentary from 160 organizations and individuals, and analyzed survey results from 743 active teachers of algebra.

In late 2007, the Panel synthesized the Final Report by drawing together the most important findings and recommendations, which were issued with the Panel's full voice. The task group and subcommittee reports in this volume carry the detailed analyses of research literature and other relevant materials from which the Panel synthesized its major findings. These supporting reports cover work carried out as part of the Panel's overall mission, but they are presented by only those members who participated in creating them. The Final Report represents findings and recommendations of the Panel as a whole.

Chapter 2: Report of the Subcommittee on Standards of Evidence

Valerie F. Reyna, Chair
Camilla Persson Benbow
A. Wade Boykin
Grover J. "Russ" Whitehurst, Ex Officio
Tyrrell Flawn, U.S. Department of Education Staff

CONTENTS

I. Introduction

The President's Executive Order calls for the National Mathematics Advisory Panel (Panel) to marshal the best available scientific evidence and offer advice on the effective use of the results of research related to proven, effective, and evidence-based mathematics instruction. The Panel's assertions and recommendations, therefore, need to be grounded in the highest quality evidence available from scientific studies. The highest-quality evidence that is actually available on some key topics, however, may not be of sufficiently high quality to support confident conclusions. So that the Panel may be systematic in identifying the quality of evidence on which its assertions and recommendations are based, criteria such as the following will be applied during the preparation and review of the final report.

II. Background: Categories of Internal and External Validity

There are three broad categories into which one can categorize research and the corresponding claims based on that research. First, there is the highest-quality scientific evidence, based on such considerations as the quality of the design, the validity and reliability of measures, the size and diversity of subject samples, and similar considerations of *internal* (scientific rigor and soundness) and *external* validity (generalizability to different circumstances and students). Hypothesis testing, especially the active search for disconfirmation, is a hallmark of high-quality research (e.g., Lewin, 1951; Platt, 1964). Hence, the Panel's strongest confidence will be reserved for studies that test hypotheses, meet the highest methodological standards (internal validity), and have been replicated with diverse samples of students under conditions that warrant generalization (external validity).

In addition to reviewing the best scientific evidence, the Panel is also charged with considering promising or suggestive findings that should be the subject of future research. Promising or suggestive studies do not meet the highest standards of scientific evidence, but they represent sound, scientific research that needs to be further investigated or extended. For example, laboratory studies showing significant effects of "desirable difficulties" (i.e., difficulties produced by challenging to-be-learned material) or of repeated testing on long-term retention could be extended to actual classrooms or existing curricula (e.g., Bjork, 1994; Roediger & Karpicke, 2006; see Cook & Campbell, 1979). The final category corresponds to statements based on values, impressions, or weak evidence; these are essentially opinions as opposed to scientifically justified conclusions. Issues such as what constitutes algebra are matters of expert opinion rather than of scientific evidence.

III. Quantity, Quality, and Balance of Evidence

A. Strong Evidence

All of the applicable high-quality studies support a conclusion (statistically significant individual effects, significant positive mean effect size, or equivalent consistent positive findings), and they include at least three independent studies with different relevant samples and settings, *or* one large high-quality multisite study. Any applicable studies of less than high quality show either a preponderance of evidence consistent with the high-quality studies (e.g., mean positive effect size) or such methodological weaknesses that they do not provide credible contrary evidence. Factors, such as error variance and measurement sensitivity, clearly influence the number of studies needed to support a conclusion (reflected in such statistics as p-rep, the probability of replicating an effect; Killeen, 2005); the number and balance of studies that are indicated above are, therefore, rules of thumb (e.g., see evidence standards applied by the What Works Clearinghouse at http://ies.ed.gov/).

B. Moderately Strong Evidence

Criteria for moderately strong evidence are the same as that for strong evidence, but with one of the following exceptions: there are only one or two high-quality studies, the effects have not been independently replicated by different researchers, or they do not involve different samples (i.e., diversity of characteristics) and settings.

C. Suggestive Evidence

Suggestive evidence is based on one of the following criteria:

a) There are some high-quality studies that support the conclusion (statistically significant effects, significant mean effects) but others that do not (nonsignificant). Those that do not are null, not negative (nonsignificant effect or mean effects, but not a significant negative effect). Any applicable moderate quality studies show a comparable pattern or better.

b) There are no high-quality studies, but all the applicable moderate-quality studies support the conclusion (statistically significant individual effects, significant positive mean effect size, or equivalent consistent positive findings), and there are at least three such studies.

D. Inconsistent Evidence

The evaluation of mixed evidence depends crucially on the quality of the designs and methods of each study. The results of high-quality designs trump inconsistent or null results of low-quality designs. Mixed results of high-quality studies, moderate-quality studies, or both, that are not consistent enough to fall into any of the previously described categories, and cannot be adjudicated by methodological criteria, are inconclusive.

E. Weak Evidence

Evidence is considered weak when only low-quality studies are available.

IV. Applying the Criteria

To apply such criteria, each study on which an assertion or recommendation is based must be characterized as “high quality,” “moderate quality,” or “low quality.” The standards for those designations will necessarily differ for the different kinds of research that are applicable to different issues and inferences (Shavelson & Towne, 2002). The primary interest of the Panel is experimental and quasi-experimental research designed to investigate the effects of programs, practices, and approaches on students’ mathematics learning and achievement. On some matters, however, the relevant studies are surveys (e.g., of students’ mathematical knowledge). On yet other matters, by necessity, the relevant sources represent compilations of practice and informed opinion (e.g., regarding the mathematical concepts essential to algebra). The methodological quality of individual studies will be categorized as part of the documentation for the database for the Panel’s work, using such definitions as the following.

For studies of the effects of interventions:

High quality. Random assignment to conditions; low attrition; valid and reliable measures.

Moderate quality. Nonrandom assignment to conditions with matching, statistical controls, or a demonstration of baseline equivalence on important variables; low attrition or evidence that attrition effects are small; valid and reliable measures. Correlational modeling with instrumental variables and strong statistical controls. Random assignment studies with high attrition.

Low quality. Nonrandom assignment without matching or statistical controls. Pre-post studies. Correlational modeling without strong statistical controls. Quasi-experimental studies with high attrition.

For descriptive surveys of population characteristics:

High quality. Probability sampling of a defined population; low nonresponse rate or evidence that nonresponse is not biasing; large sample (achieved sample size gives adequate error of estimate for the study purposes); valid and reliable measures.

Moderate quality. Purposive sampling from a defined population; face valid for representativeness; low nonresponse rate; moderate to large sample size; valid and reliable measures. Probability sample with high nonresponse rate, but evidence that nonresponse is not biasing.

Low quality. Convenience sample; high nonresponse rate or evidence that it is biasing; small sample size; invalid or unreliable measures.

For studies of tests and assessments:

Psychometric standards such as measures of validity, reliability, and sensitivity will be used to evaluate tests and assessments (e.g., Anastasi, 1968; Cronbach & Meehl, 1955).

V. Task Group Guidelines

To ensure identification of the best available evidence in the research literature, each task group has developed guidelines for the literature search that identify the relevant topics and the screening criteria to be used to select the studies the task group will consider for review. These criteria are designed to produce full or representative coverage of the highest quality and the most relevant studies in a relatively efficient manner.

A. Learning Processes Task Group

1. Topics and Content

a) Research linking mathematical content and children’s learning, and cognitive processes. Focus on children’s solving or understanding of mathematics in specific content areas (see key words) with measures of children’s learning, problem solving, or understanding that are more precisely defined than is typically found with achievement measures, e.g., trial-by-trial assessment of problem-solving strategy.

2. Coverage

a) Emphasis on the literature found in a designated set of core journals supplemented with studies on specific topics of interest (e.g., whole number division) from other peer-reviewed journals.
b) Reviews of empirical research in books or annual reviews (e.g., *Annual Review of Psychology*, *Handbook of Child Psychology*).
c) Published in English, 1990 or after; supplemented with earlier, high-citation impact work, where available.

3. Study Samples

a) Children 3 years of age to young adult.

4. Study Methods

a) Randomized experiments.
b) Quasi-experiments with nonrandom assignment to conditions.
c) Correlational studies with a measure of math processes that is predicted by or predicts some other achievement outcome or process measures.

B. Conceptual Knowledge and Skills Task Group

1. Topics and Content

a) Topics taught and assessed in mathematics, preschool to eighth grade and algebra, in the United States and internationally.
b) The relationship between math concepts and skills learned or taught at elementary and middle school levels, and later success in algebra (achievement).

2. Coverage

a) State and international curriculum frameworks for preschool to Grade 8 mathematics topics.
b) Course-level expectations in state-based curriculum frameworks for the algebra topics [synthesized by Institute for Defense Analyses Science and Technology Policy Institute (STPI) for 22 states].
c) Contents of algebra textbooks with particular attention to current and historic (1913) algebra topics (synthesized by STPI for 27 textbooks).
d) Pre-algebra (kindergarten through eighth grade) and algebra topics represented in the National Assessment of Educational Progress (NAEP), the Advanced Diploma Project (ADP), and the Singapore Curriculum.

3. Study Samples

a) Students from elementary through high school grades.

4. Study Methods

a) Descriptive (frequency) analysis from representative sets of materials nationally and internationally.
b) Criteria established by manuscript authors (e.g., Fordham report) for state mathematics frameworks.

C. Instructional Practices Task Group

1. Topics and Content

a) Effects of instructional practice, teaching strategies, and instructional materials on mathematics achievement.

2. Coverage

a) Published in a peer-reviewed journal or government report.
b) Published in English, 1976 or after.

3. Study Samples

a) Children, kindergarten through high school level.

4. Study Methods

a) Randomized experiments or quasi-experiments with techniques to control for bias (matching, statistical control) or demonstration of initial equivalence on important pretest variables.
b) Attrition of less than 30% or evidence that the remaining sample is equivalent to the original sample on important variables.

D. Teachers Task Group

1. Topics and Content

a) Relationship between teacher content knowledge and student achievement.
b) Programs of teacher education and professional development, and their effects on teacher knowledge, instructional practice, and student achievement.
c) Programs of mathematics specialist teachers at the elementary level, and effects on instruction and student achievement.
d) Programs to recruit and retain qualified teachers, and their effects on teacher quality.

2. Coverage

a) Published in a peer-reviewed journal or government report.
b) Books and book chapters.
c) Selected reports relevant to key topics.
d) Published in English.

3. Study Samples

a) Teachers of preschool through high school students.

4. Study Methods

a) Randomized experiments.
b) Quasi-experiments with techniques to control for bias (e.g., matching, statistical control) or demonstration of initial equivalence.
c) Correlational studies of natural variation with statistical controls.

VI. Procedures

A. Screening Criteria for the Literature Search

As described in the previous section, each task group has developed specific criteria for identifying and screening the research literature pertinent to its task. Those criteria give priority to high-quality scientific research but also include weaker evidence where it may be promising or suggestive, and when limited high-quality research is available. As such, the search and screening criteria do not provide an assessment of methodological quality per se; they only describe the studies each task group wishes to consider in preparing its review.

B. Documenting the Quality of the Evidence Used in the Report

The individual research studies that are considered part of the relevant research base by each task group will be evaluated as presenting high-, moderate-, or low-quality scientific evidence using the standards appropriate to the nature of the research. For some task groups, this coding will be done by Abt Associates Inc. as part of their documentation of the database of research studies on which the Panel's review is based. The body of research on which each significant claim, conclusion, and recommendation in the report is based will be characterized as strong, suggestive, or weak according to the quality, quantity, and generalizability of the collective evidence across studies. This information will guide the wording of the Final Report with regard to the confidence with which conclusions and recommendations are presented.

VII. Recommendations

The Panel's systematic reviews have yielded hundreds of studies on important topics, but only a small proportion of those studies have met methodological standards. Most studies have failed to meet standards of quality because they do not permit strong inferences about causation or causal mechanisms (Mosteller & Boruch, 2002; Platt, 1964). Many studies rely on self-report, introspection about what has been learned or about learning processes, and open-ended interviewing techniques, despite well-known limitations of such methods (e.g., Brainerd, 1973; Nisbett & Ross, 1980; Woodworth, 1948). Therefore, the Subcommittee on Standards of Evidence recommends that the rigor and amount of course work in statistics and experimental design be increased in graduate training in education. Such knowledge is essential to produce and to evaluate scientific research in crucial areas of national need, including mathematics education.

BIBLIOGRAPHY

Anastasi, A. (1968). *Psychological testing* (3rd ed.). London: Collier-Macmillan Ltd.

Bjork, R.A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe and A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp.185–205). Cambridge, MA: MIT Press.

Brainerd, C.J. (1973). Judgments and explanations as criteria for the presence of cognitive structures. *Psychological Bulletin, 79*, 172–179.

Cook, T.D., & Campbell, D.T. (1979). Quasi-experimentation: Design and analysis for field settings. Chicago: Rand McNally.

Cronbach, L., & Meehl, P. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281–302.

Killeen, P.R. (2005). An alternative to null-hypothesis significance tests. *Psychological Science, 16*, 345–353.

Lewin, K. (1951). Field theory in social science: Selected theoretical papers. New York: Harper & Row.

Mosteller, F., & Boruch, R. (Eds.). (2002). Evidence matters: Randomized trials in education research. Washington, DC: Brookings Institution Press.

Nisbett, R.E., & Ross, L. (1980). Human inference: Strategies and shortcomings of social judgment. Englewood Cliffs, NJ: Prentice-Hall.

Platt, J.R. (1964). Strong inference. *Science, 146*, 347–353.

Roediger, H.L., & Karpicke, J.D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*, 249–255.

Shavelson, R.J., & Towne, L. (2002). Scientific research in education. Washington, DC: National Academies Press.

Woodworth, R.S. (1948). Contemporary schools of psychology (2nd ed.). New York: Ronald Press.

Chapter 3: Report of the Task Group on Conceptual Knowledge and Skills

Francis "Skip" Fennell, Chair
Larry R. Faulkner
Liping Ma
Wilfried Schmid
Sandra Stotsky
Hung-Hsi Wu
Tyrrell Flawn, U.S. Department of Education Staff

CONTENTS

Figures

Tables

Abbreviations

ACT	American College Testing
ADP	American Diploma Project
CLE	course-level expectations
LSAY	High School and Beyond, Longitudinal Survey of American Youth
NAEP	National Assessment of Educational Progress
NCTM	National Council of Teachers of Mathematics
NELS	National Education Longitudinal Study
STPI	Institute for Defense Analyses Science and Technology Policy Institute
SAT	Scholastic Achievement Test
SES	socioeconomic status
TIMSS	Trends in International Math and Science Study

Executive Summary

Introduction

The National Mathematics Advisory Panel was asked to make recommendations on "the critical skills and skill progressions for students to acquire competence in algebra and readiness for higher-level mathematics." To address this particular charge, the Panel established a Task Group on Conceptual Knowledge and Skills (CKS). To guide its inquiry, deliberations, and recommendations, CKS formulated three major questions:

1) What are the major topics of school algebra?
2) What are the essential mathematical concepts and skills that lead to success in Algebra and that should be learned as preparation for Algebra?[1]
3) Does the sequence of topics prior to algebra course work or for algebra course work affect achievement in Algebra?

Methodology

The Panel was charged with determining how to use "the results of research relating to proven-effective and evidence-based mathematics instruction" and making recommendations "based on the best available scientific evidence." The Panel contracted with Abt Associates Inc. to survey the research literature for studies that addressed each task group's major questions and met standards of methodological quality.

The Task Group's literature review yielded some peer-reviewed and published studies that met standards of methodological quality and were relevant to the work of this Task Group, especially with respect to its third question. However, because of the small number of such studies, the Task Group decided to include reports that presented the best available evidence on the topic of the conceptual knowledge and skills needed for success in Algebra. Thus it supplemented the literature review with reports by national organizations and government agencies, and with analyses and comparisons of state curriculum frameworks and school textbooks developed for the Task Group by the Institute for Defense Analyses Science and Technology Policy Institute. Where there may be differences among these reports, studies, or analyses, the differences are so noted. The Task Group's recommendations on matters of definition and mathematical content were also guided by professional judgment.

[1] Algebra will be capitalized when it is referred to as a course.

Results and Conclusions

The Major Topics in School Algebra

The Major Topics in School Algebra that were developed by the Task Group on Conceptual Knowledge and Skills are shown in this section. The teaching of Algebra, like the teaching of all of school mathematics, must ensure that students are proficient in computational procedures, can reason logically and clearly, and can formulate and solve problems. For this reason, the topics listed below should not be regarded as a sequence of disjointed items, simply to be committed to memory. On the contrary, teachers and textbook writers should emphasize the connections as well as the logical progression among these topics. The topics comprise both core and foundational elements of school algebra—those elements needed for study of school algebra itself and those elements needed for study of more advanced mathematics courses. The total amount of time spent on covering them in single-subject courses is normally about 2 years, although algebra content may be and is often structured in other ways in the secondary grades. What is usually called Algebra I would, in most cases, cover the topics in Symbols and Expressions, and Linear Equations, and at least the first two topics in Quadratic Equations. The typical Algebra II course would cover the other topics, although the last topic in Functions (Fitting Simple Mathematical Models to Data), the last two topics in Algebra of Polynomials (Binomial Coefficients and the Binomial Theorem), and Combinatorics and Finite Probability are sometimes left out and then included in a precalculus course. It should be stressed that this list of topics reflects professional judgment as well as a review of other sources.

Symbols and Expressions

- Polynomial expressions
- Rational expressions
- Arithmetic and finite geometric series

Linear Equations

- Real numbers as points on the number line
- Linear equations and their graphs
- Solving problems with linear equations
- Linear inequalities and their graphs
- Graphing and solving systems of simultaneous linear equations

Quadratic Equations

- Factors and factoring of quadratic polynomials with integer coefficients
- Completing the square in quadratic expressions
- Quadratic formula and factoring of general quadratic polynomials
- Using the quadratic formula to solve equations

Functions

- Linear functions
- Quadratic functions—word problems involving quadratic functions
- Graphs of quadratic functions and completing the square
- Polynomial functions (including graphs of basic functions)
- Simple nonlinear functions (e.g., square and cube root functions; absolute value; rational functions; step functions)
- Rational exponents, radical expressions, and exponential functions
- Logarithmic functions
- Trigonometric functions
- Fitting simple mathematical models to data

Algebra of Polynomials

- Roots and factorization of polynomials
- Complex numbers and operations
- Fundamental theorem of algebra
- Binomial coefficients (and Pascal's Triangle)
- Mathematical induction and the binomial theorem

Combinatorics and Finite Probability

- Combinations and permutations, as applications of the binomial theorem and Pascal's Triangle

Critical Foundations of Algebra

The Task Group also presents three clusters of concepts and skills that it considers foundational for formal algebra course work:

1) Fluency With Whole Numbers,
2) Fluency With Fractions, and
3) Particular Aspects of Geometry and Measurement.

To prepare students for Algebra, the curriculum must simultaneously develop conceptual understanding, computational fluency, and problem-solving skills. These three aspects of learning are mutually reinforcing and should not be seen as competing for class time. The Critical Foundations identified and discussed below are not at all meant to comprise a complete preschool-to-algebra curriculum; the Task Group merely aims to recognize the Critical Foundations for the study of Algebra, whether as part of a dedicated algebra course in the seventh, eighth, or ninth grade, or within an integrated mathematics sequence in the middle and high school grades. However, these Critical Foundations *do* deserve ample time in any mathematics curriculum. The foundations are presented in three distinct clusters of concepts and skills, each of which should incorporate the three aspects of learning noted here.

Fluency With Whole Numbers

By the end of the elementary grades, children should have a robust sense of number. This sense of number must include understanding place value, and the ability to compose and decompose whole numbers. It must clearly include a grasp of the meaning of the basic operations of addition, subtraction, multiplication, and division, including use of the commutative, associative, and distributive properties; the ability to perform these operations efficiently; and the knowledge of how to apply the operations to problem solving. Computational facility rests on the automatic recall of addition and related subtraction facts, and of multiplication and related division facts. It requires fluency with the standard algorithms for addition, subtraction, multiplication, and division. Fluent use of the algorithms not only depends on the automatic recall of number facts but also reinforces it. A strong sense of number also includes the ability to estimate the results of computations and thereby to estimate orders of magnitude, e.g., how many people fit into a stadium, or how many gallons of water are needed to fill a pool.

Fluency With Fractions

Before they begin algebra course work, middle school students should have a thorough understanding of positive as well as negative fractions. They should be able to locate both positive and negative fractions on the number line; represent and compare fractions, decimals, and related percents; and estimate their size. They need to know that sums, differences, products, and quotients (with nonzero denominators) of fractions are fractions, and they need to be able to carry out these operations confidently and efficiently. They should understand why and how (finite) decimal numbers are fractions and know the meaning of percentages. They should encounter fractions in problems in the many contexts in which they arise naturally, for example, to describe rates, proportionality, and probability. Beyond computational facility with specific numbers, the subject of fractions, when properly taught, introduces students to the use of symbolic notation and the concept of generality, both being integral parts of Algebra.

Particular Aspects of Geometry and Measurement

Middle-grade experience with similar triangles is most directly relevant for the study of Algebra: Sound treatments of the slope of a straight line and of linear functions depend logically on the properties of similar triangles. Furthermore, students should be able to analyze the properties of two- and three-dimensional shapes using formulas to determine perimeter, area, volume, and surface area. They should also be able to find unknown lengths, angles, and areas.

Benchmarks for the Critical Foundations

In view of the sequential nature of mathematics, the Critical Foundations of Algebra described in the previous section require judicious placement in the grades leading up to Algebra. For this purpose, the Task Group suggests the following benchmarks as guideposts for state frameworks for school districts. There is no empirical research on the placement of these benchmarks, but they find justification in a comparison of national and international curricula. The benchmarks should be interpreted flexibly, to allow for the needs of students and teachers.

Fluency With Whole Numbers

1) By the end of Grade 3, students should be proficient with the addition and subtraction of whole numbers.
2) By the end of Grade 5, students should be proficient with multiplication and division of whole numbers.

Fluency With Fractions

1) By the end of Grade 4, students should be able to identify and represent fractions and decimals, and compare them on a number line or with other common representations of fractions and decimals.
2) By the end of Grade 5, students should be proficient with comparing fractions and decimals and common percents, and with the addition and subtraction of fractions and decimals.
3) By the end of Grade 6, students should be proficient with multiplication and division of fractions and decimals.
4) By the end of Grade 6, students should be proficient with all operations involving positive and negative integers.
5) By the end of Grade 7, students should be proficient with all operations involving positive and negative fractions.
6) By the end of Grade 7, students should be able to solve problems involving percent, ratio, and rate, and extend this work to proportionality.

Particular Aspects of Geometry and Measurement

1) By the end of Grade 5, students should be able to solve problems involving perimeter and area of triangles, and all quadrilaterals having at least one pair of parallel sides (i.e., trapezoids).
2) By the end of Grade 6, students should be able to analyze the properties of two-dimensional shapes and solve problems involving perimeter and area. They should also be able analyze the properties of three-dimensional shapes and solve problems involving surface area and volume.
3) By the end of Grade 7, students should understand relationships involving similar triangles.

To address the question of whether the sequence of topics prior to formal algebra course work or how formal algebra course work affects achievement in algebra, the Task Group examined three related sub-questions. It first looked for evidence on the effectiveness of currently used elementary and middle school mathematics curricula (including their sequence of topics) for achievement in Algebra. It found no research demonstrating that a specific multigrade sequence of mathematics topics assures success in Algebra.

The Task Group also sought evidence on whether an integrated approach or a single-subject sequence might be more effective for formal algebra course work and more advanced mathematics course work. It found no clear body of research from which one may draw conclusions.

Finally, the Task Group sought to locate evidence on the benefits or disadvantages of teaching the content of an Algebra I course to a broad range of students before Grade 9. The Task Group found a positive relationship between taking Algebra in Grade 7 or 8 and later high school mathematics achievement, regardless of students' prior achievement and school and student characteristics.

Recommendations

This Task Group affirms that Algebra is the gateway to more advanced mathematics and to most postsecondary education. All schools and teachers of mathematics must concentrate on providing a solid mathematics education to all elementary and middle school students so that all of them can enroll and succeed in Algebra. Students need to be soundly prepared for Algebra and then well taught in Algebra, regardless of the grade level at which they study it. To improve the teaching of Algebra, the Task Group proposes the following eight recommendations:

1) The Task Group recommends that school algebra be consistently understood in terms of the Major Topics of School Algebra given in this report.

2) The Major Topics of School Algebra in this report, accompanied by a thorough elucidation of the mathematical connections among these topics, should be the main focus of Algebra I and Algebra II standards in state curriculum frameworks, in Algebra I and Algebra II courses, in textbooks for these two levels of Algebra whether for integrated curricula or otherwise, and in end-of-course assessments of these two levels of Algebra. The Task Group also recommends use of the Major Topics of School Algebra in revisions of mathematics standards at the high school level in state curriculum frameworks, in high school textbooks organized by an integrated approach, and in grade-level state assessments using an integrated approach at the high school, by Grade 11 at the latest.

3) Proficiency with whole numbers, fractions, and particular aspects of geometry and measurement are the Critical Foundation of Algebra. Emphasis on these essential concepts and skills must be provided at the elementary- and middle-grade levels. The coherence and sequential nature of mathematics dictate the foundational skills that are necessary for the learning of algebra. By the nature of algebra, the most important foundational skill is proficiency with fractions (including decimals, percent, and negative fractions). The teaching of fractions must be acknowledged as critically important and improved before an increase in student achievement in Algebra can be expected.

4) The Benchmarks proposed by the Task Group should be used to guide classroom curricula, mathematics instruction, and state assessments. They should be interpreted flexibly, to allow for the needs of students and teachers.

5) International studies show that high-achieving nations teach for mastery in a few topics, in comparison with the U.S. mile-wide-inch-deep curriculum. A coherent progression, with an emphasis on mastery of key topics, should become the norm in elementary and middle school mathematics curricula. There should be a de-emphasis on a spiral approach in mathematics that continually revisits topics year after year without closure.

6) Federal and state policies should give incentives to schools to offer an authentic Algebra I course in Grade 8, and to prepare a higher percentage of students to study the content of such a course by the beginning of Grade 8. The word “authentic” is used here as a descriptor of a course that addresses algebra in the manner of Recommendation 2. Students must be prepared with the mathematical prerequisites for this course in the sense of Recommendation 3.

7) Publishers must ensure the mathematical accuracy of their materials. Those involved with developing mathematics textbooks and related instructional materials need to engage mathematicians, as well as mathematics educators, in writing, editing, and reviewing these materials.

8) Adequate preparation of students for Algebra requires their teachers to have a strong mathematics background. To this end, the Major Topics of School Algebra and the Critical Foundations of Algebra must be fundamental in the mathematics preparation of elementary and middle school teachers. Teacher education programs and licensure tests for early childhood teachers (preschool–Grade 3) should focus on the Critical Foundations of Algebra; for elementary teachers (Grades 1–5), on the Critical Foundations of Algebra and those algebra topics typically covered in an introductory Algebra course; and for middle school teachers (Grades 5–8), on the Critical Foundations of Algebra and all of the Major Topics of School Algebra.

I. Introduction

The learning of mathematics at the elementary- and middle-grade levels forms the basis for achievement in high school, college mathematics, and college courses using mathematics, and for the broad range of mathematical skills used in the workplace. Yet, on average, American students do not do as well on international mathematics tests as their peers in many developed countries. And while scores on the National Assessment of Educational Progress (NAEP) mathematics tests at Grades 4 and 8 are higher than they have ever been, a large majority of students do not score at the "proficient" and "advanced" levels (U.S. Department of Education, 2007). Nor are American students uniformly taught mathematics by teachers with an adequate grasp of the mathematics they teach (e.g., Loveless, 2004). There is broad agreement that mathematics education in the schools needs to be strengthened.

One of the major charges to the National Mathematics Advisory Panel concerned preparation for course work in algebra. A strong grasp of algebra is essential for successful participation in the contemporary American workforce; it is also necessary for entry into higher education and for the pursuit of advanced mathematics in general (e.g., Business Roundtable, 2006). The Panel was asked to make recommendations on "the critical skills and skill progressions for students to acquire competence in algebra and readiness for higher level mathematics" (Executive Order No. 13398). To address this particular charge, the Panel established a Task Group on Conceptual Knowledge and Skills. This report is the Task Group's response to this charge and describes the essential mathematical concepts and skills that students should acquire prior to and during the study of algebra.

To guide its inquiry and deliberations, the Task Group formulated three major questions:

1) What are the major topics of school algebra?[2]
2) What are the essential mathematical concepts and skills that lead to success in Algebra and that should be learned as preparation for Algebra?[3]
3) Does the sequence of topics prior to algebra course work or for algebra course work affect achievement in Algebra?

II. Methodology

The Panel was charged with determining how to use "the results of research relating to proven-effective and evidence-based mathematics instruction" and making recommendations "based on the best available scientific evidence." The Panel contracted with Abt Associates Inc. to survey the research literature for studies that addressed each task group's major questions and met standards of methodological quality.

[2] "School algebra" is a term chosen to encompass the full body of algebraic material that the Task Group expects to be covered through high school, regardless of its organization into courses and levels.
[3] Algebra will be capitalized when it is referred to as a course.

The Task Group's literature review yielded some peer-reviewed and published studies that met standards of methodological quality and were relevant to the work of this Task Group, especially with respect to its third question. However, because of the small number of such studies, the Task Group decided to include other types of reports that presented the best available evidence on the topic of the conceptual knowledge and skills needed for success in Algebra. Thus it supplemented the review of research literature with reports by national organizations and government agencies, and with analyses and comparisons of state curriculum frameworks and school textbooks developed for the Task Group by the Institute for Defense Analyses Science and Technology Policy Institute (henceforth STPI).[4] Where there may be differences among these reports, studies, or analyses, the differences are so noted. The Task Group's recommendations on matters of definition and mathematical content were also guided by professional judgment.

III. Student Achievement in Mathematics

During a time of decline in test scores on the Scholastic Achievement Test (SAT) in the 1960s, Congress legislated for the creation of NAEP. It has been the only nationally representative and continuing assessment of what American students, through high school, know and can do in all major subject areas: reading, mathematics, science, writing, U.S. history, civics, geography, and the arts. Using nationally representative samples, NAEP provides information on the academic achievement of nationwide and state populations and subpopulations, not on individual students or schools.

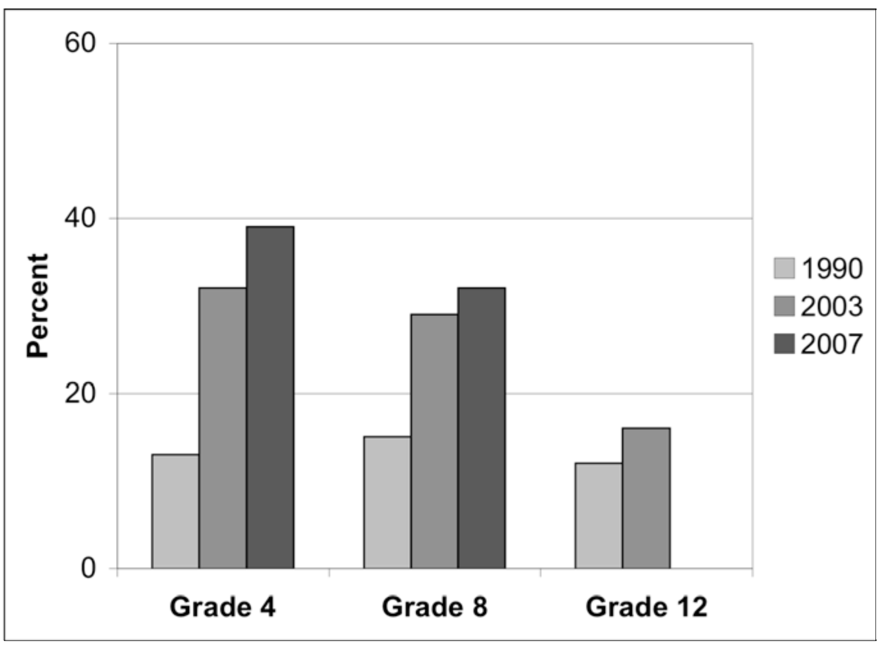
There are two types of NAEP tests. The long-term trend NAEP tests, which have been given since the late 1960s and early 1970s, assess students by age (9, 13, and 17). The main NAEP tests, which have been given since 1990, assess students in Grades 4, 8, and 12, instead of by age. For the long-term trend assessments, the same tests have been given under the same conditions since 1978. The main tests have been used since 1990 to create a second trend line that reflects more current practices in mathematics curriculum and assessment. The main NAEP tests in mathematics assess five areas: numbers and operations, measurement, geometry, data analysis and probability, and algebra; these tests have been governed by a framework that parallels the content strands in *Curriculum and Evaluation Standards for School Mathematics,* issued by the National Council of Teachers of Mathematics (NCTM) in 1989. The long-term trend NAEP tests in mathematics focus on essential concepts and skills in four areas: numbers and operations, measurement, geometry, and algebra. On both types of tests, students respond to questions of three types: multiple choice, short answer, and extended answer. However, on the long-term trend tests, most of the items are multiple choice. There is greater emphasis on extended responses on the main tests.

Student scores on both types of tests present a mixed picture of national achievement in mathematics, although the results in Grade 4 seem to allow for some optimism. As Figure 1 shows, average student scores on the main NAEP tests have increased considerably

[4] The Institute for Defense Analyses Science and Technology Policy Institute provided technical support to the National Mathematics Advisory Panel through a task order contract initiated in August 2006, Contract Number OIA-0408601 and Task Order OSTP-20-0001.

since 1990 at Grade 4 and to a lesser extent at Grade 8.[5] There have been increases at both grade levels on the long-term trend tests as well—not to the same extent as the increases on the main NAEP tests, but corroborating the trend. However, for Grade 4, U.S. results on the 1995 and 2003 Trends in International Math and Science Study (TIMSS) tests are exactly the same, raising questions about the large increase from 1990 to 2007 on the main NAEP tests (Mullis, Martin, Gonzalez, & Chrostowski, 2004).[6] Moreover, the percentage of students scoring at the proficient or advanced levels at Grade 4 and Grade 8 on the 2003 NAEP tests is well below 40%; at Grade 12, it is below 20%.[7] Not only is it not clear how valid the terms “proficient” and “advanced” are, it is also not clear how academically significant the gains are. Loveless (2004) conducted two analyses of the “arithmetic load” of 512 released items from NAEP’s mathematics tests to determine their level of difficulty. Among items assessing “problem solving,” he found that Grade 8 items were only slightly more difficult than the Grade 4 items, with many items testing arithmetic skills typically taught in Grades 1 and 2. He further noted that for these same items the algebra strand is the least-challenging content strand at both grades. In an analysis of calculator use on all released items in the number sense and algebra strands, he found that students are allowed to use calculators on items involving anything more difficult than whole numbers.

Figure 1: Percentage of Students At or Above Proficient in Mathematics Achievement On Main NAEP Test: 1990, 2003, and 2007



Source: Institute for Defense Analyses Science and Technology Policy Institute tabulations using the NAEP Data Explorer, available at: http://nces.ed.gov/nationsreportcard/nde/.

[5] See http://nces.ed.gov/nationsreportcard/nde/.
[6] The average was 518 in both years and the average age was 10.2 in both.
[7] Figure 1 shows the percent of students at or above proficient in mathematics achievement on the main NAEP test in 1990, 2003, and 2007 for Grade 4 and Grade 8. The percents in Grade 4 go from 13% in 1990, to 32% in 2003, to 39% in 2007. The percents in Grade 8 go from 15% in 1990, to 29% in 2003, to 32% in 2007. Grade 12 data are not available for 2007 as that grade was not tested that year.

It is also not clear that increased enrollment in advanced mathematics courses at the high school level is as academically significant as it appears. According to the U.S. Department of Education, the percentage of high school graduates completing Algebra II or higher rose from 39% in 1990 to 49% in 2005, with the number of students completing Calculus doubling from 1990 to 2005 (U.S. Department of Education, 1990–2005). Yet, test scores for high school students are flat on both the main and long-term trend NAEP tests, and there is no evidence that high school students are beginning their freshman year in college with stronger preparation for mathematics courses. In fact, almost one-fourth of the students enrolled in postsecondary education nationwide require placement in remedial mathematics courses, with the percentage varying by state and according to whether students are enrolled in 2-year or 4-year postsecondary institutions (U.S. Department of Education, 2004).

IV. What Are the Major Topics of School Algebra?

In response to this question, the Task Group reviewed the algebra topics addressed in several sources. They examined the algebra topics 1) in current state standards for Algebra I and Algebra II courses and for integrated curricula, 2) in current textbooks for school algebra and integrated mathematics, 3) in the algebra objectives in NAEP's 2005 Grade 12 mathematics assessment, 4) in the American Diploma Project's benchmarks for a high school exit test and its forthcoming Algebra II end-of-course test, and 5) in the algebra standards in Singapore's mathematics curriculum for Grades 7 through 10. The Task Group also developed its own list of major topics of school algebra. The Task Group presents its list first, together with a brief explanation of the logical connections among these topics, and then shows how this list compares with the algebra topics in these other sources.

A. The Major Topics of School Algebra

1. Introduction to the Topics

The Major Topics of School Algebra that were developed by the Task Group on Conceptual Knowledge and Skills are shown in this section. The teaching of Algebra, like the teaching of all school mathematics, must ensure that students are proficient in computational procedures, can reason logically and clearly, and can formulate and solve problems. For this reason, the topics listed below should not be regarded as a sequence of disjointed items, simply to be committed to memory. On the contrary, teachers and textbook writers should emphasize the connections as well as the logical progression among these topics. They comprise both core and foundational elements of school algebra—those elements needed for study of school algebra itself and those elements needed for study of more advanced mathematics courses. The total amount of time spent on covering them in single-subject courses is normally about 2 years, although algebra content may be and is often structured in other ways in the secondary grades. What is usually called Algebra I would, in most cases, cover the topics in Symbols and Expressions and Linear Equations and at least the first two topics in Quadratic Equations. The typical Algebra II course would cover the other topics, although the last topic in Functions (Fitting Simple Mathematical Models to Data), the last two topics in Algebra of Polynomials (Binomial Coefficients and the Binomial Theorem), and Combinatorics and Finite Probability are sometimes left out and then included in a precalculus course. It should be stressed that this list of topics reflects professional judgment as well as a review of what is in other sources.[8]

The Major Topics of School Algebra

Symbols and Expressions

- Polynomial expressions
- Rational expressions
- Arithmetic and finite geometric series

Linear Equations

- Real numbers as points on the number line
- Linear equations and their graphs
- Solving problems with linear equations
- Linear inequalities and their graphs
- Graphing and solving systems of simultaneous linear equations

Quadratic Equations

- Factors and factoring of quadratic polynomials with integer coefficients
- Completing the square in quadratic expressions
- Quadratic formula and factoring of general quadratic polynomials
- Using the quadratic formula to solve equations

Functions

- Linear functions
- Quadratic functions—word problems involving quadratic functions
- Graphs of quadratic functions and completing the square
- Polynomial functions (including graphs of basic functions)
- Simple nonlinear functions (e.g., square and cube root functions; absolute value; rational functions; step functions)
- Rational exponents, radical expressions, and exponential functions
- Logarithmic functions
- Trigonometric functions
- Fitting simple mathematical models to data

Algebra of Polynomials

- Roots and factorization of polynomials
- Complex numbers and operations
- Fundamental theorem of algebra
- Binomial coefficients (and Pascal's Triangle)
- Mathematical induction and the binomial theorem

Combinatorics and Finite Probability

- Combinations and permutations as applications of the binomial theorem and Pascal's Triangle

[8] An expanded version of this overview can be found in an article, *The Major Topics of School Algebra* at http://math.berkeley.edu/~wu/ and http://math.harvard.edu/~schmid/.

2. Overview of School Algebra

Although the general reader may also find the following overview useful, the Task Group briefly explains in this section the major concepts of school algebra primarily for mathematics teachers and textbook publishers. Some common pitfalls in the classroom are also pointed out. Because textbooks often omit the mathematical connections between basic concepts and skills, a good part of the discussion is devoted to these connections. The Task Group believes that it is impossible to attain a basic understanding of algebra without a grasp of such connections. One consequence of this focus is that the discussion of word problems will be somewhat abbreviated. In no way, however, should this emphasis be interpreted to mean that problem solving is considered to be less important than these connections. Indeed, the solution of multistep word problems should be part of students' routine.

a. Symbols and Expressions

Without any doubt, the foundational skill of algebra is fluency in the use of symbols. A letter or symbol, e.g., x, is used to represent a number in the same way that the pronoun "she" is used to stand for a female. Unless the context makes it absolutely clear, it is necessary in each situation to state what kind of number x represents, e.g., a positive integer or a rational number. The importance of clearly specifying what each letter stands for cannot be overemphasized. For example, the commutative law of multiplication for whole numbers can be stated as follows: for any two whole numbers m and n, $mn = nm$. As is customary, we omit the multiplication symbol x when letters are being multiplied. The truth of this statement is easily verified by noting that a rectangular array of m rows of n dots has the same number of dots as an array of n rows of m dots. However, if one now allows m and n to stand for any two *real numbers*, then the truth of the same equality would be far less simple. It would depend on a precise definition of what a real number is.[9] Thus the meaning of a symbolic statement such as $mn = nm$ depends strongly on what the symbols m and n stand for. A symbolic statement in which the meaning of the symbols is not explained is not acceptable in mathematics. This is perhaps the most basic protocol in the use of symbols: **Always specify precisely what each symbol stands for.**

[9] In school mathematics, defining a real number as a point on the **number line** is a workable compromise. This is, in fact, one argument for emphasizing the importance of the number line in the school mathematics curriculum. For example, it is far from clear why $\sqrt{2}\sqrt{3}=\sqrt{3}\sqrt{2}$, because it is difficult to claim that one knows what these numbers $\sqrt{2}$ and $\sqrt{3}$ are. To say that $\sqrt{2}$ is the number whose square is 2 is to beg the question of how one can be sure there is such a number. If one tries to write down $\sqrt{2}$ by giving its decimal expansion, then one can give no more than its initial segment, e.g., 1.414213562373, but not the *complete* expansion. The same remark applies to $\sqrt{3}$. Such being the case, what does it mean to "multiply" these two numbers when one cannot even be certain of their very existence? And why are $\sqrt{2}\sqrt{3}$ and $\sqrt{3}\sqrt{2}$ equal?

A slightly different example in the use of symbols is the following. Suppose a student tries to solve the simple linear equation $2x = 3$ in x.[10] The solution, as is well known, is $\frac{3}{2}$. If the equation is $2x = 5$, then the solution is $\frac{5}{2}$. If the equation is $7x = 3$, then the solution is $\frac{3}{7}$. And so on. Students soon notice that, regardless of the specific numbers 2 and 3 used in the first equation, the numbers 2 and 5 in the second equation, or the numbers 7 and 3 used in the third equation, the solution is always the quotient of the second number by the first. This suggests the following symbolic presentation for presenting this idea succinctly and correctly: Given fixed numbers a and b, with $a \neq 0$, the solution to the equation $ax = b$ is $\frac{b}{a}$. Note that the number a in the equation $ax = b$ is the same number in the solution $\frac{b}{a}$. The same is true for the number b. When a symbol stands for exactly the same number throughout a discussion, it is called a **constant**. If a symbol is allowed to stand for a collection of numbers in a given discussion, such as the whole numbers or the rational numbers, that symbol is called a **variable**. For example, the m and n in the statement $mn = nm$, of the commutative law of multiplication for whole numbers, are variables.

When teaching introductory Algebra, it is important to give students the correct concept of a variable as a symbol used in the way described above; the common but misleading concept of a variable as "a quantity that varies" should be avoided. Unfortunately, the incorrect definition of a variable as "a quantity that varies" is usually presented right at the beginning of school algebra.

The following are some of the standard concepts associated with the use of symbols. Let x be any number. The number obtained by performing repeated arithmetic operations and taking roots, e.g., $\frac{1+x}{7}(\sqrt{5-x^2})$, is usually called a **symbolic expression** in x. Two of the most common kinds of symbolic expression are polynomials and rational expressions. An expression in x of the form $a_5x^5 + a_4x^4 + \ldots + a_1x + a_0$ where the $a_5, a_4, \ldots, a_0$ are constants, is called a **polynomial** of degree 5 in x with **coefficients** $a_5, \ldots, a_0$. A **rational expression** in x is a quotient of two polynomials in x, e.g., $\frac{x^6 - 7x^3}{x^4 + x + 2}$. If an equality between two symbolic expressions in x is valid for *all* possible values of x under discussion, then the equality is called an **identity**. For example, the first of the following is an identity for all positive integers n and the second is an identity for all numbers x, y:

$$1 + 2 + 3 + \ldots + n = \frac{1}{2}n(n+1) \text{ for all positive integers } n,$$

$$(x + y)(x - y) = x^2 - y^2 \text{ for all numbers } x \text{ and } y.$$

[10] The concept of "solving an equation," which is explained later, is used only to illustrate one aspect of the use of symbols.

To verify these and other identities, one must remember that they are, above all, statements about *numbers* and, as such, all the knowledge about numbers that students bring to Algebra can be put to use for their verification. There is one caveat, however. Because students generally no longer have specific values of numbers for explicit arithmetic calculations (e.g., 15 × 4 = 60), all the calculations must now be carried out using only what is true for *all* numbers, regardless of what they are, i.e., the associative laws and commutative laws of addition and multiplication, and the distributive law.

One of the most important identities in introductory Algebra is

$$(x - y)(x^n + x^{n-1}y + \ldots + xy^{n-1} + y^n) = x^{n+1} - y^{n+1}$$

for all numbers x and y, and for all positive integers n. (This is a generalization of the second identity above.) The verification of this identity is a straightforward calculation. When students let $x = 1$, they get

$$(1 - y)(1 + y + y^2 + \cdots + y^{n-1} + y^n) = 1 - y^{n+1}$$

so that

$$1 + y + y^2 + \cdots + y^{n-1} + y^n = \frac{y^{n+1} - 1}{y - 1},$$

which is valid for all numbers y not equal to 1 and for all positive integers n. This is the **summation formula for finite geometric series**. This summation formula is important in mathematics and in both the natural and social sciences. The fact that it is an elementary result that can be taught at the beginning of school algebra is not generally recognized. It should be taught early because, when it is relegated to the end of Algebra II, as is done in the standard curriculum, it does not receive enough attention and is often omitted. The result is that many students go to college missing a critical piece of information.

b. Linear and Quadratic Equations

The most immediate application of the use of symbols is the solution of equations, usually linear and quadratic ones at the beginning, but more complicated ones as more sophisticated functions are defined. This discussion begins with **equations in one variable**, i.e., considering first those equations involving only one variable. The usual terminology of "solving" $3x - 1 = 4 + x$, for example, is an abbreviation of the longer statement of "determining for which numbers x the equality $3x - 1 = 4 + x$ is valid." Any such values of x are then called the **solutions** of the equation $3x - 1 = 4 + x$. For example, 5 is not a solution of $3x - 1 = 4 + x$, as $14 \neq 9$, but $\frac{5}{2}$ is a solution. Students will soon learn that the latter is the only solution possible.

Note that the equation $3x - 1 = 4 + x$ arises naturally if students try to answer the question: "What is the number if 1 less than 3 times the number is equal to 4 more than itself?" If students let x be this unknown number, then a direct symbolic transcription of the verbal data leads directly to $3x - 1 = 4 + x$. For this reason, the symbol x in an equation is sometimes called an **unknown**. In school algebra, it is therefore important to learn not only how to solve equations, but also to correctly transcribe verbal information into symbolic information. If standard textbooks on algebra serve as any indication, the latter objective has perhaps not been given the attention it deserves in the algebra classroom.

It is common for students to be taught to solve an equation by manipulating the symbolic expressions, where a symbolic expression is regarded as something distinct from anything they have ever encountered. A consequence is that ad hoc concepts, such as balancing an equation in x, have to be introduced to justify the method of solution. This line of thinking is not correct from a mathematical standpoint. The correct way to solve the equation $3x - 1 = 4 + x$, for example, involves nothing more than the consideration of numbers. The principle underlying the following solution method is applicable, not just to a linear equation, but to all equations. As a first step, students should begin by **assuming that there is a solution** to the equation. If the equation turns out to have no solution, then one would arrive at a contradiction and would know it had no solution. But if it does have a solution, one would be able to see what this solution must be. Therefore, one can call this putative solution x'. Then $3x' - 1 = 4 + x'$; note that this is an equality of two *numbers* and is therefore something with which students are completely familiar. What they do next is to try to deduce, under the assumption that they already have such a solution x', what this number x' has to be. Adding $-x' + 1$ to both sides of $3x' - 1 = 4 + x'$ to bring both terms involving x' to the left and the constants to the right, we obtain $(3x' - 1) + (-x' + 1) = (4 + x') + (-x' + 1)$, so that $2x' = 5$ and therefore $x' = \frac{5}{2}$.

Students have proved, *by using facts about numbers and nothing else*, that *if* there is a solution x', then it must be $\frac{5}{2}$. Notice that this does not say, as yet, that $\frac{5}{2}$ *is* a solution of $3x - 1 = 4 + x$. However, by verifying directly that $3(\frac{5}{2}) - 1 = 4 + \frac{5}{2}$, students reach the desired conclusion that, indeed, $\frac{5}{2}$ is a solution. It follows that $\frac{5}{2}$ is the only solution because students have already shown that any solution has to be $\frac{5}{2}$. The same reasoning shows why the solution of a **general linear equation,** $ax + b = cx + d$ (a, b, c, d being constants, and $a \neq c$), is the number $\frac{d - b}{a - c}$.

This method of solution is applicable to any equation, and, in particular, to a **quadratic equation** of one variable $ax^2 + bx + c = 0$ (a, b, c are constants and $a \neq 0$). Thus, assuming there is a solution (which is often called a **root**), students deduce what it must be, and then verify directly that any such possibility (depending on the values of the constants *a, b, c,* there may be one or two solutions) is indeed a solution. One particular detail of the solution is, however, of great interest, and it is the use of the technique of **completing the square.** This reduces the quadratic polynomial $ax^2 + bx + c$ to a simple form, from which the equation $ax^2 + bx + c = 0$ can be readily solved. The well-known **quadratic formula** is the result.

The skill of completing the square can be traced back to the Babylonians of 4,000 years ago. It is a useful skill in its own right and is the key idea that will lead to a complete clarification of the graph of a quadratic function, to be taken up in the section in this report on functions.

In the event that the quadratic polynomial $ax^2 + bx + c$ can be factored as a product, $ax^2 + bx + c = a(x - r_1)(x - r_2)$ for some numbers r_1 and r_2, then it is clear that r_1 and r_2 are the solutions of $ax^2 + bx + c = 0$. Less well known and less obvious is the *converse* of this statement, namely the fact that if r_1, r_2 are the solutions of $ax^2 + bx + c = 0$, then $ax^2 + bx + c = a(x - r_1)(x - r_2)$. This fact can be proven by using the quadratic formula, although it is not always done in textbooks. Of greater importance is the light this fact throws on the issue of factoring quadratic polynomials. When the coefficients *a*, *b*, *c* of $ax^2bx + c$ are integers, factoring such polynomials is sometimes elevated to an important skill in introductory Algebra. While some skill along this line is desirable, what this discussion has shown is that it is not necessary to emphasize it, because all such factoring can be done easily by using the quadratic formula to first locate the roots.

Next for consideration is the case of a **linear equation of two variables**, $ax + by = c$ where *a*, *b*, *c* are constants, and both *a* and *b* are not 0. A **solution** of the equation is by definition an ordered pair of numbers (x_0, y_0), so that they **satisfy the equation** in the sense that $ax_0 + by_0 = c$. This definition of a solution suggests that the collection of all solutions of $ax + by = c$ should be identified with a subset of the coordinate plane. Indeed, the **graph** of $ax + by + c = 0$ is defined to be the subset of the plane consisting of all solutions of the equation. One then proves that the graph is a straight line (or more simply, a **line**), and conversely, every straight line is the graph of an (essentially unique) linear equation of two variables. In the latter case, the equation is referred to as **the equation of the line**. The key ingredient in the proof of both facts is the concept of the **slope** of a line: Given a line *L* in the coordinate plane, its slope is the quotient $\frac{y_2 - y_1}{x_2 - x_1}$ where (x_1, y_1) and (x_2, y_2) are *any* two points on the straight line *L*. The fact that this quotient remains the same no matter which two points, (x_1, y_1) or (x_2, y_2), are chosen is stated with not a hint of justification in almost all algebra textbooks currently in use. When students are denied access to this reasoning, it is difficult, if not impossible, for them to understand the relationship between a linear equation of two variables and its graph. The result is that many students consider the different forms of the equation of a line (e.g., point-slope form, two-point form) a mystery and are confused by related computations.

The proof that the definition of a slope of a line is independent of the choice of the two points depends on considerations of similar triangles. It is therefore vitally important that students be given the opportunity to become familiar with the basic facts of similar triangles before studying algebra. This should include the fact that corresponding sides of similar triangles are in proportion, or that if two triangles have two pairs of equal angles, they are similar. Students can defer the proofs of these theorems to a later course on Euclidean geometry, but they need to be comfortable using them. Students will commonly be asked to use certain theorems before they learn why the theorems are true, (e.g., lessons on the Pythagorean Theorem or the sum of angles of a triangle as 180 degrees). Mathematics learning does not have to be *formally* linear.

With the correct definition of slope available, students are in a position to understand the relationship between slopes of lines, and the concepts of parallelism and perpendicularity. This understanding has a strong bearing on the study of simultaneous linear equations in two variables. Using the precise definition of the graph of a linear equation of two variables, one can *prove* that the solution of a pair of simultaneous linear equations is the point of intersection of the graphs of the two equations. If the graphs are parallel, they do not intersect, and therefore there is no solution to the simultaneous equations. When the parallelism of the graphs is translated into the language of slope, students arrive at the criterion for the solvability of simultaneous equations in terms of the determinant of the system of equations.

Associated with a linear equation in two variables, $ax + by = c$, are the **linear inequalities** $ax + by \geq c$ and $ax + by \leq c$. It suffices to discuss one of them, say $ax + by \leq c$. Again, the need for the definition of the **graph of a linear inequality** should be emphasized: the graph of $ax + by \leq c$ is the collection of all points (x_0, y_0) in the coordinate plane so that $ax_0 + by_0 \leq c$. Excluding the case that $b = 0$ (which can be handled easily), the graph of $ax + by = c$ is then a non-vertical line, and one can then *prove* that the graph of $ax + by \leq c$ is all the points in the plane on or below the line $ax + by = c$ if $b > 0$ and on or above the line $ax + by = c$ if $b < 0$. Either of the latter is called a **half plane**. The fact that the graphs of linear inequalities are half planes of straight lines is needed for the solution of problems related to linear programming.

c. Functions

The concept of a function is a major building block of mathematics as a whole. Like most useful skills and concepts in mathematics, functions arise from the need for solving problems, more specifically, from the need for a tool to describe natural phenomena. For example, to arrive at a complete description of the temperature of a cup of freshly brewed coffee for the first 10 minutes after it has been poured, there is no alternative except to use a function f defined on the segment [0,10] from 0 to 10 on the number line, where the unit 1 represents 1 minute, so that for each t satisfying $0 \leq t \leq 10$, $f(t)$ gives the temperature of the coffee t minutes later. Many similar examples can be given so that students get to see the *need* for functions.

Given a function f of one variable, the **graph** of f is the collection of all the points in the plane of the form $(x, f(x))$ whenever $f(x)$ makes sense. Linear functions (of one variable) are those of the form $f(x) = cx + k$, where c and k are constants, and x is any number. Clearly the graph of f is the same as the graph of the linear equation in two variables $cx - y = -k$. It follows from the earlier discussion on page 10 that the graph of a linear function is a line. The linear functions of the form $f(x) = cx$, so-called **linear functions without constant term**, occupy a special place in school mathematics. It is only through the use of such functions that students can understand the usual discussion of so-called **proportional reasoning**. To see the relevance, notice that one can express $f(x) = cx$ in an equivalent way, as follows: for any two nonzero numbers x_1 and x_2, one has the proportional relationship $\frac{f(x_1)}{x_1} = \frac{f(x_2)}{x_2}$ (and the common value is c). Whenever proportional reasoning is called for, it means only one thing: A certain function turns out to be a linear function without constant term. A common flaw in the presentation of proportional reasoning is that certain functions that come up in such problems are *assumed automatically to be linear functions without constant term*. Unfortunately, such an assumption is sometimes not warranted. For example, the reasoning that "if five people need 12.5 liters of water for a camping trip, then eight people need 20 liters" would be accepted as a sensible rule of thumb in everyday life. In a classroom, however, it should be pointed out that the argument implicitly depends, in mathematical terms, on the assumption that every person needs exactly the same amount of water for the camping trip. Such an assumption, though unreasonable in other contexts, is necessary for the translation of the given data ("five people need 12.5 liters of water") into a mathematically solvable problem. The need for simplifying assumptions of this type must be explicitly pointed out to students.

After teaching linear functions, **quadratic functions** are next. These are functions f of the form $f(x) = ax^2 + bx + c$, where a, b, c are constants and x is any number. The **zeros** of f, i.e., the numbers x' so that $f(x') = 0$, are exactly the solutions of the quadratic equation $ax^2 + bx + c = 0$, which students will know are given by the quadratic formula. For more detailed information about f, one should consider the case of a positive a, as the case of a negative a is similar. By completing the square, one can rewrite f as $f(x) = a(x+p)^2 + q$, where p and q are constants, which can be explicitly determined in terms of a, b and c. In this form, one sees that (because a square is never negative) $f(x) \geq q$, and $f(x) = q$ exactly when $x = -p$. Thus the minimum value of f is q, and this happens exactly when $x = -p$. Another simple argument using $f(x) = a(x+p)^2 + q$ also shows that the graph of f is congruent to the graph of ax^2, and that the axis of (reflection) symmetry of the graph of f is the vertical line passing through $(-p, 0)$. Thus one can obtain all the essential information about f by simply applying the technique of completing the square.

The new information about quadratic functions greatly enlarges for students the scope of word problems. It is now relatively simple to find out among rectangles with a fixed perimeter, which has the biggest area. Word problems of the following type also become accessible: Two workmen, painting at a constant rate, can paint a house together in six days. In how many days can each paint it alone if it takes one of them three days longer than the other to get it done?

Beyond quadratic functions, there are not too many things one can say about **polynomial functions** [functions f so that $f(x)$ is a fixed polynomial in x] in general without first acquiring more advanced tools. For **rational functions** (quotients of polynomial functions), a new phenomenon is the emergence of the concept of an **asymptote** of its graph.

The next major classes of functions to be considered in school algebra are the exponential functions and the associated logarithmic functions. To solve $a > 0$ but $a \neq 1$, students have to make sense of the **exponential function** $h(x) = a^x$ for all numbers x. If x is a positive integer such as 7, the meaning of $h(7) = a^7$ is clear: $h(7) = aaaaaaa$. If x is a rational number, the meaning of a^x will have to be carefully defined, first for $x = 0$, then for x a fraction, and then for x a negative fraction. In school mathematics, all these considerations fall under the headings of rational exponents and laws of exponents. What needs to be pointed out is that the consideration of exponential functions is the main justification for teaching these topics. Once students know the meaning of $h(x)$ for all rational numbers x, in the context of school mathematics, this knowledge is already sufficient, and h is then known for all numbers x, rational or not. From the basic properties of rational exponents, it can be concluded that $h(x + x') = h(x)h(x')$ for all numbers x, x'. It is more common to express the last property directly as $a^{x+x'} = a^x a^{x'}$.

Suppose now $a > 1$. (One has to consider separately the case of an a so that $0 < a < 1$, but the reasoning is similar.) These same basic properties of rational exponents show that $h(x) = a^x$ is an increasing positive function. One can also present heuristic arguments as to why $h(x)$ goes to infinity as x goes to positive infinity, and $h(x)$ goes to 0 as x goes to negative infinity. In particular, $h(x)$ can be any positive number. Because the exponential function h is increasing and takes all positive values, it has an **inverse function** $\log_a s$, called the **logarithm with base a**, which is defined for all *positive* values s. Precisely, $\log_a s$ is the number x so that $a^x = s$ (i.e., $h(x) = s$). The fact that $h(x + x') = h(x)h(x')$ for all numbers x, x' now becomes $\log_a ss' = \log_a s + \log_a s'$ for all positive numbers s, s'. The graph of $\log_a$ is obtained from the graph of a^x by reflecting across the line $y = x$, but the proof of this fact requires some geometry.

Historically, the logarithmic functions were discovered before the exponential functions. Because $\log_a ss' = \log_a s + \log_a s'$, the multiplication ss' becomes (under $\log_a$) a sum, and with the compilation of so-called log tables, the computation of products of numbers becomes much more manageable. This was one reason that made the logarithm important before the advent of computers. Nowadays, of course, the logarithm is important for quite different reasons: $\log_a$ and the exponential function appear in nature in innumerable ways, and there is no way to avoid these two functions.

A final class of functions to be considered in school algebra is the set of **periodic functions** in general and the trigonometric functions—especially sine and cosine—in particular. A function f defined on the number line is periodic of period k for some positive number k if $f(x + k) = f(x)$ for all numbers x. The importance of such functions is clear once the periodic nature of many natural phenomena is realized.

d. Algebra of Polynomials

Thus far, a polynomial in a number *x* is a special kind of symbolic expression in *x*. One may therefore regard a polynomial as a function which assigns to each *x* the number given by that symbolic expression. At a more advanced level, a polynomial is a purely formal object and not a function. This is because algebra at an advanced level is an abstract study of structure and ceases being generalized arithmetic. School mathematics should introduce students to such abstract considerations at some point.

This introduction can include the following: Let *X* be a *symbol*, i.e., it no longer stands for a number. We will now prescribe how *X* should behave under addition and multiplication. Consider all sums of the form $f(X) = a_n X^n + a_{n-1} X^{n-1} + \cdots a_1 X + a_o$ where *n* is a whole number, the a_i's are constants (i = 0,1, ... , *n*), and $a_n \neq 0$. To avoid confusion with polynomials, such an *f*(*X*) is called a **polynomial form**, but *n* will continue to be referred to as its **degree** and the a_n, ... a_0 as its **coefficients**. Now define the addition and multiplication of polynomial forms by requiring that the sum or product of any two polynomial forms should be equal to the polynomial form normally obtained if *X* were an ordinary number. Thus, for polynomial forms, addition and multiplication are associative, commutative, and distributive.

Introducing *polynomial forms* rather than just polynomials offers additional clarity by separating considerations of the coefficients a_i from those of the symbol *X*. As far as the arithmetic operations on polynomial forms are concerned, the preceding definitions imply that addition and multiplication will continue to be associative, commutative, and distributive if one replaces the real number coefficients by, for example, **complex number** coefficients. By doing that, the two kinds of polynomial forms are denoted by **R**[*X*] and **C**[*X*], where **R** and **C** denote, for obvious reasons, the real and complex numbers, respectively. On the other hand, *X* is now free to assume an existence of its own, and can now take on values other than real or complex numbers. In more advanced courses, for instance, *X* can be a square matrix, and polynomials of a square matrix are an integral component of linear algebra.

R and **C** share an important property, which is that in each case, every nonzero number has a **multiplicative inverse**, i.e., if $a \neq 0$, there is a *b* so that $ab = ba = 1$. Therefore the important **division algorithm**, which is the exact analog among polynomial forms of division with remainder among whole numbers, in either **R**[*X*] or **C**[*X*], is valid. Consequently, the **factor theorem** for real or complex polynomial forms is valid, which can be explained as follows. First, we say a number *c* (whether real or complex) is a **root** of a polynomial form *f*(*X*) if the number *f*(*c*) obtained by replacing *X* with *c* in *f*(*X*) is equal to 0. Then the factor theorem states that given an *f*(*X*) in **R**[*X*] (respectively, **C**[*X*]), a number *c* in **R** (respectively, **C**) is a **root** of *f*(*X*) if and only if the linear polynomial form (*X*-*c*) in **R**[*X*] (respectively, **C**[*X*]) divides *f*(*X*).

Having emphasized the formal similarity between **R**[*X*] and **C**[*X*], students can now be pointed to an essential difference between the two as a consequence of the difference between real and complex numbers. For complex numbers, the **fundamental theorem of algebra** applies: Every complex polynomial form of positive degree has a complex root. The proof of this theorem requires some advanced ideas, but as previously referenced in connection with similar triangles, the importance of the theorem justifies that school students learn it and use it even if they will not see how it is proved.[11] By repeated application of the factor theorem, students will see that *every complex polynomial form of degree n, where n is a positive integer, is equal to the product of n linear complex polynomial forms*. If $f(X)$ is a real polynomial form, i.e., $f(X)$ is an element in $\mathbf{R}[X]$, it can be regarded as a complex polynomial form and therefore is also the product of n linear complex polynomial forms [n is the degree of $f(X)$]. However, since $f(X)$ is a *real* polynomial form, one may wish to have a conclusion involving only real polynomial forms rather than complex ones. With this in mind, one applies the fundamental theorem of algebra with more care and concludes that *every polynomial form with real coefficients is the product of* real *linear polynomial forms and* real *quadratic polynomial forms without real roots.* The reasoning is very instructive.

So far, polynomial forms with one symbol X has been the focus, but there is no reason not to consider polynomial forms in more than one symbol. A case in point is the very natural question of whether there is a formula for $(X+Y)^n$, where X and Y are two symbols and n is a positive integer. The main impetus behind this question is the simple identity $(X+Y)^2 = X^2 + 2XY + Y^2$, which answers the question for $n = 2$. Additional effort by the students reveals that $(X+Y)^3 = X^3 + 3X^2Y + 3XY^2 + Y^3$. If one is persistent and computes the 4th, 5th, and even 6th powers, one would perceive a certain pattern and come up with a guess that the expansion of $(X+Y)^n$ must involve the so-called **binomial coefficients**. The precise result is the **binomial theorem**, and a common proof of this theorem uses the technique of **mathematical induction**. The latter is an integral part of school algebra.

e. Combinatorics and Finite Probability

In the process of proving the binomial theorem, one comes into contact with the basic properties of the binomial coefficients, the Pascal Triangle, and, consequently, simple facts about finite probability. Indeed, finite probability is fertile ground for applications of ideas in algebra as well as a rich source of problems.

B. Algebra Topics in Curriculum Sources

As mentioned in the introduction to Section V, the Task Group's judgment in formulating the Major Topics of School Algebra was informed by careful examinations of relevant source material. Now that the list itself has been presented and amplified by a discussion of important linkages among topics, the discussion returns to the various categories of source material. The following sections will show how the Task Group's Major Topics of School Algebra align with the actual content of current standards, teaching materials, and assessment tools.

[11] Students who take a college course in complex analysis will see then how it is proved.

1. State Standards for Algebra I and Algebra II

Because the responsibility for public education rests mainly with the states (and with local communities), requirements for the number of years of study in high school mathematics vary by state and across local communities within a state. Most states require 3 years or units of mathematics, a few require 2, and a small but growing number require 4 (Newton, Larnell, & Lappan, 2006). In addition, each state (except for Iowa, which has a draft of its secondary standards out for review) has its own mathematics standards or guidelines for Grades preschool through 8 or higher, as well as its own state tests. In the name of local control, local school districts also have their own curriculum standards or expectations, identifying what they believe to be the essential elements of mathematics content and instruction.

To determine common elements in algebra education across the states, the Task Group analyzed the content of state-based curriculum frameworks with specific attention to their standards, objectives, or course level expectations (CLEs). As of June 2006, the 22 states in Figure 2 provided standards for Algebra I and II courses. (A few other states provided only integrated curriculum standards at the high school level, while most states at that time did not provide any standards in high school mathematics.) However, in some of these states, (e.g., North Carolina) algebra standards are also offered as part of an integrated approach in the high school mathematics curriculum and may differ in coverage and level of difficulty from the algebra standards in their single-subject courses. An integrated approach may be generally defined as one in which the topics of high school mathematics are presented in some order other than the customary sequence in the Unites States of yearlong courses in Algebra I, Geometry, Algebra II, and Precalculus. In some states, algebra standards are offered only as part of an integrated approach rather than for single-subject courses.

Figure 2: States With Standards for Algebra I and II Courses

Alabama	Michigan
Arkansas	Mississippi
California	North Carolina
District of Columbia (counted as a state in NAEP)	Oklahoma
Florida	Oregon
Georgia	South Carolina
Hawaii	Tennessee
Indiana	Texas
Kentucky	Utah
Maryland	Virginia
Massachusetts	West Virginia

Source: The algebra curriculum of each state is available to the public on each state's department of education Web site. This figure was prepared for the Task Group by Institute for Defense Analyses Science and Technology Policy Institute in June 2006.

The frameworks for the Algebra I and Algebra II standards in these 22 states contain 300 different course level expectations (CLEs), which were organized into 31 major topics (Institute for Defense Analyses Science and Technology Policy Institute, in press, a). After tallying the frequency with which the CLEs occurred under each topic on a state-by-state basis, the Task Group found 13 broad topics included by at least 15 of these 22 states (see Table 1). Results for Algebra I and II were combined for this analysis.

Table 1: Frequency Counts for Broad Topics in 22 States' Standards for Algebra I and II Courses

Topic	**Number of States (*n* = 22)**
Linear equations and slope	21
Systems of equations	20
Evaluating, interpreting and representing data	19
Analyzing, interpreting and representing functions and relations	19
Inequalities	19
Real number operations	19
Solving quadratic problems	18
Exponents, roots, radicals, and absolute values	17
Operations with polynomials	16
Exponential functions and equations	16
Probability	16
Complex number operations	15
Rational equations and functions	15

Source: Institute for Defense Analyses Science and Technology Policy Institute, Brief Number 2, in press, a.

Note: Twenty-two states are shown, which represents the available information at the time.

Table 2 compares the Major Topics of School Algebra with the algebra topics in 20 Algebra I and Algebra II mathematics frameworks and in 3 high school integrated mathematics frameworks. These 23 sets of algebra topics come from 21 states that had content expectations of Algebra I or II, or an integrated math curriculum in Algebra I or II, or both. The Algebra I or II topics are in the 20 states from which the Task Group was able to obtain frameworks explicitly for Algebra. The algebra topics in the 3 sets of integrated mathematics frameworks come from 3 randomly selected states with integrated mathematics frameworks: Florida, North Carolina, and Georgia (Georgia no longer has standards for Algebra I and Algebra II courses). The comparisons do not reflect depth of treatment in the frameworks or the classroom. Nor do they necessarily reflect actual classroom content.

Table 2: Major Topics of School Algebra Covered by State Algebra or Integrated Mathematics Frameworks, by State and Two-Thirds Composite*

Major Topics of School Algebra	CA	GA (int)	AL	IN	MA	AR	FL	FL (int)	DC	MI	TN	MS	NC	NC (int)	VA	WV	UT	SC	HI	OK	OR	TX	MD	Two-thirds
Symbols and Expressions																								
Polynomial expressions																								
Rational expressions																								
Arithmetic and geometric sequences and series																								
Linear Relations																								
Real numbers as points on the number line																								
Linear equations and their graphs																								
Solving problems with linear equations																								
Linear inequalities and their graphs																								
Graphing and solving systems of simultaneous linear equations																								
Quadratic Relations																								
Factors and factoring of quadratic polynomials with integer coefficients																								
Completing the square in quadratic expressions																								
Quadratic formula and factoring of general quadratic polynomials																								
Using the quadratic formula to solve equations																								
Functions																								
Linear functions																								
Quadratic functions – word problems involving quadratic functions																								
Graphs of quadratic functions and completing the square																								
Polynomial functions (including graphs of basic functions)																								
Simple nonlinear functions (e.g., square and cube root functions; absolute value; rational functions; step functions)																								
Rational exponents, radical expressions, and exponential functions																								
Logarithmic functions																								
Trigonometric functions																								
Fitting simple mathematical models to data																								
Algebra of Polynomials																								
Roots and factorization of polynomial forms																								
Complex numbers and operations																								
Fundamental theorem of algebra																								
Binomial coefficients (and Pascal's Triangle)																								
Mathematical induction and the binomial theorem																								
Combinatorics and finite probability																								

Composite Key							
	Explicit		Not Found		Implicit/Incomplete		2/3 of States

*Integrated mathematics frameworks are identified as (int).

Source: Table created for the Task Group by Institute for Defense Analyses Science and Technology Policy Institute in June 2007.

Note: In this analysis, two-thirds of states included algebra as a single subject course or as integrated in the states' algebra framework.

2. Algebra I and Algebra II Textbooks

The Task Group examined an older algebra textbook as well as a number of current textbooks for Algebra I and II. Four of the five sets of Algebra I and II textbooks were national editions published by several major textbook publishers; a fifth set was the California edition published by Holt, Rinehart and Winston. The goal was to determine if there were any substantive differences between the content of a California edition and national editions. In addition, the Task Group also wanted to find out how an edition of the Algebra I and II textbooks authored by Mary Dolciani, dominant in the American market for many years, compared with current textbooks as well as with the Major Topics of School Algebra.

The Task Group's comparison of the major algebra topics in these five sets of Algebra I and Algebra II textbooks with the Major Topics of School Algebra appears in Appendix A. The Task Group's analysis of a textbook used in 1913 (see Figure 3) shows that the major topics, as derived from the table of contents, have remained much the same since the early 1900s except for the addition of material to modern textbooks addressing trigonometry, probability, and statistics (Institute for Defense Analyses Science and Technology Policy Institute, in press, a).

Figure 3: Topics in a 1913 High School Algebra Textbook

1. Definitions of Elementary Terms	14. Factors and Multiples
2. The Equation	15. Fractions
3. Addition	16. Ratio, Proportion, and Variation
4. Subtraction	17. Graphs of Linear Equations
5. Factoring	18. Systems of Linear Equations
6. Multiplication	19. Involution and Evolution
7. Division	20. Radicals and Exponents
8. Equations	21. Quadratic Equations
9. Type Products*	22. Systems of Quadratic Equations
10. Review and Extension of Processes	23. Graphs of Quadratic Equations
11. Exponents and Roots	24. Proportion, Variation and Limits
12. Logarithms	25. Series
13. Imaginary and Complex Numbers	26. Geometric Problems for Algebraic Solutions

*This chapter treats the identities $(a \pm b)^2 = a^2 \pm 2ab + b^2$, $(a+b)(a-b) = a^2 - b^2$ and related third-order identities, which are organized into "types."

Source: Young, 1913.

3. Singapore's Mathematics Curriculum for Grades 7–10

Singapore's fourth- and eighth-graders have consistently outperformed all other countries' students on the mathematics portion of the TIMSS (Gonzales et al., 2004). Singapore's compulsory secondary curriculum begins in Grade 7 and extends through the 10th year of schooling. Figure 4 shows the algebra standards addressed in Grades 7 through 10 (Singapore Ministry of Education, 2006). Table 3 shows the comparison of the list of Major Topics of School Algebra with the topics in Singapore's secondary curriculum.

Figure 4: Singapore's 2007 Algebra Standards for Grades 7–10

Algebraic Representation and Formulae
Using letters to represent numbers
Interpreting notations such as $\frac{3 \pm y}{5}$ *as* $(3 \pm y) \div 5$ *or* $\frac{1}{5 \times (3 \pm y)}$
Evaluation of algebraic expressions and formulae
Translation of simple "real-world" situations into algebraic expressions
Recognizing and representing number patterns (including finding an algebraic expression for the n^{th} term)
Algebraic Manipulation
Addition and subtraction of linear algebraic expressions
Simplification of linear algebraic expressions, e.g., $-2(3x - 5) + 4x$
Factorization of linear algebraic expressions of the form: $ax + ay$ (where a is a constant) and $ax + bx + kay + kby$ (where a, b and k are constants)
Expansion of the product of algebraic expressions
Changing the subject of a formula
Finding the value of an unknown quantity in a given formula
Recognizing and applying the special products
Factorization of algebraic expressions
Multiplication and division of simple algebraic fractions
Addition and subtraction of algebraic fractions with linear or quadratic denominator
Functions and Graphs
Cartesian coordinates in two dimensions
Graph of a set of ordered pairs
Linear relationships between two variables (linear functions)
Gradient of a linear graph as the ratio of the vertical change to the horizontal change (positive and negative gradients)
Graphs of linear equations in two unknowns
Graphs of quadratic functions and their properties: positive or negative coefficient of x^2, maximum and minimum points and symmetry
Sketching of the graphs of quadratic functions given in the form $y = \pm (x - p)^2 + q$ *and* $y = \pm (x - a)(x - b)$
Graphs of functions of the form y = ax^n where $n = -2, -1, 0, 1, 2, 3,$ and simple sums of no more than three of these
Graphs of exponential functions $y = kax$ where a is a positive integer
Estimation of gradients of curves by drawing tangents
Solutions of Equations and Inequalities
Solving linear equations of one unknown (including fractional coefficients)
Solving simple inequalities (e.g., $3x \leq 5$)
Solving simple fractional equations that can be reduced to linear equations
Formulating a linear equation in one unknown to solve problems
Solving simultaneous linear equations in two unknowns by substitution and elimination methods, and graphical method
Solving quadratic equations in one unknown by factorization
Formulating a pair of linear equations in two unknowns or a quadratic equation in one unknown to solve problems
Solving quadratic equations in one unknown by: use of formula, completing the square for $y = x\ 2 + px + q$, or graphical methods
Solving fractional equations that can be reduced to quadratic equations
Solving linear inequalities in one unknown and representing the solution set on the number line

Continued on p. 3-21

Figure 4: Singapore's 2007 Algebra Standards for Grades 7–10, continued

Matrices
Display of information in the form of a matrix of any order
Interpreting the data in a given matrix
Product of a scalar quantity and a matrix
Problems involving the calculation of the sum and product (where appropriate) of two matrices
Ratio, Rate, and Proportion
Map scales (distance and area)
Direct and inverse proportion
Set Language and Notation
Use of set language
Union and intersection of two sets
Venn diagrams
Numbers and the Four Operations
Positive, negative, zero, and fractional indices
Laws of indices
Quadratic Equations and Inequalities
Conditions for a quadratic equation to have: two real roots, two equal roots, and no real roots
Related conditions for a given line to: intersect a given curve, be a tangent to a given curve, and not intersect a given curve
Solution of quadratic inequalities, and the representation of the solution set on the number line
Conditions for $r\ ax^2 + bx + c$ to be always positive (or always negative)
Relationships between the roots and coefficients of the quadratic equation $ax^2 + bx + c = 0$
Indices and Surds
Four operations on indices and surds
Rationalizing the denominator
Solving equations involving indices and surds
Polynomials
Multiplication and division of polynomials
Use of remainder and factor theorems
Factorization of polynomials
Solving cubic equations
Simultaneous Equations in Two Unknowns
Solving simultaneous equations with at least one linear equation, by substitution
Expressing a pair of linear equations in matrix form and solving the equations by inverse matrix method
Partial Fractions
Partial fractions including cases where the denominator is no more complicated than: $(ax + b)(cx + d)$, $(ax + b)(cx + d)^2$, and $(ax + b)(x^2 + c^2)$
Binomial Expansions
Use of the binomial theorem for positive integer n
Use of the notations $n!$ and $({}^{n}{}_{r})$
Use of the general term $({}^{n}{}_{r})\ a^{n-r}b^{r}$, $0 < r \leq n$
Exponential, Logarithmic, and Modulus Functions
Functions a^x, e^x, $\log_a x$, in x, and their graphs
Laws of logarithms
Equivalence of $y = a^x$ and $x = \log_a y$
Change of base of logarithms
Function x and graph of $f(x)$, where $f(x)$ is linear, quadratic, or trigonometric
Solving simple equations involving exponential, logarithmic, and modulus functions

Source: Singapore Ministry of Education, 2006.

Table 3: Comparison of Major Topics of School Algebra With Singapore's Secondary Mathematics Curriculum

Major Topics of School Algebra	**Singapore's Secondary Curriculum**
Symbols and Expressions	
Polynomial expressions	
Rational expressions	
Arithmetic and finite geometric series	
Linear Equations	
Real numbers as points on the number line	
Linear equations and their graphs	
Solving problems with linear equations	
Linear inequalities and their graphs	
Graphing and solving systems of simultaneous linear equations	
Quadratic Equations	
Factors and factoring of quadratic polynomials with integer coefficients	
Completing the square in quadratic expressions	
Quadratic formula and factoring of general quadratic polynomials	
Using the quadratic formula to solve equations	
Functions	
Linear functions	
Quadratic functions and word problems involving quadratic functions	
Graphs of quadratic functions and completing the square	
Polynomial functions, including graphs of basic functions	
Simple nonlinear functions (e.g., square and cube root functions, absolute value, rational functions, step functions)	
Rational exponents, radical expressions, and exponential functions	
Logarithmic functions	
Trigonometric functions	
Fitting simple mathematical models to data	
Algebra of Polynomials	
Roots and factorization of polynomials	
Complex numbers and operations	
Fundamental theorem of algebra	
Binomial coefficients (and Pascal's Triangle)	
Mathematical induction and the binomial theorem	
Combinatorics and Finite Probability	
Combinations and permutations as applications of the binomial theorem and Pascal's Triangle	

Note: A shaded cell shows agreement among the Major Topics of School Algebra and Singapore's secondary mathematics curriculum.

Source: The data in this table are STPI tabulations using data available from the Singapore Ministry of Education, 2007.

C. Algebra Topics in Assessment Sources

1. National Assessment of Educational Progress Test Objectives

NAEP developed two sets of mathematics objectives of interest to this Task Group, one for Grade 8 and one for Grade 12. The 2005 *NAEP Mathematics Framework for the 2005 National Assessment of Educational Progress* report describes a proposed special study at Grade 8 to "examine the breadth and depth of Grade 8 students' understanding of proportionality and other fundamental topics in algebra" (National Assessment Governing Board, 2004). This document includes a list of objectives for Algebra and a list of objectives for proportionality that are to be used for this special assessment. However, because this assessment has not yet been scheduled, the Task Group did not examine its objectives. Figure 5 shows the algebra objectives addressed in NAEP's Grade 12 assessment based on its 2005 assessment framework. Table 4 shows the comparison of the algebra topics in this set of objectives with the Major Topics of School Algebra.

Figure 5: Algebra Objectives for National Assessment of Educational Progress' Grade 12 Mathematics Assessment

Patterns, relations, and functions
Recognize, describe, or extend arithmetic, geometric progressions, or patterns using words or symbols
Express the function in general terms (either recursively or explicitly), given a table, verbal description, or some terms of a sequence
Identify or analyze distinguishing properties of linear, quadratic, inverse ($y = k/x$), or exponential functions from tables, graphs, or equations
Determine the domain and range of functions given various contexts
Recognize and analyze the general forms of linear, quadratic, inverse, or exponential functions (e.g., in $y = ax + b$, recognize the roles of a and b)
Express linear and exponential functions in recursive and explicit form given a table or verbal description
Algebraic representations
Translate between different representations of algebraic expressions using symbols, graphs, tables, diagrams, or written descriptions
Analyze or interpret relationships expressed in symbols, graphs, tables, diagrams, or written descriptions
Graph or interpret points that are represented by one or more ordered pairs of numbers on a rectangular coordinate system
Perform or interpret transformations on the graphs of linear and quadratic functions
Use algebraic properties to develop a valid mathematical argument
Use an algebraic model of a situation to make inferences or predictions
Given a "real-world" situation, determine if a linear, quadratic, inverse, or exponential function fits the situation (e.g., half-life bacterial growth)
Solve problems involving exponential growth and decay
Variables, expressions, and operations
Write algebraic expressions, equations, or inequalities to represent a situation
Perform basic operations, using appropriate tools, on algebraic expressions (including grouping and order of multiple operations involving basic operations, exponents, roots, simplifying, and expanding)
Write equivalent forms of algebraic expressions, equations, or inequalities to represent and explain mathematical relationships
Equations and inequalities
Solve linear, rational, or quadratic equations or inequalities
Analyze situations or solve problems using linear or quadratic equations, or inequalities symbolically or graphically
Recognize the relationship between the solution of a system of linear equations and its graph
Solve problems involving more advanced formulas [e.g., the volumes and surface areas of three-dimensional solids; or such formulas as: $A = P(1+r)^t$, $A = Pe^{rt}$]
Given a familiar formula, solve for one of the variables
Solve or interpret systems of equations or inequalities

Source: The data in this figure are Institute for Defense Analyses Science and Technology Policy Institute tabulations based on information from National Assessment Governing Board (2004).

Table 4: Comparison of the Major Topics of School Algebra With the 2005 NAEP Grade 12 Algebra Topics

Major Topics of School Algebra	NAEP Grade 12 Algebra Topics
Symbols and Expressions	
Polynomial expressions	
Rational expressions	
Arithmetic and finite geometric series	
Linear Equations	
Real numbers as points on the number line	
Linear equations and their graphs	
Solving problems with linear equations	
Linear inequalities and their graphs	
Graphing and solving systems of simultaneous linear equations	
Quadratic Equations	
Factors and factoring of quadratic polynomials with integer coefficients	(see note)
Completing the square in quadratic expressions	
Quadratic formula and factoring of general quadratic polynomials	
Using the quadratic formula to solve equations	
Functions	
Linear functions	
Quadratic functions and word problems involving quadratic functions	
Graphs of quadratic functions and completing the square	
Polynomial functions, including graphs of basic functions	
Simple nonlinear functions (e.g., square and cube root functions, absolute value, rational functions, step functions)	
Rational exponents, radical expressions, and exponential functions	
Logarithmic functions	
Trigonometric functions	
Fitting simple mathematical models to data	
Algebra of Polynomials	
Roots and factorization of polynomials	
Complex numbers and operations	
Fundamental theorem of algebra	
Binomial coefficients (and Pascal's Triangle)	
Mathematical induction and the binomial theorem	
Combinatorics and Finite Probability	
Combinations and permutations as applications of the binomial theorem and Pascal's Triangle	

Note: A shaded cell shows agreement among the Major Topics of School Algebra and 2005 NAEP.

"Factors and factoring of quadratic polynomials with integer coefficients" is subsumed under one of the Grade 12 objectives in the 2005 NAEP assessment framework. "Completing the square in quadratic expressions" is not explicitly a part of the 2005 framework at Grade 12. "Quadratic formula and factoring of general quadratic polynomials" appears partially in the 2005 Grade 12 framework. "Using the quadratic formula to solve equations" is not explicitly included as an objective at Grade 12. However, since there is an objective at Grade 12 on solving quadratic equations, students would have to utilize the quadratic formula or complete the square to solve a quadratic equation (P. Carr, personal communication, May 25, 2007).

Source: The data in this table are Institute for Defense Analyses Science and Technology Policy Institute tabulations using data available from National Assessment Governing Board, 2004.

2. American Diploma Project Benchmarks and Test Objectives

The American Diploma Project (ADP) Benchmarks describe the mathematics content and skills that ADP suggests all students should master by the time they leave high school if they are to be successful in college and work. Achieve Inc. developed these benchmarks based on research in colleges, universities, and high-performance workplaces across the country (Achieve Inc., 2007). The ADP benchmarks include five strands. The algebra strand is subdivided into six clusters, most of which include a number of benchmarks (up to eight). The six cluster headings are as follows:

1) Perform basic operations on algebraic expressions fluently and accurately.

2) Understand functions, their representations, and their properties.

3) Apply basic algebraic operations to solve equations and inequalities.

4) Graph a variety of equations and inequalities in two variables, demonstrate understanding of the relationships between the algebraic properties of an equation and the geometric properties of its graph, and interpret a graph.

5) Solve problems by converting the verbal information given into an appropriate mathematical model involving equations or systems of equations, apply appropriate mathematical techniques to analyze these mathematical models, and interpret the solution obtained in written form using appropriate units of measurement.

6) Understand the binomial theorem and its connections to combinatorics, Pascal's Triangle, and probability.

Currently, 30 states are working with Achieve Inc. to align their standards with the needs of college and work, as represented by the ADP Benchmarks. Thirteen of these states are also collaborating on an end-of-course Algebra II test that is based on the ADP Benchmarks and can serve as an indicator of readiness for credit-bearing college mathematics courses. Figure 6 shows the core algebra topics and the topics in the optional modules for this test.

Figure 6: Topics to Be Assessed in the American Diploma Project Algebra II End-of-Course Test

Topics in Core Test Modules	Topics in Optional Test Modules
Operations on Numbers and Expressions	**Data and Statistics**
Real numbers	Summarization and comparison of data sets
Complex numbers	Interpretation and communications through data
Algebraic expressions	**Probability**
Equations and Inequalities	Permutations, combinations, and probability
Linear equations and inequalities	Probability distributions
Nonlinear equations and inequalities	**Logarithmic Functions**
Polynomial and Rational Functions	Logarithmic expressions and equations
Quadratic functions	Logarithmic functions
Higher-order polynomial and rational functions	**Trigonometric Functions**
Exponential Functions	Trigonometric functions
Exponential functions	**Matrices**
Function Operations and Inverses	Matrix arithmetic
Function operations and composition	Solving systems of equations using matrices
Inverse functions	Matrix transformations
Piecewise functions	Vectors
	Conic Sections
	Conic sections
	Sequences and Series
	Arithmetic and geometric sequences and series
	Other types of iteration and recursion

Source: Achieve Inc., 2007.

Table 5 shows how ADP's high school benchmarks, the core topics in its Algebra II end-of-course test, and the topics in the optional modules for this Algebra II test compare with the Major Topics of School Algebra. There are three topic areas that require additional explanation. While the Major Topics of School Algebra lists these topics as algebra, the ADP addresses them in their Benchmarks as categories outside of algebra. Specifically, these areas include the following: 1) under the category of Linear Equations, real numbers as points on the number line are categorized by ADP as part of the Number Sense and Numerical Operations strand, and not in algebra as categorized by the Major Topics of School Algebra, 2) under the category of Functions, trigonometric functions are considered by ADP to be in the Geometry strand rather than algebra, and 3) under the category of Algebra of Polynomials, complex numbers are categorized as Number Sense and Numerical Operations (while operations are not in the Number Sense and Numerical Operations category) and not as algebra. Therefore, the lack of shading to represent agreement in these cases on what is considered algebra simply means that the ADP addresses them at other places in their Benchmarks.

Also of note, under the category of Algebra of Polynomials, the matching for those topics only refers to that fact the binomial theorem is discussed in ADP, but no proof by mathematical induction is required.

Table 5: Comparison of Major Topics of School Algebra With American Diploma Project's High School Algebra Benchmarks, Core Topics in Its Algebra II Test, and the Topics in the Optional Modules for Its Algebra II Test

<table>
<tr><th>Major Topics of School Algebra</th><th>ADP Algebra Benchmarks</th><th>ADP Algebra II Core</th><th>ADP Algebra II Optional</th></tr>
<tr><td colspan="4">Symbols and Expressions</td></tr>
<tr><td>Polynomial expressions</td><td bgcolor="#cccccc"></td><td bgcolor="#cccccc"></td><td></td></tr>
<tr><td>Rational expressions</td><td bgcolor="#cccccc"></td><td bgcolor="#cccccc"></td><td></td></tr>
<tr><td>Arithmetic and finite geometric series</td><td bgcolor="#cccccc"></td><td></td><td bgcolor="#cccccc"></td></tr>
<tr><td colspan="4">Linear Equations</td></tr>
<tr><td>Real numbers as points on the number line</td><td></td><td bgcolor="#cccccc"></td><td></td></tr>
<tr><td>Linear equations and their graphs</td><td bgcolor="#cccccc"></td><td bgcolor="#cccccc"></td><td></td></tr>
<tr><td>Solving problems with linear equations</td><td bgcolor="#cccccc"></td><td bgcolor="#cccccc"></td><td></td></tr>
<tr><td>Linear inequalities and their graphs</td><td bgcolor="#cccccc"></td><td bgcolor="#cccccc"></td><td></td></tr>
<tr><td>Graphing and solving systems of simultaneous linear equations</td><td bgcolor="#cccccc"></td><td bgcolor="#cccccc"></td><td></td></tr>
<tr><td colspan="4">Quadratic Equations</td></tr>
<tr><td>Factors and factoring of quadratic polynomials with integer coefficients</td><td bgcolor="#cccccc"></td><td bgcolor="#cccccc"></td><td></td></tr>
<tr><td>Completing the square in quadratic expressions</td><td bgcolor="#cccccc"></td><td bgcolor="#cccccc"></td><td></td></tr>
<tr><td>Quadratic formula and factoring of general quadratic polynomials</td><td bgcolor="#cccccc"></td><td bgcolor="#cccccc"></td><td></td></tr>
<tr><td>Using the quadratic formula to solve equations</td><td bgcolor="#cccccc"></td><td bgcolor="#cccccc"></td><td></td></tr>
<tr><td colspan="4">Functions</td></tr>
<tr><td>Linear functions</td><td></td><td bgcolor="#cccccc"></td><td></td></tr>
<tr><td>Quadratic functions, word problems involving quadratic functions</td><td bgcolor="#cccccc"></td><td bgcolor="#cccccc"></td><td></td></tr>
<tr><td>Graphs of quadratic functions and completing the square</td><td bgcolor="#cccccc"></td><td bgcolor="#cccccc"></td><td bgcolor="#cccccc"></td></tr>
<tr><td>Polynomial functions, including graphs of basic functions</td><td></td><td bgcolor="#cccccc"></td><td></td></tr>
<tr><td>Simple nonlinear functions (e.g., square and cube root functions, absolute value, rational functions, step functions)</td><td bgcolor="#cccccc"></td><td bgcolor="#cccccc"></td><td></td></tr>
<tr><td>Rational exponents, radical expressions, and exponential functions</td><td bgcolor="#cccccc"></td><td bgcolor="#cccccc"></td><td></td></tr>
<tr><td>Logarithmic functions</td><td bgcolor="#cccccc"></td><td></td><td bgcolor="#cccccc"></td></tr>
<tr><td>Trigonometric functions</td><td></td><td bgcolor="#cccccc"></td><td bgcolor="#cccccc"></td></tr>
<tr><td>Fitting simple mathematical models to data</td><td bgcolor="#cccccc"></td><td></td><td bgcolor="#cccccc"></td></tr>
<tr><td colspan="4">Algebra of Polynomials</td></tr>
<tr><td>Roots and factorization of polynomials</td><td bgcolor="#cccccc"></td><td bgcolor="#cccccc"></td><td></td></tr>
<tr><td>Complex numbers and operations</td><td></td><td bgcolor="#cccccc"></td><td></td></tr>
<tr><td>Fundamental theorem of algebra</td><td></td><td></td><td></td></tr>
<tr><td>Binomial coefficients and Pascal's Triangle</td><td bgcolor="#cccccc"></td><td></td><td bgcolor="#cccccc"></td></tr>
<tr><td>Mathematical induction and the binomial theorem</td><td bgcolor="#cccccc"></td><td></td><td></td></tr>
<tr><td colspan="4">Combinatorics and Finite Probability</td></tr>
<tr><td>Combinations and permutations as applications of the binomial theorem and Pascal's Triangle</td><td bgcolor="#cccccc"></td><td></td><td bgcolor="#cccccc"></td></tr>
</table>

Note: A shaded cell shows agreement among the Major Topics of School Algebra and the three ADP categories.

Source: Table created for the Task Group by Institute for Defense Analyses Science and Technology Policy Institute from information available from Achieve Inc., 2007.

D. Comparisons

To show how the 27 Major Topics of School Algebra first listed on page 5 compare with current practices, they were matched against algebra topics listed in 1) U.S. state standards for Algebra I and Algebra II courses, 2) current algebra textbooks, 3) Singapore's 2007 algebra standards for Grades 7 through 10, 4) NAEP's assessment objectives for its 2005 Grade 12 test, and 5) the ADP benchmarks for a high school exit test, its core Algebra II end-of-course test and its optional modules for this test. In Tables 2, 3, 4, and 5, the Major Topics of School Algebra served as row headings and the comparison sources served as column headings. The corresponding cell was shaded or filled in when a comparison source clearly included that specific Major Topic of School Algebra. It is important to note that a shaded cell simply means coverage, not extent of coverage.

Potential sources of error in this analysis are the different ways in which the 27 topics may be worded in each document. Some topics do not appear to be covered in a comparison source, but they may be covered under another topic in the comparison source. For example, although none of the comparison sources explicitly covers polynomial functions, some sources include these functions under such headings as rational equations and functions or operations with polynomials. The level of detail possible for this analysis did not allow for reconciliation of misalignments of this type.

As Tables 2 and 3, and Appendix A show, the three comparison sources providing topics for algebra course work (state algebra standards, algebra textbooks, and Singapore's secondary mathematics curriculum) include most, if not all, of the Major Topics in School Algebra. Overall, almost all the Major Topics of School Algebra can be found in the state standards for Algebra I and II; moreover, a majority of the topics appear in at least two-thirds of the available frameworks examined. In Singapore's secondary mathematics curriculum, only two topics do not appear to be covered, and they are the fundamental theorem of algebra, and combinatorics and finite probability (Table 3), although it is possible that these topics are covered after Grade 10. All the Major Topics of School Algebra appear in almost every set of Algebra I and II textbooks that the Task Group examined (Appendix A), whether national or state editions. In addition, all the Major Topics of School Algebra were addressed in the Dolciani-authored Algebra I and II textbooks (Dolciani, Swanson, & Graham, 1986; Dolciani, Sorgenfrey, Brown, & Kane, 1988). A striking and significant difference lies in the number of topics and page length of all current Algebra I textbooks, each of which has close to 1,000 pages and attempts to address far more topics than the more focused and much slimmer texts of 20 years ago. For example, the now out-of-print Dolciani algebra textbooks, which were among the most widely used textbooks of their day, had far fewer pages and focused on far fewer topics. It is not clear how many of the topics in current Algebra I and II textbooks students can realistically study in the course of one year, and, more importantly, to what depth they study the major algebra topics.

On the other hand, comparisons with sources that provide assessment standards or objectives show gaps. The NAEP algebra objectives for its current Grade 12 test do not include many of the Major Topics of School Algebra, such as real numbers as points on a number line, all the topics listed under the algebra of polynomials, and combinatorics and finite probability (Table 4). However, the National Assessment Governing Board has revised the Grade 12 objectives for the mathematics test to be administered in 2009. Several of the Major Topics of School Algebra not included on the 2005 NAEP test will be assessed on the 2009 test, including arithmetic and finite geometric series, logarithmic functions, trigonometric functions, binomial coefficients (and Pascal's Triangle), and mathematical induction and the binomial theorem (P. Carr, personal communication, May 24, 2007). It is important to remember that these assessment objectives were designed, as are all NAEP assessment objectives, expressly for the purpose of assessment (in this case, of high school mathematics), not for the development of curriculum frameworks.

A comparison with the ADP's core topics for its Algebra II end-of-course test (Table 5) also shows gaps. The ADP's list of core topics for its Algebra II end-of-course test does not explicitly include such subjects as arithmetic and finite geometric series or linear equations and their graphs, which are typically taught prior to Algebra II. But the list also omits some traditional Algebra II topics, such as logarithmic functions, the binomial theorem and Pascal's Triangle, and mathematical induction. The gaps are fewer when the topics in the optional modules for this test are included. In particular, the optional modules do cover arithmetic logarithmic functions, fitting simple mathematical models to data, the binomial theorem and Pascal's Triangle, mathematical induction, and combinatorics and finite probability. However, these topics will be assessed only if a state chooses to test the module in which they appear.

In sum, most of the Major Topics in School Algebra are addressed in state algebra standards for Algebra I and Algebra II, albeit inconsistently across the 21 states. They are all addressed in almost all the algebra textbooks that were examined. And they are addressed almost completely in Singapore's algebra standards for Grades 7 through 10. The Major Topics of School Algebra have the least amount of coverage in assessment objectives, for NAEP's current Grade 12 test and for ADP's forthcoming Algebra II end-of-course test of core topics.

E. Observations Regarding Rigor in Algebra Textbooks

The Task Group commissioned a systematic examination of leading Algebra I and Algebra II textbooks for mathematical accuracy. The results of the survey are described in Appendix B. They reveal a systemic problem: Textbook publishers, their authors, and editorial staff do not pay sufficient attention to mathematical accuracy. It should be emphasized that the Task Group is not asking for rigor in a formal mathematical sense. The mathematics should be presented in an age-appropriate fashion, yet be clear and accurate. Circular definitions or the omission of a definition of an important notion being introduced must be avoided, and can be avoided without making the material less accessible.

Many of the problems uncovered by the textbook examination will not be apparent to most students, or even to their teachers. However, such problems tend to affect students' learning in both overt and subtle ways. Mathematical reasoning, accuracy, and clarity of thought are learned by example. Mathematically flawed textbooks hinder this learning process.

V. What Are the Essential Mathematical Concepts and Skills That Lead to Success in Algebra and Should Be Learned As Preparation for Algebra?

The mathematics that children learn from preschool through the middle grades provide the basic foundation for Algebra and more advanced mathematics course work. What is taught at particular grade levels is determined at the local and state level, and reflects the interests of a variety of national, state, and local agencies and organizations, as well as parents and the general public. In the past, there has been no research base to guide them. However, the results of TIMSS and other international tests showing student achievement across the participating countries have led to international comparisons of curricula and provided much information on what high-achieving countries teach their students in elementary and middle school.

To suggest what essential concepts and skills should be learned as preparation for algebra course work, the Task Group reviewed the skills and concepts listed 1) in the Grades 1 through 8 curricula of the highest-performing countries on TIMSS, 2) in NCTM's *Curriculum Focal Points for Prekindergarten through Grade 8 Mathematics: A Quest for Coherence* (National Council of Teachers Mathematics, 2006), 3) in Grades K through 8 in the six highest-rated state curriculum frameworks in mathematics, 4) in a 2007 American College Testing (ACT) survey (American College Testing, 2007), and 5) in a Panel-sponsored survey of 743 teachers of introductory Algebra across the country who were asked what students need to learn to be prepared for success in Algebra (Hoffer, Venkataraman, Hedburg, & Shagel, 2007). The Task Group also took into consideration the structure of mathematics itself, which requires teaching a sequence of major topics. Based on these sources and considerations, the Task Group proposes three clusters of concepts and skills, defined later as the Critical Foundations on page 40, reflecting their judgment about the most essential mathematics for students to learn thoroughly prior to algebra course work. It should be noted that there is no direct empirical evidence to support the effectiveness of any lists discussed in this section for success in algebra course work.

A. International Approaches to Pre-Algebra Education

1. Mathematics Topics Taught in Grades 1 Through 8 in the TIMSS Top-Performing Countries

One of the richest sources of information on mathematics curricula in other countries is the work of William Schmidt and his colleagues, who used data drawn from TIMSS. Schmidt, Houang, and Cogan (2002) compared the mathematics topics and the grade levels at which they were taught in the six highest-performing countries (Singapore, Japan, Korea, Hong Kong, Flemish Belgium, and the Czech Republic), which they called the "A+ countries." Figure 7 shows the composite mathematics curriculum profile for Grades 1 through 8 from a paper that is slightly revised from 2002 (Schmidt et al., 2002; Schmidt & Houang, 2007).

Figure 7: Mathematics Topics Intended From Grade 1 to Grade 8 by a Majority of TIMSS 1995 Top-Performing Countries

Mathematics Topics Intended	Grade 1	2	3	4	5	6	7	8
1. Whole Number: Meaning	●	●	●	●	●			
2. Whole Number: Operations	●	●	●	●	●			
3. Measurement Units	●	●	●	●	●	●	●	
4. Common Fractions			●	●	●	●		
5. Equations & Formulas			●	●	●	●	●	●
6. Data Representation and Analysis			●	●	●	●		●
7. 2-D Geometry: Basics			●	●	●	●	●	●
8. 2-D Geometry: Polygons and Circles				●	●	●	●	●
9. Measurement: Perimeter, Area and Volume				●	●	●	●	●
10. Rounding and Significant Figures				●	●			
11. Estimating Computations				●	●	●		
12. Whole Numbers: Properties of Operations				●	●			
13. Estimating Quantity and Size				●	●			
14. Decimal Fractions				●	●	●		
15. Relation of Common and Decimal Fractions				●	●	●		
16. Properties of Common and Decimal Fractions					●	●		
17. Percentages					●	●		
18. Proportionality Concepts					●	●	●	●
19. Proportionality Problems					●	●	●	●
20. 2-D Geometry: Coordinate Geometry					●	●	●	●
21. Geometry: Transformations						●	●	●
22. Negative Numbers, Integers, and Their Properties						●	●	
23. Number Theory							●	●
24. Exponents, Roots and Radicals							●	●
25. Exponents and Orders of Magnitude							●	●
26. Measurement: Estimation and Errors							●	
27. Constructions Using Straightedge and Compass							●	●
28. 3-D Geometry							●	●
29. Geometry: Congruence and Similarity								●
30. Rational Numbers and Their Properties								●
31. Patterns, Relations and Functions								●
32. Proportionality: Slope and Trigonometry								●

Note:

● Individual topics intended by more than half of the top-achieving countries: Singapore, Japan, Korea, Hong Kong, Flemish Belgium, and the Czech Republic.

(grey shading) Collection of topics intended by more than half of top-achieving countries.

Source: Schmidt & Houang, 2007.

2. Differences in Curriculum Approaches Between TIMSS Top-Performing Countries and the United States

In 2003, the International Association for the Evaluation of Educational Achievement reported on its survey of educators in Singapore, Japan, Flemish Belgium, Chinese Taipei, and Korea to learn more about their mathematics curriculum in Grades 4 and 8 (Mullis et al., 2004). The mathematics curricula in these countries show different entry and exit points for many topics. That is, they introduce and complete the study of many topics at different grade levels. For example, patterns of numbers or shapes is taught between Grades 1 and 5 in Singapore, followed by missing numbers in an equation between Grades 2 and 5. In contrast, these concepts are generally not studied until Grade 4 in Japan or Chinese Taipei, and even later in Flemish Belgium.

There seem to be two major differences between the curricula in top-performing countries and U.S. curricula: in the number of mathematical concepts or topics presented at each grade level and in the expectations for learning. U.S. curricula typically include many topics at each grade level, with each receiving relatively light development, while top-performing countries present fewer topics at each grade level but in greater depth. In addition, U.S. curricula generally review and extend at successive grade levels many (if not most) topics already presented at earlier grade levels, while the TIMSS top-performing countries are more prone to expect proficiency in what is taught at each grade level. These critical differences distinguish a spiral curriculum (common in many subjects in U.S. curricula) from one built on proficiency—a curriculum that expects proficiency in the topics that are presented before more complex or difficult topics are introduced.

In addition, the mathematics curricula in these top-performing countries show study of topics that are not in many U.S. curricula for these grade levels. For example, simple linear equations and simultaneous linear equations are taught in Grades 7 through 8 in the Singapore curriculum, Grades 7 through 9 in Japan, and Grade 8 in Chinese Taipei. A majority of U.S. state curriculum frameworks do not present these algebra concepts in Grades 7 and 8, although other algebra topics are taught in some states before Grade 9 (Newton et al., 2006).[12]

It is important to note that these A+ countries, as well as many other countries, teach Algebra in Grade 8, if not earlier. For example, Singapore begins study of Algebra in Grade 7. Schmidt et al. (2002) note that the "A+ composite curriculum portrays an evolution from an early emphasis on arithmetic in Grades 1 through 4 to more advanced Algebra and Geometry beginning in Grades 7 and 8. Grades 5 and 6 serve as a transitioned stage in which such topics as proportionality and coordinate geometry are taught, providing a bridge to the study of Algebra and Geometry" (p. 5).

[12] According to Newton et al.'s analysis of state frameworks in 2006, 12 states specify two-step equations as a grade-level expectation in Grade 7, and 16 do so in Grade 8. It is not clear what the overlap is among states.

B. National Approaches to Pre-Algebra Education

1. National Council of Teachers of Mathematics Curriculum Focal Points

In September 2006, NCTM released *Curriculum Focal Points for Prekindergarten through Grade 8 Mathematics: A Quest for Coherence* (*Focal Points*), a document that drew on an analysis of national and international programs to respond to a need for coherence in U.S. mathematics curricula and to offer direction for teachers. The *Focal Points* are suggested grade-level areas of emphasis—the concepts, skills, and procedures that connect important mathematics topics from grade to grade, and form the foundation for more advanced mathematics, beginning with Algebra. NCTM's Web site[13] provides a description of the purposes for the *Focal Points*, their grade-level connections, and the mathematics defining them.

Figure 8 shows a comparison of the *Focal Points* with the composite curriculum of the A+ countries (Schmidt & Houang, 2007). After comparing NCTM's 1989 standards with this composite curriculum, Schmidt and Houang (2007) judge the *Focal Points* as representing a "movement toward more coherent standards." However, as Figure 8 shows, the *Focal Points* recommends study in the primary grades of much more than arithmetic topics (e.g., 2-D and 3-D geometry; transformations; and patterns, relations, and functions). As noted earlier, and as the composite curriculum in Figure 7 shows, the A+ countries concentrate on arithmetic topics in Grades 1 and 2.

[13] http://www.nctm.org/standards/focalpoints.aspx?id=282

Figure 8: Mathematics Topics Intended From Grade 1 to Grade 8 in the 2006 NCTM *Focal Points* Compared With the Topics Intended by a Majority of TIMSS 1995 Top-Achieving Countries

Mathematics Topics Intended	Grade 1	2	3	4	5	6	7	8
1. Whole Numbers: Meaning	●░	●░	●░	●░	●░	●		
2. Whole Numbers: Operations	●░	●░	●░	●░	●░			
3. Measurement: Units	●░	●░	●░	●░	●░	●░	●░	●
4. Common Fractions			●░	●░	●░	●░	●	
5. Equations & Formulas			░	●░	●░	●░	●░	●░
6. Data Representation & Analysis	●		●░	●░	●░	░	●░	●░
7. 2-D Geometry: Basics			●░	●░	░	░	░	●░
8. 2-D Geometry: Polygons & Circles	●	●	●	●░	●░	░	●░	●░
9. Perimeter, Area & Volume			●	●░	●░	●░	●░	●░
10. Rounding and Significant Figures				░	░			
11. Estimating Computations		●	●	●░	●░	░		
12. Properties of Operations	●	●	●	●░	●░			
13. Estimating Quantity and Size				●░	░			
14. Decimal Fractions				●░	●░	●░		
15. Relationships of Common & Decimal Fractions			●	●░	░	●░	●	
16. Properties of Common and Decimal Fractions					●░	●░		
17. Percentages					░	░	●	
18. Proportionality Concepts		●			░	●░	●░	●░
19. Proportionality Problems					░	●░	●░	●░
20. 2-D Coordinate Geometry					●░	░	░	●░
21. Transformations	●		●	●		░	░	░
22. Negative Numbers, Integers and Their Properties					●	░	●░	
23. Number Theory	●				●		●░	░
24. Exponents, Roots and Radicals							░	●░
25. Exponents and Orders of Magnitude							░	●░
26. Estimation and Errors		●	●		●		░	
27. Constructions w/ Straightedge and Compass							░	░
28. 3-D Geometry	●	●			●		●░	●░
29. Congruence and Similarity	●		●					●░
30. Rational Numbers and Their Properties							●	░
31. Patterns, Relations and Functions	●	●	●	●	●	●		●░
32. Slope and Trigonometry							●	●░

In the table, ● marks a dot and ░ marks a cell shaded grey.

Note:

Collection of topics intended by more than half of top-achieving countries

2006 NCTM Focal Points (*Curriculum Focal Points for Prekindergarten through Grade 8 Mathematics: A Quest for Coherence*)

Source: Schmidt & Houang, 2007.

2. Skills and Concepts in the Six Highest-Rated State Curriculum Frameworks

Also compared were the mathematics topics in the six highest-rated state curriculum frameworks, as judged by Klein et al. (2005), with the curriculum profile of the A+ countries (Schmidt & Houang, 2007). The Task Group used the highest-rated frameworks as noted in Klein et al. because this evaluation is the most comprehensive and in-depth review of state mathematics frameworks to date. The reviewers rated each state's standards for content, clarity, reason, and negative qualities, assigning different weights to each criterion for the overall assessment. The six highest-rated states are, in rank order, California, Indiana, Massachusetts, Alabama, New Mexico, and Georgia. A shaded cell in Table 6 indicates that the topic in the left-hand column appears in the mathematics standards of at least one of the six states at the given grade level. The 27 topics in Table 6 were among the 30 developed by the Task Group for examining state curriculum frameworks in mathematics and reflect the topics in the Schmidt et al. analysis (Institute for Defense Analyses Science and Technology Policy Institute, in press, b). They have been placed as closely as possible in the order in which the mathematics topics in the composite curriculum for the A+ countries first appear (Figure 7), an order that presumably reflects increasing difficulty or complexity.

Table 6: K Through 8 Grade-Level Expectations in the Six Highest-Rated State Curriculum Frameworks in Mathematics Compared With the Topics Intended by a Majority of TIMSS 1995 Top-Performing Countries*

		Grade								
Topics		**K**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	Whole Number Meaning	**6**	**5**	**5**	**4**	**6**	**2**			
2	Whole Number Operations	**3**	**4**	**5**	**6**	**3**				
3	Measuring and Units of Measurement	**6**	**6**	**2**	**3**	**4**	**2**	**6**		
4	Common Fractions	**3**	**1**	**4**	**4**	**5**	**3**			
5	Equations and Formulas				**2**	**2**	**2**	**4**	**6**	**6**
6	Collecting, Evaluating, Interpreting, and Representing Data	**4**	**5**	**5**	**3**	**4**	**5**	**5**	**3**	**4**
7	2-D Geometry Basics		**4**	**5**	**5**	**4**	**2**	**3**		
8	Polygons and Circles	**6**	**6**		**1**	**1**	**2**	**2**		**1**
9	Perimeter, Area, and Volume				**2**	**2**	**5**	**4**	**5**	**4**
10	Rounding				**4**	**5**	**4**			
11	Estimating Computations and Determine Reasonableness				**3**	**3**	**3**	**3**	**1**	
12	Properties of Whole Number Operations				**5**	**2**	**1**	**3**	**5**	**4**
13	Decimal Fractions and Decimals			**6**	**2**	**4**	**1**	**1**	**2**	**4**
14	Percentages						**5**	**4**	**2**	**1**
15	Ratios and Proportions							**2**	**1**	**2**
16	2-D Coordinate Geometry				**3**	**4**	**2**	**4**	**2**	
17	Geometric Transformation				**1**	**3**	**1**	**1**	**3**	**3**
18	Integers & Their Properties						**4**	**4**		
19	Number Theory						**5**	**3**	**1**	**1**
20	Exponents, Roots, Radicals, and Absolute Values							**3**	**5**	**6**
21	3-D Geometry				**6**	**2**	**2**	**1**	**3**	**1**
22	Congruence and Similarity		**1**	**2**	**2**	**5**	**2**	**2**	**4**	**3**
23	Rational Numbers and Their Properties							**2**	**6**	**5**
24	Patterns	**6**	**3**	**3**	**3**	**3**	**4**	**4**	**4**	**2**
25	Functions and Relations							**2**	**2**	**4**
26	Slope and Rates of Change							**1**	**2**	**3**
27	Probability		**1**	**3**	**3**	**3**	**3**	**5**	**1**	**4**

Note: A shaded cell indicates that at least one state specified an objective in this area at the given grade level.

STPI used 30 topics for its original analysis of these six state frameworks. Twenty-five of these topics agree with the TIMSS topics in Figure 7. Three of the remaining five topics were eliminated chiefly because they appeared to overlap with existing TIMSS topics. It is important to note that how STPI defined its 30 topics may differ from how these topics were defined by Schmidt et al. (2002) since it is not completely clear from the latter's writings how the 30 topics in Figure 7 were defined. A shaded cell indicates that at least one state specified an objective in this area at the given grade level.

The numbers in the shaded cells refer to the number of states within the six states that list that topic in their curriculum expectations.

Source: Institute for Defense Analyses Science and Technology Policy Institute, in press, b.

As Table 6 shows, there are wide variations across these six states in grade-level placement for these 27 topics. When the grade-level placement of the topics in these six highest-rated states is compared with the order of topics in Figure 7, there appears to be a relationship between the increasing complexity of the topics classified by TIMSS under numbers (Items 1, 2, 4, 10, 11, 12, 13, 14, 15, 18, 19, and 23) and under algebra (Items 5, 20, 25, and 26) and their introduction at increasingly higher grade levels. Nevertheless, Table 6 suggests that these six states (and probably most, if not all, of the others) spend a great deal of time in the primary grades on topics other than arithmetic (in the case of these six states, patterns, probability, and data analysis).

Patterns are labeled as an “algebra topic” in elementary and middle school mathematics curricula and assessments in this country; yet patterns are not emphasized in the curriculum of the high-performing countries in TIMSS; nor are patterns a topic of major importance in the Major Topics of School Algebra. The prominence given to patterns in K through 8 mathematics education in the U.S. is thus not supported by either the mathematical considerations or the data from TIMSS.

There is yet another striking difference with Figure 7. Not only do the A+ countries concentrate on arithmetic in the early grades, they also introduce geometry topics gradually from Grades 3 to 8, for the most part adding only one new geometry topic at each grade level. The seven geometry topics for the TIMSS high-performing countries in Figure 7 (Items 7, 8, 9, 16, 17, 21, and 22) first appear in Grades 3, 4, 4, 5, 6, 7, and 8, respectively. In contrast, in Table 6, in one or more of the highest-rated sets of state mathematics standards, geometry topics first appear in Grades 1, K/1, 3, 3, 3, 3, and 1 respectively. Indeed in these six states, there is little resemblance to the order of difficulty suggested in the composite curriculum of the A+ countries.

This comparison speaks to the reality of the spiral curriculum and the excessive number of topics taught at the elementary grades, even in the state frameworks that many consider as models for other states. Table 6 also helps make the point that a state’s mathematics standards, however highly their quality may be judged, do not necessarily correlate with student achievement in the state. These six states exhibit a wide range of student achievement on the 2007 NAEP mathematics tests for Grades 4 and 8. The quality of a state’s assessments and the extent to which its standards drive sound school curricula, as well as appropriate programs for teacher preparation and professional development, are intervening variables that strongly influence achievement. They may well override the quality of its standards.

C. Surveys of What College and Secondary Teachers See as Essential Concepts and Skills

1. Findings From the ACT Curriculum Survey

In 2007, ACT issued a report containing the results of its 2005–06 curriculum survey of a nationally representative sample of middle and high school teachers, high school counselors, and postsecondary regular and remedial course instructors in four major subjects (ACT, 2007). The survey indicates what instructors at postsecondary institutions believe is important and necessary for their entering students to know, and what middle and high school teachers are teaching. It therefore identifies the gap between postsecondary expectations and secondary school practice. For mathematics, ACT received responses from about 2,400 teachers, instructors, and counselors for an average response rate of 17%, with categorical response rates ranging from 11% (middle or junior high school teachers) to 26% (postsecondary instructors).

The responses of the instructors of postsecondary mathematics remedial courses were closer to the ratings of postsecondary mathematics entry-level-course instructors than to the ratings of high school mathematics teachers. On ratings of individual skills by mathematics strand on a 1 through 5 scale, postsecondary instructors for both regular and remedial mathematics courses rated the “4 skills that make up basic operations and applications” as most important (4.15) and the “12 skills that make up probability, statistics, and data analysis” as least important (1.76). On their rank ordering of eight mathematics strands in terms of importance, the postsecondary instructors rated basic operations and applications first and probability, statistics, and data analysis last. With respect to the course work needed for success in postsecondary mathematics, the content that both sets of postsecondary instructors rated the highest and that was being covered the least in instruction in arithmetic or in Algebra I courses was “solving quadratic equations and factoring (80%), working with rational exponents (41%), and using the quadratic formula (68%)” (ACT, 2007, p. 19).

2. Findings From the National Mathematics Advisory Panel Survey

As part of its deliberations in 2006, the Panel set as a priority a process to obtain information from a large national sample of teachers of introductory Algebra on, among other things, their views on their students’ mathematics preparation. The Panel developed a survey and gathered comments from 743 teachers (Hoffer, Venkataraman, Hedburg, & Shagle, 2007). About 28% of the Algebra I teachers were teaching at the middle or junior high school level, while almost all of the others were teaching in high schools.

The survey findings show that these algebra teachers generally describe their students’ backgrounds for Algebra I as weak. The two areas in which teachers report their students having the poorest preparation are 1) rational numbers and operations involving fractions and decimals and 2) solving word problems. The most frequent type of suggestion among the 578 teachers who responded in writing to an open-ended question was a greater focus in the elementary grades on proficiency with basic mathematics concepts and skills (Hoffer et al., 2007, p. 13). To a question on students using calculators in the early grades, those who wrote in a response specifically mentioned that they would like to see less use of calculators before students take their Algebra I class (p. 13). A number of teachers (n = 46) also mentioned student success in a “pre-Algebra” curriculum in the middle school as a requirement before they are allowed to take Algebra I (p. 14). In response to 10 options describing the challenges they face, a majority of the teachers (62%) saw “working with unmotivated students” as the “single most challenging aspect of teaching Algebra I successfully” (p. 23). The written-in responses, however, most frequently mentioned handling different skill levels in a single classroom (p. 23). In fact, a substantial number of teachers consider mixed-ability groupings to be a “moderate” (30%) or “serious” (23%) problem, an item with a combined rating of 53% for “moderate” and “serious” second only to the combined rating of 64% for “too little parent/family support” (p. 25). Finally, a majority of teachers favorably rated their Algebra I textbooks, with 90% agreeing or agreeing strongly that “the textbook includes the appropriate topics and content to teach the course” (Hoffer et al., 2007, Appendix D: Means and Confidence Intervals for Items in the National Survey of Algebra Teachers, p. 9-64).

D. Critical Foundations for Success in Algebra

The Task Group reviewed the concepts and skills indicated for teaching and learning in the elementary and middle grades 1) in the national curricula of the highest-performing countries on the TIMSS tests, 2) in the highest-rated U.S. state standards, 3) in NCTM’s *Focal Points*, 4) in the 2007 ACT curriculum survey, and 5) in the compiled ratings of the 743 algebra teachers surveyed for the Panel. The Task Group also took into consideration the structure of mathematics itself. Its structure requires teaching a sequence of major topics (from whole numbers to fractions, from positive numbers to negative numbers, and from the arithmetic of rational numbers to algebra) and an increasingly complex progression from explicit number computations to symbolic computations. The structural reasons for this sequence and its increasing complexity dictate what must be taught and learned before students take course work in Algebra. Based on all these considerations, the Task Group proposes the following three clusters of concepts and skills. The clusters reflect their judgment about what students need to learn thoroughly prior to algebra course work.

To prepare students for Algebra, the curriculum must simultaneously develop conceptual understanding, computational fluency, and problem-solving skills. These three aspects of learning are mutually reinforcing and should not be seen as competing for class time. The Critical Foundations identified and discussed below are not meant to comprise a complete preschool-to-algebra curriculum. However, the Task Group aims to recognize the Critical Foundations for the study of Algebra, whether as part of a dedicated algebra course in the seventh, eighth, or ninth grade, or within an integrated mathematics sequence in the middle and high school grades. These Critical Foundations deserve ample time in any mathematics curriculum. The foundations are presented in three distinct clusters of concepts and skills, each of which should incorporate the three aspects of learning noted here.

1. Fluency With Whole Numbers

By the end of the elementary grades, children should have a robust sense of number. This sense of number must include understanding place value, and the ability to compose and decompose whole numbers. It must clearly include a grasp of the meaning of the basic operations of addition, subtraction, multiplication, and division, including use of the commutative, associate, and distributive properties; the ability to perform these operations efficiently; and the knowledge of how to apply the operations to problem solving. Computational facility rests on the automatic recall of addition and related subtraction facts, and of multiplication and related division facts. It requires fluency with the standard algorithms for addition, subtraction, multiplication, and division. Fluent use of the algorithms not only depends on the automatic recall of number facts but also reinforces it. A strong sense of number also includes the ability to estimate the results of computations and thereby to estimate orders of magnitude, e.g., how many people fit into a stadium, or how many gallons of water are needed to fill a pool.

2. Fluency With Fractions

Before they begin algebra course work, middle school students should have a thorough understanding of positive as well as negative fractions. They should be able to locate both positive and negative fractions on the number line; represent and compare fractions, decimals, and related percents; and estimate their size. They need to know that sums, differences, products, and quotients (with nonzero denominators) of fractions are fractions, and they need to be able to carry out these operations confidently and efficiently. They should understand why and how (finite) decimal numbers are fractions and know the meaning of percentages. They should encounter fractions in problems in the many contexts in which they arise naturally, for example, to describe rates, proportionality, and probability. Beyond computational facility with specific numbers, the subject of fractions, when properly taught, introduces students to the use of symbolic notation and the concept of generality, both being an integral part of Algebra (Wu, 2001).

3. Particular Aspects of Geometry and Measurement

Middle-grade experience with similar triangles is most directly relevant for the study of algebra: Sound treatments of the slope of a straight line and of linear functions depend logically on the properties of similar triangles. Furthermore, students should be able to analyze the properties of two- and three-dimensional shapes using formulas to determine perimeter, area, volume, and surface area. They should also be able to find unknown lengths, angles, and areas.

E. Benchmarks for the Critical Foundations

In view of the sequential nature of mathematics, the Critical Foundations of Algebra described in the previous section require judicious placement in the grades leading up to Algebra. For this purpose, the Task Group suggests the following benchmarks as guideposts for state frameworks, for state assessments, and for school districts. There is no empirical research on the placement of these benchmarks, but they find justification in a comparison of national and international curricula. The benchmarks should be interpreted flexibly, to allow for the needs of students and teachers.

Fluency With Whole Numbers

1. By the end of Grade 3, students should be proficient with the addition and subtraction of whole numbers.

2. By the end of Grade 5, students should be proficient with multiplication and division of whole numbers.

Fluency With Fractions

1. By the end of Grade 4, students should be able to identify and represent fractions and decimals, and compare them on a number line or with other common representations of fractions and decimals.

2. By the end of Grade 5, students should be proficient with comparing fractions and decimals and common percents, and with the addition and subtraction of fractions and decimals.

3. By the end of Grade 6, students should be proficient with multiplication and division of fractions and decimals.

4. By the end of Grade 6, students should be proficient with all operations involving positive and negative integers.

5. By the end of Grade 7, students should be proficient with all operations involving positive and negative fractions.

6. By the end of Grade 7, students should be able to solve problems involving percent, ratio, and rate, and extend this work to proportionality.

Particular Aspects of Geometry and Measurement

1. By the end of Grade 5, students should be able to solve problems involving perimeter and area of triangles all quadrilaterals having at least one pair of parallel sides (i.e., trapezoids).

2. By the end of Grade 6, students should be able to analyze the properties of two-dimensional shapes and solve problems involving perimeter and area, and analyze the properties of three-dimensional shapes and solve problems involving surface area and volume.

3. By the end of Grade 7, students should be familiar with the relationship between similar triangles and the concept of the slope of a line.

VI. Does the Sequence of Mathematics Topics Prior to and During Algebra Course Work Affect Algebra Achievement?

As Schmidt et al. (2002) point out, the highest-performing countries in the TIMSS study employ somewhat differing curricula in the elementary grades. They conclude that a coherent, focused, and effective mathematics curriculum can be achieved in different ways. Over the past 15 years, many studies in the United States have examined the effects of recently developed mathematics curricula on student achievement at or across various grade levels. However, a search of the literature did not turn up any studies that sought to provide evidence on the effectiveness of these curricula (including their sequence of topics) for achievement in Algebra. A committee authorized by the National Academy of Sciences reviewed nearly 700 evaluations of 13 National Science Foundation-sponsored kindergarten through 12 mathematics curricula and 6 commercially developed mathematics curricula. Of these 700 evaluations, 147 met the committee's minimum criteria for scientific effectiveness and relevance, but they "did not permit one to determine the effectiveness of individual programs with a high degree of certainty, due to the restricted number of studies for any particular curriculum, limitations in the array of methods used, and the uneven quality of the studies" (Confrey & Stohl, 2004, p. 189).

There is no body of sound evidence showing particular multiyear mathematics curricula as more effective than others in preparing students in algebra course work. Thus, there is no basis in research for preferring a particular sequence.[14]

Beyond this central question are two related matters: a) whether an integrated approach or a single-subject sequence might be more effective for algebra course work and more advanced mathematics course work, and b) whether there are benefits to teaching the content of an Algebra I course before Grade 9. The next two subsections cover these issues.

A. Benefits of an Integrated or Single-Subject Approach For the Study of Algebra

The Task Group sought to examine if the differences in the way in which topics are sequenced in an integrated and single-subject approach for the study of Algebra lead to differences in algebra achievement. An integrated approach is defined as one in which the topics of high school mathematics are presented in some order other than the customary sequence in the United States of year-long courses in Algebra I, Geometry, Algebra II, and Precalculus.

[14] It should be noted that the What Works Clearinghouse (WWC), managed by the U.S. Department of Education, reviews studies evaluating the effectiveness of current mathematics (and other) programs as part of an ongoing process, using standards that it formulated to rate the quality of the studies it reviews. WWC has found five middle school mathematics curricula supported by what it rated as high-quality research (see http://www.whatworks.ed.gov/). It is not clear how much weight should be attached to the ratings for these five curricula because for most of these curricula there are, so far, only one to three studies in all contributing to the ratings, and some of the studies contributing to the ratings did not find statistically significant results. None of these studies sought to determine the effectiveness of a particular multiyear mathematics curriculum implemented prior to formal algebra course work for success in Algebra.

The curricula of most higher-performing nations in the TIMSS study do not follow the single-subject sequence of Algebra I, Geometry, and Algebra II, but they also differ from the approach used in most U.S. integrated curricula. Instead, Algebra, Geometry, and Trigonometry are divided into blocks. The teaching of each block typically extends over several months and aims for mathematical closure. As a result, the need to revisit essentially the same material over several years, often referred to as “spiraling,” is avoided. For an example, interested readers may consult the following Japanese textbooks for mathematics in Grades 9–11: The University of Chicago Mathematics Project, 1992; Kodaira, 1996, 1997.

The Task Group reviewed the most relevant studies to determine differences in achievement between curricular approaches that differentially shape the sequence of topics prior to or during the study of Algebra. They uncovered no research that clearly compared the use of textbooks featuring an integrated approach to the use of textbooks reflecting more of a single-subject approach. Nor did the researchers at the GE Foundation’s Urban Institute, who found 156 comparison studies of 18 middle and high school mathematics curricula that met their criteria for inclusion (Clewell et al., 2004). Their criteria included 1) an experimental or quasi-experimental design, 2) comparison groups, and 3) measures of student achievement that included but were not limited to test scores. Although the researchers found that students using six of the curricula scored higher than comparison students on both a majority of the standardized, state tests used, or both, and on a majority of the curriculum-based tests used, the reviewers stress that *only* 3 of the 156 studies they examined provided details on what was in the comparison curricula. Thus, it is not clear with what these six significantly more effective curricula were being compared in the other 153 studies, and there is no clear body of research from which conclusions on this question can be drawn.

There is no basis in research for preferring either a single-subject sequence or an integrated sequence for the teaching of school mathematics at the level of Algebra or above.

Although there seems to be little descriptive material on the differences in mathematical content between integrated and single-subject approaches for high school algebra, the differences may be striking. STPI compared North Carolina’s 2003 standards for an integrated approach to high school mathematics with the state’s standards for the single-subject sequence in high school (North Carolina State Board of Education, 2003; Institute for Defense Analyses Science and Technology Policy Institute, in press, c). It found that the integrated mathematics sequence for Grades 9 through 12 includes 1) all of the course objectives for the Algebra I course, 2) 7 out of 8 of the course objectives for the geometry course, and 3) 9 of the 15 course objectives for the Algebra II course. The comparison determined that North Carolina students completing *4 years* of the integrated mathematics sequence would not complete all the course objectives addressed by students in the *3-year* Algebra I, Geometry, and Algebra II sequence. In this case, at least, high school students enrolled in mathematics courses using textbooks featuring an integrated approach may not be in a position to take more advanced mathematics course work in their senior year, as can high school students at present who are able to enroll in an Algebra II course in their sophomore or junior year.

B. Research on the Timing of Algebra Course Work

The final question the Task Group addressed concerns the benefits of teaching the content of an Algebra I course before Grade 9, that is, of giving students an opportunity to learn more mathematics before Grade 9 than is expected in many, if not most, state curriculum frameworks. This specific question is relevant to the Task Group's broader question of the math necessary prior to algebra because the content of Algebra I is provided to some or all students in Grade 8 or even Grade 7 in many other countries (as well as to many students in this country), and the specific grade at which it is offered may well affect the grade-by-grade sequence of topics in a school's curriculum prior to algebra course work. The Task Group does not indicate specific grade levels for teaching or learning the essential concepts and skills in the three clusters proposed in the previous section, to allow schools the option of working out a coherent sequence of these concepts and skills at differing elementary grade levels, depending on when they choose to provide the content of an Algebra I course.

According to information gathered for the 2005 Grade 8 NAEP tests, about 39% of U.S. students have completed a 1- or 2-year Algebra I course in Grade 8 or have taken Algebra I in Grade 7.[15] Although clear and current international data across a wide range of countries on the timing of algebra course work cannot be located, it is clear from TIMSS data and the work of Schmidt et al. (2002) that students in the A+ countries study Algebra as well as Geometry in Grades 7 and 8. In contrast, in a large number of U.S. schools, an algebra course is not available in those grades. Schmidt et al. determined that "while 80% of eighth-graders had access to a 'regular' math course, only 66.5% of eighth-graders attend schools that even offer an algebra course. That is, a full third of eighth-graders don't even have such a course as an option" (p. 14).

Yet, a report from the U.S. Department of Education (1997) articulated the need to "provide all students the opportunity to take Algebra I or a similarly demanding course that includes fundamental algebraic concepts in the 8th grade and more advanced math and science courses in all four years of high school." It urged schools to "build the groundwork for success in algebra by providing a rigorous curriculum in grades K–7 that moves beyond arithmetic and prepares students for the transition to algebra."

The U.S. Department of Education report (1997) was the background for the Task Group's interest in finding research evidence on the long-term benefits of completing Algebra I before Grade 9. A search of the literature produced six studies that met the Panel's design criteria and included Algebra or mathematics achievement as an outcome (Jones, Davenport, Bryson, Bekhuis, & Zwick, 1986; Lee, Burkam, Chow-Hoy, Smerdon, & Geverdt, 1998; Ma, 2000, 2005; Smith, 1996; and Wilkins & Ma, 2002). Smith's study used algebra achievement as an outcome, but the others used general tests that measured student performance on a variety of mathematical concepts and skills. All of the studies were analyses of large national data sets [High School and Beyond (HS&B), Longitudinal Survey of American Youth (LSAY), National Education Longitudinal Study (NELS): 88, and the High School Effectiveness Study] and all examined the relationships between high school mathematics achievement and

[15] See http://nces.ed.gov/nationsreportcard/nde/viewresults.asp.

students' course-taking patterns in mathematics. Because students are never randomly assigned to specific course-taking patterns in mathematics in any school, one cannot definitively determine whether student achievement is the result of the courses that students take, whether their course-taking patterns result from their achievement, or if both their course-taking patterns and their achievement are the result of other, unmeasured factors. Nevertheless, these six studies are informative and appear to have used rigorous methods to analyze the course-taking patterns in mathematics that students chose. They all controlled for important school and student characteristics, including prior achievement, although they did not control for exactly the same variables, and some studies controlled for more variables than others.

Four of the six studies highlight the relationship between the timing of Algebra I and mathematics achievement (Ma, 2000, 2005; Smith, 1996; and Wilkins & Ma, 2002). Three of the four (two of which used the same LSAY data set) found that students who took Algebra prior to starting high school tended to have higher mathematics achievement in high school (Ma, 2000, 2005; Smith) than those who did not. Ma (2000) examined the effect of course taking in one year on achievement in the following year. Controlling for socioeconomic status (SES), gender, age, and prior mathematics achievement, he found that students who took Pre-Algebra or Algebra I in Grade 7 had higher average mathematics achievement in Grade 8 than those who did not take these courses in Grade 7; in addition, taking Pre-Algebra in Grade 7 had a greater effect on Grade 8 achievement than taking Algebra I in Grade 7. Those who took Pre-Algebra, Algebra I, Algebra I Honors, or Geometry in Grade 8 had higher average mathematics achievement in Grade 9 than those who did not take one of these courses; in addition, boys did better than girls. Of these four courses, taking Algebra I in Grade 8 had the largest impact on Grade 9 achievement, followed by Pre-Algebra, Algebra I Honors, then Geometry. However, the mathematics courses that students took in Grade 9 did not predict Grade 10 achievement.

Smith's study (1996) compared students who took Algebra prior to starting high school with those who took it at the beginning of high school (Grade 9). She, too, found that students who took Algebra early had higher mathematics achievement scores in Grade 12, even after controlling for background characteristics and mathematics achievement in Grade 10. Smith also found that students who took Algebra early, on average, took more advanced mathematics courses in high school.

The final two studies (Jones et al., 1986; Lee et al., 1998) also highlight the relationship between the number of mathematics courses taken in high school and students' mathematics achievement. Controlling for prior mathematics achievement, verbal ability, and SES, Jones et al. found that, on average, students who took a larger number of advanced mathematics courses in high school (Algebra I or higher) had higher mathematics achievement in Grade 12 than students who took fewer courses. Conversely, Lee et al. found that, after controlling for characteristics of students and schools, students who took more low-level courses (lower than Algebra I), on average, had low mathematics achievement scores in Grade 12 and reached lower levels of mathematics course work by Grade 12 than students who took more high-level courses. Together, findings from these three studies (Smith, 1996; Jones et al.; and Lee et al.) suggest that students who take more mathematics courses at the level of Algebra 1 or higher have, on average, higher mathematics achievement in high school than students who take fewer courses, controlling for background characteristics.

It is important to note that these six studies drew on four national data sets. Three analyzed LSAY data (Ma, 2000, 2005; Wilkins & Ma, 2002), two used HS&B data (Smith, 1996; Jones et al., 1986), and one used data from NELS: 88 and the High School Effectiveness Study (Lee et al., 1998). The consistency of their findings is striking. The studies by Ma and others provide some evidence that there are long-term benefits for Grade 7 or 8 students with the requisite mathematical background for algebra if they can take an authentic Algebra course in Grade 7 or 8: higher mathematics achievement in high school and the opportunity to take advanced mathematics course work in Grade 11 or 12.

If students have the opportunity to take Algebra I in Grade 7 or 8, they will be able to enroll in precalculus, calculus, or other advanced mathematics courses in Grade 11 or 12. The importance of being able to take a calculus course before graduation, or at the least a precalculus course (or a post-Algebra II course that includes trigonometry), is underscored by a 2003 survey of admission requirements for Massachusetts public and private institutions of higher education offering 4-year engineering programs (Massachusetts Department of Education, 2003).

In sum, there is no research demonstrating that a specific multigrade sequence of mathematics topics assures success in Algebra. Nor is there a body of research from which one may draw conclusions about the relative effectiveness of either an integrated or a single-subject approach to the study of Algebra and more advanced mathematics. However, research evidence, as well as the experience of other countries, supports the value of preparing a higher percentage of students than the U.S. does at present to complete an Algebra I course or its equivalent by Grade 7 or 8, and of providing such course work in Grade 7 or 8.

VII. Recommendations

This Task Group affirms that Algebra is the gateway to more advanced mathematics and to most postsecondary education. All schools and teachers of mathematics must concentrate on providing a solid mathematics education to all elementary and middle school students so that all of them can enroll and succeed in Algebra. Students need to be soundly prepared for Algebra and then well taught in Algebra, regardless of the grade level at which they study it. To improve the teaching of Algebra, the Task Group proposes the following eight recommendations:

1) The Task Group recommends that school algebra be consistently understood in terms of the Major Topics of School Algebra given in this report on page 5.

2) The Major Topics of School Algebra, accompanied by a thorough elucidation of the mathematical connections among these topics, should be the main focus of Algebra I and Algebra II standards in state curriculum frameworks, in Algebra I and Algebra II courses, in textbooks for these two levels of Algebra whether for integrated curricula or otherwise, and in end-of-course assessments of these two levels of Algebra. The Task Group also recommends use of the Major Topics of School Algebra in revisions of mathematics standards at the high school level in state

curriculum frameworks, in high school textbooks organized by an integrated approach, and in grade-level state assessments using an integrated approach at the high school level, by Grade 11 at the latest.

3) Proficiency with whole numbers, fractions, and particular aspects of geometry and measurement are the Critical Foundations of Algebra (p. 40). Emphasis on these essential concepts and skills must be provided at the elementary- and middle-grade levels. The coherence and sequential nature of mathematics dictate the foundational skills that are necessary for the learning of algebra. By the nature of algebra, the most important foundational skill is proficiency with fractions (including decimals, percent, and negative fractions). The teaching of fractions must be acknowledged as critically important and improved before an increase in student achievement in Algebra can be expected.

4) The Benchmarks proposed by the Task Group on page 42 should be used to guide classroom curricula, mathematics instruction, and state assessments. They should be interpreted flexibly, to allow for the needs of students and teachers.

5) International studies show that high-achieving nations teach for proficiency in a few topics, in comparison with the U.S. mile-wide-inch-deep curriculum. A coherent progression, with an emphasis on proficiency in key topics, should become the norm in elementary and middle school mathematics curricula. What should be avoided in mathematics is an approach that continually revisits topics year after year without closure.

6) All school districts should ensure that all prepared students have access to an authentic algebra course—and should prepare more students than at present to enroll in such a course by Grade 8. The word "authentic" is used here as a descriptor of a course that addresses algebra consistently with the Major Topics of School Algebra. Students must be prepared with the mathematical prerequisites for this course according to the Critical Foundations and the Benchmarks.

7) Publishers must ensure the mathematical accuracy of their materials. Those involved with developing mathematics textbooks and related instructional materials need to engage mathematicians, as well as mathematics educators, in writing, editing, and reviewing these materials.

8) Teacher education programs and licensure tests for early childhood teachers, including all special education teachers at this level, should fully address the topics on whole numbers, fractions, and the appropriate geometry and measurement topics in the Critical Foundations of Algebra, as well as the concepts and skills leading to them; for elementary teachers, including elementary-level special education teachers, all topics in the Critical Foundations of Algebra and those topics typically covered in an introductory Algebra course; and for middle school teachers, including middle school special education teachers, the Critical Foundations of Algebra and all of the Major Topics of School Algebra.

BIBLIOGRAPHY

Achieve, Inc. (2007). *Mathematics benchmarks*. *The American diploma project*. Retrieved June 1, 2007 from http://www.achieve.org/node/187.

American College Testing (2007). *ACT national curriculum survey 2005–2006.* Iowa City, IA: ACT.

Business Roundtable. (2006). *Tapping America's potential: The education for innovation initiative*. Washington, DC: Business Roundtable.

Clewell, B.C., de Cohen, C.C., Campbell, P.B., Perlman, L., Deterding, N., Manes, S., et al. (2004). *Review of evaluation studies of mathematics and science curricula and professional development models.* Washington, DC: Urban Institute, GE Foundation.

Confrey, J., & Stohl, V. (Eds.). (2004). *On evaluating curricular effectiveness: Judging the quality of K–12 mathematics evaluations*. Washington, DC: The National Academies Press.

Dolciani, M.P., Sorgenfrey, R.H., Brown, R.G., & Kane, R.B. (1988). *Algebra and trigonometry – Book 2*. Boston: Houghton-Mifflin.

Dolciani, M.P., Swanson, R.A., & Graham, J.A. (1986). *Algebra 1.* Boston: Houghton-Mifflin.

Executive Order No. 13398, 3 C.F.R. 216 (2006).

Gonzales, P., Guzmán, J.C., Partelow, L., Pahlke, E., Jocelyn, L., Kastberg, D., et al. (2004). *Highlights from the Trends in International Mathematics and Science Study (TIMSS) 2003.* (NCES 2005–005). Washington, DC: U.S. Government Printing Office. Retrieved June 1, 2007 from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2005005.

Hoffer, T.B., Venkataraman, L., Hedberg, E.C., & Shagle, S. (2007). *Final report of the national survey of algebra teachers for the National Math Panel*. Chicago: University of Chicago.

Institute for Defense Analyses Science and Technology Policy Institute. (in press, a). High school algebra education in the United States: 2006, information brief number 1. *Elementary and secondary mathematics expectations, testing and selected outcomes*.

Institute for Defense Analyses Science and Technology Policy Institute. (in press, b). K–8 mathematics education in the United States, information brief number 2. *Elementary and secondary mathematics expectations, testing and selected outcomes*.

Institute for Defense Analyses Science and Technology Policy Institute. (in press, c). The integrated mathematics curriculum: A case study from the state of North Carolina, information brief number 2. *Elementary and secondary mathematics expectations, testing and selected outcomes*.

Jones, L.V., Davenport, E.C., Bryson, A., Bekhuis, T., & Zwick, R. (1986). Mathematics and science test scores as related to courses taken in high school and other factors. *Journal of Educational Measurement, 23*(3), 197–208.

Klein, D., Braams, B.J., Parker, T., Quirk, W., Schmid, W., & Wilson, W.S. (2005). *The State of State Math Standards 2005*. Retrieved August 1, 2007 from http://www.edexcellence.net/foundation/publication/publication.cfm?id=338&pubsubid=1118#1118.

Kodaira, K. (Ed.). (1996). *Mathematics 1: Japanese Grade 10.* Providence, RI: American Mathematical Society.

Kodaira, K. (Ed.). (1997). *Mathematics 2: Japanese Grade 11.* Providence, RI: American Mathematical Society.

Lee, V.E., Burkam, D.T., Chow-Hoy, T., Smerdon, B.A., & Geverdt, D. (1998). *High school curriculum structure: Effects on coursetaking and achievement in mathematics for high school graduates.* Working Paper No. 98-09. Washington, DC: U.S. Government Printing Office.

Loveless, T. (2004). *The 2004 Brown Center report on American education: How well are American students learning?* (Vol. 1, No. 5). Washington, DC: Brookings Institution.

Ma, X. (2000). A longitudinal assessment of antecedent course work in mathematics and subsequent mathematical attainment. *Journal of Educational Research, 94*(1), 16–28.

Ma, X. (2005). Early acceleration of students in mathematics: Does it promote growth and stability of growth in achievement across mathematical areas? *Contemporary Educational Psychology, 30*(4), 439–460.

Massachusetts Department of Education. (2003). *Mathematics and science courses required or recommended for admission into engineering and engineering technology programs at Massachusetts institutions of higher education*. Retrieved on (August 31, 2007) from http://www.doe.mass.edu/boe/sac/tech/.

Mullis, I.V.S., Martin, M.O., Gonzalez, E.J., & Chrostowski, S.J. (2004). *TIMSS 2003 International Mathematics Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eight Grades.* Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

National Assessment Governing Board. (2004). *Mathematics framework for the 2005 National Assessment of Educational Progress* (Chapter 3). Retrieved June 1, 2007 from http://www.nagb.org/frameworks/m_framework_05/chap3.html.

National Council of Teachers of Mathematics. (2006). *Curriculum focal points for prekindergarten through Grade 8 mathematics: A quest for coherence*. Reston, VA: National Council of Teachers of Mathematics.

Newton, J., Larnell, G., & Lappan, G. (2006). Analysis of K–8 grade-level algebra learning expectations. In B. Reys (Ed.), *The intended mathematics curriculum as represented in state-level curriculum standards: Consensus or confusion*. Charlotte, NC: Information Age Publishing.

North Carolina State Board of Education. (2003). Mathematics: Standard course of study and grade level competencies. Retrieved October 1, 2007 from http://community.learnnc.org/dpi/math/archives/SCSMATH.pdf.

Schmidt, W., & Houang, R. (2007). Lack of focus in mathematics: Symptom or cause? In T. Loveless (Ed.), *Lessons learned: What international assessments tell us about math achievement* (pp. 65–84). Washington, DC: Brookings Institution Press.

Schmidt, W., Houang, R., & Cogan, L. (2002, Summer). A coherent curriculum: The case of mathematics. *American Educator*.

Singapore Ministry of Education. (2006). *Education at secondary schools*. Retrieved on June 1, 2007 from http://www.moe.gov.sg/esp/schadm/sec1/Edu_at_Sec_Schs.htm.

Singapore Ministry of Education (2007). *Secondary mathematics syllabuses.* Retrieved on June 1, 2007 from http://www.moe.gov.sg/cpdd/doc/2007%20Sec%20Math%20Syllabuses.pdf.

Smith, J.B. (1996). Does an extra year make any difference? The impact of early access to algebra on long-term gains in mathematics achievement. *Educational Evaluation and Policy Analysis*, *18*(2), 141–153.

The University of Chicago Mathematics Project. (1992). *Japanese Grade 9 mathematics, UCSMP textbook translation.* Chicago: The University of Chicago.

U.S. Department of Education, National Center for Education Statistics. (1990–2005). *The high school transcript study*. Washington, DC: Author.

U.S. Department of Education. (1997). Mathematics equals opportunity (white paper prepared for U.S. Secretary of Education Richard W. Riley). Retrieved on October 1, 2007 from www.ed.gov/pubs/math/index.html.

U.S. Department of Education, National Center for Education Statistics. (2004). Remedial course-taking. *The condition of education.* Retrieved August 21, 2004, from http://nces.ed.gov/programs/coe/2004/section5/indicator31.asp.

U.S. Department of Education, National Center for Educational Statistics. (2007). The nation’s report card: Mathematics 2007. Washington, DC: Author.

Wilkins, J.L., & Ma, X. (2002). Predicting student growth in mathematical content knowledge. *Journal of Educational Research, 95*(5), 288–298.

Wu, H. (2001). How to prepare students for algebra. *American Educator*, *25*, 2, 10–17. Retrieved October 1, 2007 from http://www.aft.org/pubs-reports/american_educator/summer2001/index.html.

Young, J.W.A. (1913). *A high school algebra*. New York: D. Appleton and Company.

APPENDIX A

Table A-1: Comparison of the Major Algebra Topics in Five Sets of Algebra I and Algebra II Textbooks With the List of Major Topics of School Algebra

Major Topics of School Algebra		Glencoe McGraw-Hill 2008	Glencoe McGraw-Hill: Calif. Edition 2005	Prentice Hall 2007	Holt, Rinehart and Winston 2007	Houghton Mifflin 1988
Symbols and Expressions	Polynomial Expressions	Variables and Expressions	Variables and Expressions	Using Variables	Variables and Expressions	Variables and Equations
		Writing Equations	Expressions and Formulas	Properties of Numbers	Simplifying Expressions	
		Solving Equations by Using Addition and Subtraction	Commutative and Associative Properties	Adding Rational Numbers	Order of Operations	
		Expressions and Formulas	The Distributive Property	Subtracting Rational Numbers	Adding and Subtracting Polynomials	
		Order of Operations	Adding and Subtracting Rational Numbers	Multiplying and Dividing Rational Numbers		
		The Distributive Property	Multiplying Rational Numbers	Exponents and Order of Operations		
		Commutative and Associative Properties	Dividing Rational Numbers	The Distributive Property		
		Multiplying Monomials	Operations With Polynomials	Adding and Subtracting Polynomials		
		Adding and Subtracting Polynomials				
		Identity and Equality Properties	Order of Operations			
	Rational Expressions	Dividing Monomials	Rational Expressions	Solving Rational Equations	Simplifying Algebraic Expressions	Algebraic Fractions
		Rational Expressions	Multiplying Rational Expressions	Algebraic Expressions	Simplifying Rational Expressions	Adding and Subtracting Fractions
		Multiplying Rational Expressions	Dividing Rational Expressions	Rational Expressions	Multiplying and Dividing Rational Expressions	Working With Rational Expressions
		Rational Expressions With Like Denominators	Dividing Polynomials	Adding and Subtracting Rational Expressions	Adding and Subtracting Rational Expressions	Fractional Equations
		Rational Expressions With Unlike Denominators	Rational Expressions With Like Denominators	Solving Rational Equations	Solving Rational Equations	Polynomial Long Division
		Rational Equations and Functions	Rational Expressions With Unlike Denominators	Dividing Polynomials	Multiplying and Dividing Rational Expressions	Mixed Expressions
		Dividing Rational Expressions	Mixed Expressions and Complex Fractions	Simplifying Rational Expressions	Adding and Subtracting Rational Expressions	Rational Algebraic Expressions
		Mixed Expressions and Complex Fractions	Solving Rational Equations	Multiplying and Dividing Rational Expressions	Adding and Subtracting Polynomials	
		Multiplying and Dividing Rational Expressions	Solving Rational Equations and Inequalities			
		Adding and Subtracting Rational Expressions	Multiplying and Dividing Rational Expressions			
		Dividing Polynomials	Adding and Subtracting Rational Expressions			
			Operations With Polynomials			

Continued on p. 3-54

Table A-1, continued

Major Topics of School Algebra		Glencoe McGraw-Hill 2008	Glencoe McGraw-Hill: Calif. Edition 2005	Prentice Hall 2007	Holt, Rinehart and Winston 2007	Houghton Mifflin 1988
Symbols and Expressions	Arithmetic and Geometric Sequences and Series	Arithmetic Sequences	Arithmetic Sequences	Arithmetic Sequences	Arithmetic Sequences	Arithmetic and Geometric Series
		Arithmetic Series	Geometric Sequences	Geometric Sequences	Geometric Sequences	Sequences
		Geometric Series	Arithmetic Series	Geometric Series	Introduction to Sequences	
		Infinite Geometric Series	Infinite Geometric Series	Arithmetic Series	Series and Summation Notation	
		Recursion and Special Sequences	Recursion and Special Sequences		Arithmetic Sequences and Series	
		Geometric Sequences	Geometric Series		Geometric Sequences and Series	
					Mathematical Induction and Infinite Geometric Series	
Linear Equations	Real Numbers as Points on the Number Line	Properties of Real Numbers	Rational Numbers on the Number Line	Exploring Real Numbers	Adding and Subtracting Real Numbers	Numbers on a line
			Properties of Real Numbers	Properties of Real Numbers	Multiplying and Dividing Real Numbers	Operating With Real Numbers
					Properties of Real Numbers	Rational Numbers
						Irrational Numbers
						Working With Real Numbers: Addition and Subtraction
						Dividing Real Numbers
	Linear Equations and Their Graphs	Solving Equations With the Variable on Each Side	Writing Equations in Slope-Intercept Form	Point-Slope Form and Writing Linear Equations	The Slope Formula	Linear Equations
		Geometry: Parallel and Perpendicular Lines	Writing Equations	Slope-Intercept Form	Slope-Intercept Form	Linear Equations and Their Graphs
		Graphing Equations in Slope-Intercept Form	Identity and Equality Properties	Parallel and Perpendicular Lines	Rate of Change and Slope	
		Solving for a Specific Variable	Relations	Linear Equations	Point-Slope Form	Solving Equations and Solving Problems
		Solving Equations by Using Multiplication and Division	Solving Equations by Using Addition and Subtraction	Solving Equations	Using Intercepts	Transforming Equations: Addition and Subtraction
		Solving Multistep Equations	Geometry: Parallel and Perpendicular Lines	Equations With Variables on Both Sides	Slopes of Parallel and Perpendicular Lines	Transforming Equations: Multiplication and Division
		Linear Equations	Graphing Linear Equations	Solving Two-Step Equations	Solving Linear Equations and Inequalities	Slope of a Line
		Writing Linear Equations	Slope-Intercept Form	Solving Multistep Equations	Solving for a Variable	Slope-Intercept Form of a Linear Equation
		Writing Point-Slope Form	Solving Equations	Equations and Problem Solving	Solving Equations by Adding or Subtracting	Determining an Equation of a Line
		Similar Triangles	Solving Equations by Using Multiplication and Division	Proportions and Similar Figures	Solving Equations by Multiplying or Dividing	
		Rate of Change and Slope	Solving Multistep Equations		Solving Two-Step and Multistep Equations	
		Writing Equations in Slope-Intercept Form	Solving Equations With the Variable on Each Side		Solving Equations With Variables on Both Sides	

Continued on p. 3-55

Table A-1, continued

Major Topics of School Algebra		Glencoe McGraw-Hill 2008	Glencoe McGraw-Hill: Calif. Edition 2005	Prentice Hall 2007	Holt, Rinehart and Winston 2007	Houghton Mifflin 1988
Linear Equations	Linear Equations and their Graphs	Slope	Solving Equations and Formulas		Square Roots and Real Numbers	
		Linear Graphs	Equations as Relations			
			Writing Equations in Point-Slope Form			
			Linear Equations			
			Similar Triangles			
			Slope			
			Writing Linear Equations			
	Solving Problems With Linear Equations	Weighted Averages	Percent of Change	Percent of Change	Applications of Proportions	Ratio and Proportion
		Weighted Averages	Weighted Averages	Rate of Change and Slope		Percent Problems
						Variation and Proportion
		Percent of Change				Mixture and Work Problems
	Linear Inequalities and Their Graphs	Solving Compound Inequalities	Solving Inequalities by Addition and Subtraction	Solving Inequalities Using Multiplication and Division	Solving Linear Inequalities	Inequalities in One Variable
		Graphing Inequalities in Two Variables	Solving Inequalities by Multiplication and Division	Inequalities and Their Graphs	Graphing and Writing Inequalities	Inequalities in Two Variables
		Solving Inequalities	Solving Multistep Inequalities	Solving Inequalities Using Addition and Subtraction	Solving Inequalities by Adding or Subtracting	Working With Inequalities
		Graphing Inequalities	Solving Compound Inequalities	Linear Inequalities	Solving Inequalities by Multiplying or Dividing	Working With Absolute Value
		Solving Equations	Graphing Inequalities in Two Variables	Solving Multistep Inequalities	Solving Two-Step and Multistep Inequalities	Solving Inequalities
		Solving Inequalities by Addition and Subtraction	Solving Inequalities	Compound Inequalities	Solving Inequalities With Variables on Both Sides	Graphing Linear Inequalities
		Solving Inequalities by Multiplication and Division	Graphing Inequalities	Two-Variable Inequalities	Solving Compound Inequalities	Solving Systems of Linear Inequalities.
		Solving Inequalities Involving Absolute Value	Graphing Systems of Inequalities	Solving Inequalities	Linear Inequalities in Two Variables	
		Solving Multistep Inequalities	Linear Programming	Absolute Value Equations and Inequalities	Solving Systems of Linear Inequalities	
		Solving Compound and Absolute Value Inequalities	Solving Systems of Inequalities by Graphing	Linear Programming	Linear Programming	
		Linear Programming		Systems of Inequalities		
		Solving Systems of Inequalities by Graphing				
		Graphing Systems of Inequalities				
	Graphing and Solving Systems of Simultaneous Linear Equations	Elimination Using Addition and Subtraction	Graphing Systems of Equations	Solving Systems by Graphing	Solving Systems by Graphing	Using Two Variables
		Solving Systems of Equations Algebraically	Substitution	Solving Systems Using Substitution	Solving Systems by Substitution	Solving Systems of Linear Equations
		Graphing Systems of Equations	Elimination Using Addition and Subtraction	Solving Systems Using Elimination	Solving Systems by Elimination	Linear Systems

Continued on p. 3-56

Table A-1, continued

Major Topics of School Algebra		Glencoe McGraw-Hill 2008	Glencoe McGraw-Hill: Calif. Edition 2005	Prentice Hall 2007	Holt, Rinehart and Winston 2007	Houghton Mifflin 1988
Linear Equations	Graphing and Solving Systems of Simultaneous Linear Equations	Solving Systems of Equations in Three Variables	Elimination Using Multiplication	Applications of Linear Systems	Solving Special Systems	Systems of Equations
		Applying Systems of Linear Equations	Solving Systems of Equations by Graphing	Systems of Linear Inequalities	Using Graphs and Tables to Solve Linear Systems	Systems of Linear Equations in Three Variables
		Elimination Using Multiplication	Solving Systems of Equations Algebraically	Graphing Systems of Equations	Using Algebraic Methods to Solve Linear Systems	
		Solving Systems of Equations by Graphing	Solving Systems of Equations in Three Variables	Solving Systems Algebraically	Solving Linear Systems in Three Variables	
				Systems With Three Variables	Linear Equations in Three Dimensions	
Quadratic Equations	Factors and Factoring of Quadratic Polynomials with Integer Coefficients	Substitution	Factors and Greatest Common Factors	Solving Quadratic Equations	Factors and Greatest Common Factors	Factoring Integers
		Monomials and Factoring	Factoring Using the Distributive Property	Factoring to Solve Quadratic Equations	Factoring by Greatest Common Factors	Monomial Factors of Polynomials
		Factoring Trinomials: x^2+bx+c	Factoring Trinomials: x^2+bx+c	Multiplying and Factoring	Factoring x^2+bx+c	Multiplying Binomials Mentally
		Factoring Trinomials: ax^2+bx+c	Factoring Trinomials: ax^2+bx+c	Multiplying Binomials	Factoring ax^2+bx+c	Differences of Squares
		Factoring Using the Distributive Property	Factoring Differences of Squares	Factoring Trinomials of the Type x^2+bx+c	Factoring Special Products	Squares of Binomials
		Solving Quadratic Equations by Factoring	Perfect Squares and Factoring	Factoring Trinomials of the Type ax^2+bx+c	Choosing a Factoring Method	Factoring Pattern for x^2+bx+c, c positive
		Factoring Differences of Squares	Solving Quadratic Equations by Graphing	Factoring by Grouping	Solving Quadratic Equations by Graphing and Factoring	Factoring Pattern for x^2+bx+c, c negative
		Perfect Squares and Factoring	Solving Quadratic Equations by Factoring		Solving Quadratic Equations by Factoring	Factoring Pattern for ax^2+bx+c
						Factoring by Grouping
					Solving Quadratic Equations by Using Square Roots	Using More Than One Method of Factoring
						Solving Equations by Factoring
						Solving Problems by Factoring
						Quadratic Equations With Perfect Squares
	Completing the Square in Quadratic Expressions	Completing the Square	Solving Quadratic Equations by Completing the Square	Completing the Square	Completing the Square	Completing the Square
		Solving Quadratic Equations by Completing the Square	Completing the Square	Factoring Quadratic Expressions		
				Completing the Square		

Continued on p. 3-57

Table A-1, continued

Major Topics of School Algebra		Glencoe McGraw-Hill 2008	Glencoe McGraw-Hill: Calif. Edition 2005	Prentice Hall 2007	Holt, Rinehart and Winston 2007	Houghton Mifflin 1988
Quadratic Equations	**Quadratic Formula and Factoring of General Quadratic Polynomials**	The Quadratic Formula and the Discriminant	The Quadratic Formula and the Discriminant	The Quadratic Formula	The Quadratic Formula and the Discriminant	Quotients and Factoring
				Using the Discriminant	The Quadratic Formula	Products and Factors
						The Quadratic Formula
						The Quadratic Formula and the Discriminant
	Using the Quadratic Formula to Solve Equations	Solving Quadratic Equations by Using the Quadratic Formula	Solving Quadratic Equations by Using the Quadratic Formula	Not Available	Not Available	Solving Quadratic Equations
			Solving Equations Using Quadratic Techniques			Quadratic Equations
						Roots of Quadratic Equations
						Using Quadratic Equations
Functions	**Linear Functions**	Representing Relations	Graphs and Functions	Patterns and Functions	Identifying Linear Functions	Functions and Relations
		Proportional and Nonproportional Relationships	Ratios and Proportions	Ratio and Proportion	Transforming Linear Functions	Functions
		Relations and Functions	Functions	Relations and Functions	Writing Function	Functions and Relations
		Functions and Graphs	Writing Equations from Patterns	Relating Graphs to Events	Introduction to Parent Functions	Functions Defined by Equations
		Linear Functions	Classes of Functions	Functions Rules, Tables, and Graphs	Graphing Relationships	Functions Defined by Tables and Graphs
		Operations on Functions	Relations and Functions	Writing a Function Rule	Relations and Functions	Direct Variation
		Representing Functions	Operations on Functions	Describing Number Patterns	Introduction to Function	
		Direct, Joint, and Inverse Variation	Slope and Direct Variation	Relations and Functions	Relations and Functions	Direct and Inverse Variation Involving Squares
				Families of Functions	Rates, Ratios, and Proportions	Joint and Combined Variation
				Mathematical Patterns	Graphing Linear Functions	
				Direct Variation	Writing Linear Functions	
				Applying Linear Functions	Function Notations	
				Standard Form	Direct Variation	
					Variation Functions	
					Operations With Functions	
					Multiple Representations of Functions	
					Transforming Quadratic Functions	
					Using Transformations to Graph Quadratic Functions	
					Transforming Linear Functions	
					Proportional Reasoning	

Continued on p. 3-58

Table A-1, continued

Major Topics of School Algebra		Glencoe McGraw-Hill 2008	Glencoe McGraw-Hill: Calif. Edition 2005	Prentice Hall 2007	Holt, Rinehart and Winston 2007	Houghton Mifflin 1988
Functions	**Quadratic Functions – Word Problems Involving Quadratic Functions**	Not Available	Not Available	Quadratic Functions	Properties of Quadratic Functions in Standard Form	Not Available
				Quadratic Functions	Characteristics of Quadratic Functions	
	Graphs of Quadratic Functions	Graphing Quadratic Functions	Circles	Properties of Parabolas	Graphing Functions	Quadratic Functions and Their Graphs
		Parabolas	Analyzing Graphs of Quadratic Functions	Transforming Parabolas	Parabolas	Conic Sections: Circles and Parabolas
		Analyzing Graphs of Quadratic Functions	Parabolas	Parabolas	Graphing Quadratic Functions	Linear and Quadratic Functions
		Circles	Graphing Quadratic Functions	Circles	Solving Quadratic Equations by Graphing	
		Solving Quadratic Equations by Graphing	Solving Quadratic Equations by Graphing	Exploring Quadratic Graphs	Circles	
		Graphing Quadratic Functions	Graphing Quadratic Functions		Identifying Quadratic Functions	
			Classes of Functions			
		Solving Quadratic Equations by Graphing				
	Polynomial Functions (including graphs of basic functions)	Analyzing Graphs of Polynomials Functions	Polynomial Functions	Polynomial Functions	Investigating Graphs of Polynomial Functions	Products of Polynomials
		Polynomial Functions	Analyzing Graphs of Polynomial Functions		Transforming Polynomial Functions	Polynomial Division
			Graphing Polynomial Functions			
	Simple Nonlinear Functions Simple Nonlinear Functions (e.g., square and cube root functions; absolute value; rational functions; step functions)		Square Roots and Real Numbers	Graphing Rational Functions	Transforming Functions	Applying Fractional Equations
		Inverse Variation		Finding and Estimating Square Roots	Rational Functions	Inverse Functions and Equations
		Special Functions	Graphing Rational Functions	Graphing Absolute Value Equations	Solving Absolute-Value Equations and Inequalities	The Reciprocal of a Number
		Square Root Functions and Inequalities	Solving Absolute Value Equations	Absolute Value Equations and Inequalities	Absolute-Value Functions	Inverse Variation
		nth Roots	Solving Compound and Absolute Value Inequalities	Absolute Value Functions and Graphs	Rational Functions	
		Solving Absolute Value Equations	Square Root Functions and Inequalities	Inverse Relations and Functions	Solving Rational Equations and Inequalities	
		Inverse Functions and Relations	Direct, Joint, and Inverse Variation	Inverse Variation	Radical Functions	
		Direct, Joint, and Inverse Variation	Inverse Functions and Relations	Rational Functions and Their Graphs	Solving Radical Equations and Inequalities	
		Solving Rational Equations and Inequalities	Classes of Functions	The Reciprocal Function Family	Inverse Variation	

Continued on p. 3-59

Table A-1, continued

Major Topics of School Algebra		Glencoe McGraw-Hill 2008	Glencoe McGraw-Hill: Calif. Edition 2005	Prentice Hall 2007	Holt, Rinehart and Winston 2007	Houghton Mifflin 1988
Functions	Simple Nonlinear Functions Simple Nonlinear Functions (e.g., square and cube root functions; absolute value; rational functions; step functions)	Graphing Rational Functions	Inverse Variation	Graphing Square Root and Other Radical Functions	Inverses of relations and Functions	
				Graphing Square Root Functions	Piecewise Functions	
					Functions and Their Inverses	
					Square-Root Functions	
					Solving Nonlinear Systems	
					Square Roots	
	Rational Exponents, Radical Expressions, and Exponential Functions	Exponential Growth and Decay	Exponential Functions	Zero and Negative Exponents	Integer Exponents	Radical Expressions
		Rational Exponents	Growth and Decay	Multiplication Properties of Exponents	Multiplication Properties of Exponents	Problems Involving Exponents
		Simplifying Radical Expressions	Simplifying Radical Expressions	More Multiplication Properties of Exponents	Division Properties of Exponents	Using the Laws of Exponents
		Operations With Radical Expressions	Operations With Radical Expressions	Division Properties of Exponents	Radical Expressions	Exponential Functions
		Growth and Decay	Radical Equations	Exponential Functions	Adding and Subtracting Radical Expressions	Negative Exponents
		Properties of Exponents	Radical Expressions	Exponential Growth and Decay	Multiplying and Dividing Radical Expressions	Roots and Radicals
		Exponential Functions	Exponential Growth and Decay	Simplifying Radicals	Powers and Exponents	Powers of Monomials
		Operations With Radical Expressions	Exponential Functions	Operations With Radical Expressions	Exponential Functions	Rational exponents
		Solving Radical Equations and Inequalities	Rational Exponents	Solving Radical Equations	Exponential Growth and Decay	Exponential Growth and Decay
		Classes of Functions	Radical Equations and Inequalities	Rational Exponents	Linear, Quadratic, and Exponential Models	Roots of Real Numbers
		Radical Equations	Properties of Exponents	Properties of Exponential Functions	Solving Radical Equations	
		Exponential Functions	Roots of Real Numbers	Multiplying and Dividing Radical Expressions	Exponential Functions, Growth, and Decay	
			nth Roots	Binomial Radical Expressions	Radical Expressions and Rational Exponents	
				Solving Square Root and Other Radical Equations	Properties of Exponents	
				Roots and Radical Expressions	Rational Exponents	
				Choosing a Linear, Quadratic, or Exponential Model		
				Function Operations		
	Logarithmic Functions	Logarithms and Logarithmic Functions	Logarithms and Logarithmic Functions	Logarithmic Functions as Inverses	Logarithmic Functions	Logarithmic Functions
		Properties of Logarithms	Properties of Logarithms	Properties of Logarithms	Properties of Logarithms	The Natural Logarithm Function

Continued on p. 3-60

Table A-1, continued

Major Topics of School Algebra		Glencoe McGraw-Hill 2008	Glencoe McGraw-Hill: Calif. Edition 2005	Prentice Hall 2007	Holt, Rinehart and Winston 2007	Houghton Mifflin 1988
Functions	**Logarithmic Functions**	Common Logarithms	Common Logarithms	Exponential and Logarithmic Equations	Exponential and Logarithmic Equations and Inequalities	
		Base *e* and Natural Logarithms	Base *e* and Natural Logarithms	Natural Logarithms	The Natural Base, *e*	
					Transforming Exponential and Logarithmic Functions	
	Trigonometric Functions	Graphing Trigonometric Functions	Trigonometric Ratios	Trigonometric Ratios	Trigonometric Ratios	Trigonometric Functions
		Trigonometric Functions of General Angles	Graphing Trigonometric Functions	Angles of Elevation and Depression	Graphs of Sine and Cosine	Triangle Trigonometry
		Circular Functions	Translations of Trigonometric Graphs	Solving Trigonometric Equations Using Inverses	Graphs of Other Trigonometric Functions	Circular Functions and Their Graphs
		Inverse Trigonometric Functions	Circular Functions	Right Triangle and Trigonometric Ratios	Solving Trigonometric Equations	
		Translations of Trigonometric Graphs	Inverse Trigonometric Functions	Radian Measure	Inverses of Trigonometric Functions	
		Solving Trigonometric Equations	Solving Trigonometric Equations	The Tangent Function		
		Verify Trigonometric Functions		The Sine Function		
				The Cosine Function		
				Translating Sine and Cosine Function		
				Exploring Periodic Data		
				Reciprocal Trigonometric Functions		
	Fitting Simple Mathematical Models to Data	Scatter Plots and Lines of Fit	Modeling “Real-World” Data: Using Scatter Plots	Scatter Plots	Scatter Plots and Trend Lines	
			Statistics: Displaying and Analyzing Data	Scatter Plots and Equations of Lines	Curve Fitting With Linear Models	
			Statistics: Analyzing Data by Using Tables and Graphs	Using Linear Models	Modeling “Real-World” Data	
			Statistics: Scatter Plots and Lines of Fit			
			Statistics: Using Scatter Plots			
	Roots and Factorization of Polynomial Forms	The Remainder and Factor Theorems	Factoring Polynomials	Theorems About Roots of Polynomial Equations	Factoring Polynomials	Factors of Polynomials
		Polynomials	The Remainder and Factor Theorems	Polynomials and Linear Factors	Polynomials	Theory of Polynomial Equations
		Adding and Subtracting Polynomials	Multiplying Monomials	Multiplying Special Cases	Adding and Subtracting Polynomials	Solving Polynomial Equation
		Multiplying a Polynomial by a Monomial	Dividing Monomials	Dividing Polynomials	Special Products of Binomials	Adding and Subtracting Polynomials

Continued on p. 3-61

Table A-1, continued

Major Topics of School Algebra		Glencoe McGraw-Hill 2008	Glencoe McGraw-Hill: Calif. Edition 2005	Prentice Hall 2007	Holt, Rinehart and Winston 2007	Houghton Mifflin 1988
Functions	Roots and Factorization of Polynomial Forms	Operations with Polynomials	Polynomials	Solving Polynomial Equations	Multiplying Polynomials	Multiplying Monomials
		Solving Polynomial Equations	Adding and Subtractions Polynomials	Factoring Special Cases	Dividing Polynomials	Multiplying a Polynomial by a Monomial
		Multiplying Polynomials	Multiplying a Polynomial by a Monomial	Adding and Subtracting Polynomials		Multiplying Two Polynomials
		Dividing Polynomials	Multiplying Polynomials			The Remainder and Factor Theorems
		Rational Zero Theorem	Monomials			
			Polynomials			
			Rational Zero Theorem			
			Dividing Polynomials			
	Complex Numbers and Operations	Complex Numbers	Complex Numbers	Complex Numbers	Operations With Complex Numbers	Polar Coordinates and Complex Numbers
					Complex Numbers and Roots	Real Numbers and Complex Numbers
	Fundamental Theorem of Algebra	Roots and Zeros	Roots and Zeros	The Fundamental Theorem of Algebra	Finding Real Roots of Polynomial Equations	Theory of Polynomial Equations
					Fundamental Theorem of Algebra	
	Binomial Coefficients (and Pascal's Triangle)	The Binomial Theorem	The Binomial Theorem	The Binomial Theorem		Binomial Expansion
	Mathematical Induction and the Binomial Theorem	The Binomial Theorem	The Binomial Theorem	The Binomial Theorem	Mathematical Induction and Infinite Geometric Series	Binomial Expansion
		Proof and Mathematical Induction	Exponential and Binomial Distribution		Binomial Distributions	
			Proof and Mathematical Induction			
Combinatorics and Finite Probability	Combinatorics and Finite Probability	Probability of Compound Events	Permutations and Combinations	Probability of Compound Events	Combinations and Permutations	Probability
		Probability Simulations	Probability of Compound Events	Counting Methods and Permutations	Compound Events	Permutations and Combinations
		Probability	Probability: Simple Probability and Odds	Combinations	Permutations and Combinations	Fundamental Counting Principles
		Multiplying Probabilities	Probability Simulations	Probability	Independent and Dependent Events	
		Adding Probabilities	Multiplying Probabilities			
		Counting Outcomes	The Counting Principle	Conditional Probability		
		The Counting Principle	Permutations and Combinations	Permutations and Combinations		
		Permutations and Combinations	Counting Outcomes	Probability of Multiple Events		
			Probability			
			Adding Probabilities			

Note: The Major Topics of School Algebra can be found on page 5. The chapter headings of each textbook reviewed are sorted into each Major Topic of School Algebra category as applicable. If a column only has one box under each Major Topic of School Algebra, it means that that particular textbook had only one chapter or section that was applicable to the specific Major Topic of School Algebra. If a column is empty, that book did not have a chapter or section that fit.

Source: Institute for Defense Analysis Science and Technology Policy.

APPENDIX B: Errors in Algebra Textbooks

The National Mathematics Advisory Panel commissioned a mathematician to look systematically for mathematical errors in

A. two widely used algebra textbooks, one Algebra I and one Algebra II, and
B. a chapter on linear equations in each of three other popular Algebra I textbooks.

A summary of the results is provided below.

(A) **Error density** of an Algebra I and Algebra II textbook is defined to be the following quotient expressed as a percent:

$$\frac{\text{the total number of errors}}{\text{the total number of pages in the book}}$$

It was found that for the review noted above:

- Algebra I book has error density 50.2%, and
- Algebra II book has error density 41%.

This means that, for the Algebra I book, there is on average at least one error every two pages. The Algebra II book is slightly better in this regard, with about four errors in every 10 pages on average.

The analysis also provides additional information regarding the errors found within the Algebra I and Algebra II books. There are three types:

Type I: lack of clarity, minor errors, or misprints.
Type III: a gap in a logical argument or an error on a conceptual level.
Type II: an error that falls between those two types of errors.

The following table summarizes the error densities of these errors in both books:

Table B-1: Error Densities of Errors in Algebra I and Algebra II Textbooks

Book	Type I	Type II	Type III
Algebra I	20.4%	19.5%	10.3%
Algebra II	12.1%	19.4%	9.6%

An example of a Type I error is the statement that all lines with the same slope are parallel; the correct statement should be: *two distinct lines with the same slope are parallel*. Two examples of a Type II error are:

> pointing out that the method of solving a radical equation leads to an **extraneous solution** but without explaining exactly how or why, and stating that two functions are **inverse functions** of each other (e.g., exp and log) without giving their precise domains of definition.

Several examples of Type III errors are provided here; these are more serious errors:

- Graphing a function with a discrete domain of definition (e.g., the price of *n* articles) as a (continuous) straight line;
- Interpreting an event with a probability of 0 as an impossible event, and an event with a probability 1 as one that will definitely occur without specifying that this holds only for a finite sample space;
- Giving the first few terms of a pattern and extending it to the *n*-th terms as if the extension is unique;
- Using technical terms (e.g., **linear regression**) in a problem without giving their definitions;
- Conflating the *definition* of the **negative powers** and **rational powers** of a number with a *theorem*;
- Defining the **slope** of a line using two points on the line without pointing out the independence of the choice of the two points used, and later on;
- Pointing out such an independence without indicating that there is an explanation;
- Proving a general theorem (e.g., a law of exponents) by use of only two or three examples; and
- Giving the *procedure* of the long division of polynomials without explaining what it is about, i.e., never defining **division with a remainder**.

Readers should keep in mind that the error density of Type III errors is about 10% in these two Algebra books, i.e., students are going to find one such error every ten pages on average. This is definitely a cause for concern for both students and teachers.

(B) In this portion of the analysis, one chapter on linear equations in each of three other Algebra I textbooks is analyzed. These three books are referred to as b1, b2, and b3. Because the corresponding chapter in the Algebra I textbook in (A) is also reviewed, this book is referred to as a. Here are the findings of the *error densities* in the chapter on linear equations in these four books:

Table B-2: Error Densities in Chapters on Linear Equations

Book	Type I	Type II	Type III	Overall
a	21.7%	16.7%	6.7%	45%
b1	21.2%	6%	6%	33.3%
b2	14.9%	9.2%	3.5%	27.6%
b3	2.9%	2.9%	4.4%	10.3%

Note that two errors of Type III in the book b3 were inadvertently left out by the contractor, but the above computations of error densities did take these overlooked errors into account. The Type III errors involved are the following: One is on not mentioning the fact that the definition of slope of a line is independent of the choice of the two points in the definition, and the other is on not giving an explanation when this independence is mentioned in an example.

This table leaves open the question of whether the book b3, is in fact, significantly better than the rest of the available texts, regarding errors. An independent careful reading of this book suggests that, like the others, there is a concern relative to error frequency. This analysis again raises concern for teachers, students and all others using textbooks. It is imperative that authors, editors and publishers produce mathematically accurate textbooks.

Chapter 4: Report of the Task Group on Learning Processes

David C. Geary, Chair
A. Wade Boykin
Susan Embretson
Valerie Reyna
Robert Siegler
Daniel B. Berch, Ex Officio
Jennifer Graban, U.S. Department of Education Staff

CONTENTS

Figures

Tables

Abbreviations

CTBS	Comprehensive Test of Basic Skills
ECLS	Early Childhood Longitudinal Study
ELS	Education Longitudinal Study
GPA	Grade Point Average
GRE	Graduate Record Exam
HS&B	High School and Beyond
IRT	Item Response Theory
LD	Learning Disability
MLD	Mathematics Learning Disability
NALS	National Adult Literacy Survey
NAAL	National Assessment of Adult Literacy
NAEP	National Assessment of Education Progress
NCES	National Center for Education Statistics
NCTM	National Council of Teachers of Mathematics
NELS	National Education Longitudinal Study
OECD	Organisation for Economic Co-operation and Development
PISA	Programme for International Student Assessment
SAT	Scholastic Aptitude Test
SES	Socioeconomic Status
TEMA	Test of Early Mathematics Ability
TIMSS	Trends in International Mathematics and Science Study

Executive Summary

The charge of the Learning Processes Task Group is to address what is known about how children learn and understand areas of mathematics related to algebra and preparation for algebra. This summary provides brief overviews of and recommendations from the corresponding in-depth reviews provided later in the report. The reviews cover the areas of 1) General Principles of Cognition and Learning; 2) Social, Motivational, and Affective Influences on Learning; 3) What Children Bring to School; 4) Mathematical Development in Content Areas of a) Whole Number Arithmetic; b) Fractions; c) Estimation; d) Geometry; and e) Algebra; 5) Differences Among Individuals and Groups; and 6) The Brain Sciences and Mathematics Learning. For the mathematical content areas included in the reviews, the recommendations are organized around classroom practices, training of teachers and researchers, curriculum, and future research efforts.

General Principles: From Cognitive Processes to Learning Outcomes

Cognitive science is the study of the processes that underlie learning and cognition and is a foundational component of scientifically informed educational practice. There is a large body of high-quality research on learning mechanisms that can be directly applied to the classroom to improve student learning and achievement; however, this research at present is not being optimally used.

The two main classes of cognitive mechanism that control learning are information processing operations and mental representations. Students also engage in metacognitive processing, which controls information-processing operations such as selecting strategies for effective problem solving.

Information processing begins when a student encounters information and lasts until that information is acted upon and a response is made. The process starts with *attention*, without which information is lost. Information that is the focus of attention becomes available to learners' *working memory*, and with practice the information can be transferred to long-term memory. Deficiencies or superiorities in working memory capacities are major contributors to learning disabilities or accelerated learning, respectively. Improving the effectiveness of working memory can be assisted by achieving automaticity.

Mental representations are represented in different ways in the brain, including *declarative* knowledge, *procedural* knowledge, and *conceptual* knowledge.

The number line is a core tool in modern mathematics and is used in many contexts. One important cognitive mechanism in mathematics learning is the so-called *mental* number line.

Memories occur in either *verbatim* or *gist* form. Verbatim recall of math knowledge is an essential feature of math education, and it requires a great deal of time, effort, and practice. Gist memory is the form of memory that is typically relied on in reasoning. A combination of gist knowledge and verbatim knowledge is critical for success in math.

Social, Motivational, and Affective Influences on Learning

Children's goals and beliefs about learning are related to their mathematics performance. Mastery-oriented students are focused on learning the material and show better long-term academic development in mathematics and the pursuit of difficult academic tasks. Performance-oriented students are focused on grades and show less persistence on complex tasks. When students are told that beliefs about effort and ability can be changed, they are shown to undergo a significant rebound in their mathematics grades.

Young children's intrinsic motivation to learn is positively correlated with academic outcomes in mathematics and other domains and is related to mastery goals. Extrinsic motivation is related to performance goals.

Students' attributions or beliefs about the causes of their success and failure have been repeatedly linked to their engaging and persisting in learning activities. Students' self-regulation improves math learning.

Anxiety is an emotional reaction that is related to low math achievement, failure to enroll in advanced mathematics courses, and poor scores on standardized tests of math achievement. Math anxiety creates a focus of limited working memory on managing anxiety reaction rather than on solving the math problem, but it can be reduced by therapeutic interventions.

Vygotsky's characterization of the learning process as one of social induction may be applicable to the sharing of informal mathematics knowledge when it is embedded in everyday practices.

Recommendations

The Task Group recommends extension of experimental studies that have demonstrated that: 1) children's beliefs about the relative importance of effort and ability can be changed; 2) increased emphasis on the importance of effort is related to greater engagement and persistence on mathematics tasks; and 3) improved mathematics grades result from these changed beliefs.

The Task Group recommends studies that experimentally assess the implications of the relation between intrinsic motivation and mathematics learning.

The Task Group recommends experimental and longitudinal studies that assess the relative contributions of self-efficacy (i.e., the belief that one has the specific skills needed to be successful, which differs from self-esteem) factors to mathematics learning.

Although self-regulation (i.e., making goals, planning, monitoring, and self evaluating progress) appears promising, research is needed to establish the causal relation between these processes and the ability to learn a wider range of mathematics knowledge and skills.

The Task Group recommends research that assesses the potential risk factors of anxiety; it also recommends development of promising interventions for reducing debilitating mathematics anxiety.

A shortage of controlled experiments makes the usefulness of Vygotsky's approach for improving mathematics learning difficult to evaluate, and thus its utility in mathematics classrooms and mathematics curricula needs to be scientifically tested.

What Children Bring to School

Mathematical learning begins at birth and continues through the time children first arrive at school. The amount of mathematical knowledge students bring to school has important consequences for their long-term learning, as children who start kindergarten behind their peers tend to stay behind throughout their schooling.

Mathematical development begins in the first months of infancy, as people posses an innate nonverbal sense of number that provides a foundation for learning the verbal number system.

While 3- and 4-year old children may be able to count from 1 to 10, many have only mastered the superficial form of counting without understanding counting's purpose. By kindergarten, most children begin to understand the magnitudes of the numbers from 1 to 10.

By the start of kindergarten, most children also can retrieve from memory answers to a few basic addition and subtraction facts, know a variety of other procedures for solving simple addition and subtraction problems, and show some understanding of basic arithmetic concepts. Children of this age also choose effectively among strategies; use measurement strategies that reflect basic understanding of *more than*, *less than*, and *equal to*; and show basic geometrical knowledge of simple shapes.

Mathematical knowledge during preschool and kindergarten is predictive of mathematical knowledge in third, fifth, and eighth grade. Students who are at risk for low mathematics achievement tend to come from single-parent families with low-parental education levels, families where English is not the primary language, and families living in poverty. African-American and Hispanic children are more likely than other children to enter kindergarten with poor mathematical knowledge.

Effective instructional programs designed to improve mathematical knowledge of preschool children focus on forming mental representations of numbers, such as the mental number line, the language of numbers, and tools found through computer software programs.

Recommendations

Research that scales up interventions to improve the mathematical knowledge of preschoolers and kindergartners, especially those from at-risk backgrounds, and research that evaluates the utility of these interventions in classroom settings are urgently needed.

Mathematical Development in Content Areas

Whole Number Arithmetic

The mastery of whole number arithmetic is a critical step in children's mathematical education. The road to mastery involves learning arithmetic facts, algorithms, and concepts.

The quick and efficient solving of simple arithmetic problems is achieved when children retrieve answers from long-term memory or retrieve related information that allows them to quickly reconstruct the answer. Retention of these facts requires repeated practice.

Research indicates that learning of addition and multiplication facts is easier to achieve than learning of subtraction and division facts, due to the commuted relation within addition and multiplication pairs. Children and many adults in the United States have not reached the point of fast and efficient recall of simple arithmetic problems.

Algorithms range in complexity from counting as a way to solve simple addition problems to the lengthy sequence of steps involved in solving division problems. Learning of complex algorithms is highly dependent on working memory resources and requires repeated use of the algorithm extended over time. Mastery of standard algorithms is dependent on committing these problem-solving steps to long-term procedural memory, at which point the algorithm can be executed automatically with little demand on working memory resources. Algorithms that are mastered are less prone to disruption due to anxiety or in contexts such as high-stakes testing.

The core concepts that children should understand and use when solving arithmetic problems include mathematical equality, the commutative and associative properties of addition and multiplication, the distributive property of multiplication, identity elements for addition and multiplication, the composition of numbers, connections between arithmetic and counting, and the inverse relation between addition and subtraction and between multiplication and division.

Conceptual understanding is critical for children's ability to identify and correct errors, for appropriately transferring algorithms to solve novel problems, and for understanding novel problems in general.

Recommendations

Training

For teachers to take full advantage of formative assessments, they must have a better understanding of children's learning and the sources of children's conceptual and procedural errors in the content areas they are teaching. The development of courses in mathematical cognition for inclusion in teacher training programs will be necessary to address this goal.

Programs that support cross-disciplinary pre-doctoral and post-doctoral training in cognition, education, and mathematics are needed to ensure that a sufficient number of researchers study children's mathematical learning, and have the background needed to bridge the gap between laboratory studies and classroom practice.

Curricula

Although definitive conclusions cannot be drawn at this time due to lack of relevant, large-scale experimental studies, the research that has been conducted suggest that effective practice should: 1) present more difficult problems more frequently than less difficult problems, 2) highlight the relations among problems, 3) order practice problems in ways that reinforce core concepts, and 4) include key problems that support formative assessments.

Research

Although much is known about some areas of children's arithmetical cognition and learning, further research is needed in the areas of children's learning of complex algorithms; the relation between conceptual knowledge and procedural learning; and on the learning of core concepts, including the base-10 number system, the distributive property of multiplication, and identity elements, among others.

Studies that focus on the translation of cognitive measures of children's learning into formative assessments that are easily understood by teachers and easily used in the classroom are needed.

Funding priorities that target areas of deficit in children's arithmetical cognition and learning are recommended, along with priorities that encourage projects that bridge the gap between basic research and classroom practice.

Fractions

Fractions, decimals, and proportions are introduced into the mathematics curriculum as early as elementary school, and yet solving problems with these quantities remains difficult for many adults. Understanding and manipulating fractions is crucial for further progress in mathematics and for tasks of everyday life.

A fraction is defined as a point on the number line, based on the concept of a part-whole relation, with the unit segment [0,1] (the segment from 0 to 1) serving as a whole. From this mathematical definition of a fraction, other definitions can be derived, such as the division interpretation.

Difficulties with fractions extend beyond those with learning disabilities in mathematics. The failure to attain basic facility with fractions constitutes an obstacle to progress to more advanced topics in mathematics, including algebra and, presumably, to career paths that require mathematical proficiency, as well as interfering with potential life-and-death aspects of daily functioning (e.g., understanding and adhering to medical regimen).

To accurately assess competence, it is important to separate children's understanding of formal fractional notation from their intuitive ability to understand fractional relations and perform calculations using fractional quantities. Young children reveal a nascent ability to understand ratios, and preschool children's experiences with and understanding of part-whole relations among sets of physical objects may contribute to an early understanding of simple ratios.

Similarly, the ability to manipulate fractions is also present early. Research shows that sharing forms the basis for preschool age children's ability to partition a quantity into roughly equal parts through a process of distributive counting. This does not mean that they understand the inverse relation among quantities, but with a few lessons they are able to appreciate and generalize the inverse relation.

Studies of elementary and middle school-aged children have focused on the acquisition of conceptual knowledge, computational skills, and the ability to use both of these abilities in conjunction with reading comprehension to solve word problems involving fractional quantities. Scores on items assessing conceptual knowledge have consistently been shown to explain unique variance (beyond general intellectual and reading abilities) in performance on computational fraction problems, word problems that include fractions, and estimation tasks with fractional quantities.

Many errors on fraction computation problems could be classified as involving a faulty procedure. Children's accuracy at recognizing formal procedural rules for fractions and automatic retrieval of basic arithmetic facts predicts computational skills, above and beyond the influence of intelligence, reading skills, and conceptual knowledge. Research also shows that on-task time influences performance through its effect on conceptual knowledge.

Motivation also has positive effects on fraction learning. Learning goals rather than performance goals may produce higher self-efficacy, skill, and other achievement outcomes in students. Performance goals with self-evaluation components may be more effective than without. Early levels of basic arithmetic skills may predict those children who will later have difficulty with fractions, and building such skills may enhance performance on fraction computation problems.

Proportional reasoning involves the coordination of two ratio quantities, and early, informal competence can be detected if children are able to use perceptual cues to judge relative numerosity.

A fraction's lack of fit with properties of counting adds to the relative difficulty of learning the concept. Because of the property of infinite divisibility, fractions, unlike counting numbers, do not form a sequence in which each number has a fixed successor. Therefore, it has been argued that the one-to-one and stable-order principles that are important to counting are misleading when children attempt to generalize from whole numbers to fractions.

Pictorial representations, without sufficient emphasis on the nature of wholes in part-whole relations and the importance of equal-sized parts, also may be an obstacle to learning fractions. Number line representation may be more effective. Words also seem to influence the mental representations that children form concerning fractions, particularly when language demarcates parts and wholes in fraction names.

Research on working-memory demands of different tasks shows that different fraction interpretations entail different information-processing demands. Quotient interpretations of fractions are more demanding of memory resources than part-whole interpretations because they involve a more complex series of mappings.

Individual differences in working memory have been associated with performance on fraction tasks; and effects of working memory were independent of effects of conceptual knowledge. While conceptual knowledge carries the greatest weight in predicting performance on all three outcome measures (computation, estimation, and word problems), working memory affected only word problems and only indirectly affected computation through knowledge of basic arithmetic facts.

Recommendations

Classroom

Children should begin fraction instruction with the ability to quickly and easily retrieve basic arithmetic facts. Instruction focusing on conceptual knowledge of fractions is likely to have the broadest and largest impact on problem-solving performance. Procedural knowledge is also essential, however, and although it must be learned separately, it is likely to enhance conceptual knowledge and vice versa.

Successful interventions should include the use of fraction names that demarcate parts and wholes, the use of pictorial representations that are mapped onto the number line, and composite representations of fractions that are linked to representations of the number line. Conceptual and procedural knowledge about fractions less than one do not necessarily transfer to fractions greater than one, and must be taught separately. Appropriate intuitions about sharing, part-whole relations, and proportional relations can be built on in classrooms to support acquisition of conceptual and procedural knowledge of fractions.

Training

Training of teachers should include sufficient coverage of the scientific method so that teachers are able to critically evaluate the evidence for proposed pedagogical approaches and to be informed consumers of the scientific literature. Teachers should be aware of common conceptions and misconceptions involving fractions and of effective interventions involving fractions.

New funding should be provided to train future researchers, to begin new interdisciplinary degree programs with rigorous quantitative training, and to establish support mechanisms for career shifts that encourage rigorous researchers in related fields to focus on education.

Curriculum

The curriculum should allow for sufficient time on task to ensure acquisition of conceptual and procedural knowledge of fractions and of proportional reasoning, with the goal for students being one of learning rather than performance. However, there should be ample opportunity in the curriculum for accurate self-evaluation. The curriculum should include representational supports that have been shown to be effective and tap the full range of conceptual and procedural knowledge.

Research

An area for future study is the relation between the rudimentary understanding of very simple fractional relations and the learning of formal mathematical fractional concepts and procedures. In addition, research is needed to uncover the mechanisms that contribute to the emergence of formal competencies. Research on understanding and learning of fractions should be integrated with what is known and with emerging knowledge in other areas of basic research, such as neuroscience, cognition, motivation, and social psychology. The absence of a coherent and empirically supported theory of fraction tasks is a major stumbling block to developing practical interventions to improve performance in this crucial domain of mathematics.

Classroom-relevant research need not be conducted physically in classrooms, and constraints on funding that require that relevant research be performed in classrooms should be removed. Conversely, many interventions demonstrated to be effective in experiments should be scaled up and evaluated in classrooms.

Estimation

Estimation may be used more often in everyday life than any other quantification process. It is also quite strongly related to other aspects of mathematical ability, such as arithmetic skill and conceptual understanding of computational procedures, and to overall math achievement test scores. It usually requires going beyond rote application of procedures and applying mathematical knowledge in flexible ways.

The Task Group focuses on *numerical estimation*, the process of translating between alternative quantitative representations, at least one of which is inexact and at least one of which is numerical. This category includes many prototypic forms of estimation, including computational, number line, and numerosity.

Many children have highly distorted impressions of the goals of estimation, especially the goals of computational estimation. Accurate computational estimation requires understanding of the simplification principle and the proximity principle. Research shows that students understand the principle of simplification, but they show little if any understanding of the importance of generating an estimate close in magnitude to the correct answer.

Development of computational estimation skills begins surprisingly late and proceeds slowly but does improve considerably with age and experience. From early in the development of computational estimation, individual children use a variety of strategies including rounding, truncating, prior compensation, post-compensation, decomposition, translation, and guessing. Rounding is the most common approach and compensation tends to be among the least common, although it is among the most useful.

Both children and adults adapt their strategy choices to problem characteristics. The range and appropriateness of computational estimation strategies increase with age and mathematical experience. The sophistication of strategies used also changes, and in particular, compensation shows especially substantial growth with age and experience.

The number line task has proved highly informative, not only for improving understanding of estimation but also for providing useful information about children's understanding of the decimal number system more generally.

Children use two primary mental representations of numerical magnitude on number line estimation tasks, including linear representation and logarithmic representation. With age and experience, children progress from using the less accurate logarithmic representation to the more accurate linear one on the number line task.

Both children and adults show substantial individual differences in skill at computational estimation that are associated with broader individual differences in mathematical understanding and general mathematical ability.

Playing board games with linearly arranged, consecutively numbered, equal-size spaces leads children to shift from logarithmic to linear representations of numerical magnitude. These games are particularly effective in improving low-income preschoolers' numerical knowledge and reducing disparities in the numerical knowledge brought to school by children from low-income homes and those from middle-income homes.

Another procedure that is effective for improving elementary school children's number line estimation is to provide students with feedback on their estimates.

Recommendations

Classroom

Teachers should broaden instruction in computational estimation beyond rounding. They should insure that students understand that the purpose of estimation is to approximate the correct value and that rounding is only one of several means for accomplishing this goal.

Teachers should provide examples of alternative procedures for compensating for the distortions introduced by rounding, emphasize that there are many reasonable procedures for estimating rather than just a single correct one, and discuss reasons why some procedures are reasonable and others are not.

Teachers in facilities serving preschoolers from low-income backgrounds should be made aware of the usefulness of numerical board games for improving the children's knowledge of numbers and of the importance of such early knowledge for long-term educational success.

Teachers should not assume that children understand the magnitudes represented by fractions even if the children can perform arithmetic operations with them. Examining children's ability to perform novel estimation tasks, such as estimating the positions of fractions on number lines, can provide a useful tool for assessing children's knowledge of fractions. Providing feedback on such number line estimates can improve children's knowledge of the fractions' magnitudes.

Training

Teachers in preservice and in-service programs should be informed of the tendency of elementary school students to not fully understand the magnitude of large whole numbers, and they should be taught how to assess individual students' understanding and research-based techniques for improving the children's understanding.

Teachers should be made aware of the inadequate understanding by elementary school, middle school, and high school students of the magnitudes of fractions. Teachers also should be familiarized with the usefulness of feedback on number line estimates of the magnitudes of fractions for overcoming these difficulties.

Curriculum

Textbooks need to explicitly explain that the purpose of estimation is to produce accurate approximations. Illustrating multiple useful estimation procedures for a single problem and explaining how each procedure achieves the goal of accurate estimation are useful means for achieving this goal. Contrasting these procedures with others that produce less accurate estimates and explaining why the one set of procedures produces more accurate estimates than the other are also likely to be helpful.

Research

Research is needed regarding simple instruments that teachers can use in the classroom for assessing children's estimation skills, and regarding instruction that can efficiently improve children's estimation.

Research is needed on how the inadequate representations of whole number numerical magnitudes that have been identified by studies of estimation influence learning of other mathematical skills, such as arithmetic.

Research is needed on how children can be taught to accurately estimate the magnitudes of fractions and on how learning to estimate those magnitudes influences acquisition of other numerical skills involving fractions, such as arithmetic and algebra.

Research is needed on how estimation is used by students (e.g., to solve complex problems) and by adults in everyday life and in professional tasks (e.g., to rule out implausible answers).

Geometry

Geometry is the branch of mathematics concerned with properties of space, and of figures and shapes in space. Euclidean geometry is the domain typically covered in mathematics curricula in the United States, although a separate year-long course is not usually taught until high school. Units on geometry as well as measurement are frequently included in middle school mathematics classes, whereas only the latter tends to be emphasized in the elementary grades.

The Conceptual Knowledge and Skills Task Group found that the single aspect of geometry that is most directly relevant for early learning of algebra is that of similar triangles. NCTM's *Focal Points* and some state frameworks also underscore the importance of this aspect of geometry.

To understand the mathematics underlying the proof that the slope of a straight line is independent of the choice of the points selected, students must successfully develop a conceptual understanding of the following: points, lines, length, angle, right triangle, correspondence, ratio, proportion, translation, reflection, rotation, dilation, congruence, and similarity.

One of the earliest and most influential theories of the development of spatial and geometric concepts was put forth by Piaget and Inhelder, who proposed that young children initially conceptualize space and spatial relations topologically as characterized by the following properties: proximity, order, separation, and enclosure. With development, children subsequently begin to represent space in relation to different points of view, and then sometime between middle and late childhood the Euclidean conceptual system emerges permitting preservation of metric relationships such as proportion and distance. The consensus of research is that evidence supporting this developmental model is comparatively weak.

The van Hiele model (1986) has been the dominant theory of geometric reasoning in mathematics education for the past several decades. According to this model the learner moves sequentially through five levels of understanding: Level 0: Visualization/Recognition, Level 1: Description/Analysis, Level 2: Informal Deduction or Ordering, Level 3: Formal Deduction, and Level 4: Rigor. The majority of high school geometry courses are taught at Level 3.

Research shows that the van Hiele theory provides a generally valid description of the development of students' geometric reasoning, yet this area of research is still in its infancy.

A common misconception that impedes learning includes the belief that shapes with the same perimeter must have the same area. Initial formal instruction may inadvertently promote this misconception as a consequence of students being presented with the concepts of perimeter and area pertaining to the same shapes and kinds of problems.

In addition, there is a common misconception that the linear (or proportional) model can pertain to situations where it is, in fact, not applicable. Research has found that only a long-term classroom intervention can produce a positive effect in overcoming the illusion of linearity.

Recommendations

Classroom

Teachers should recognize that from early childhood through the elementary school years, the spatial visualization skills needed for learning geometry have already begun to develop. Proper instruction is needed to ensure that children adequately build upon and make explicit this core knowledge for subsequent learning of formal geometry.

Training

Teachers need to learn more about the latest research concerning the development of children's spatial abilities, in general, and their geometric conceptions and misconceptions, in particular, to capitalize on their strengths and aid them in overcoming their weaknesses.

Researchers investigating geometry learning need to have a firm grounding in cognitive development and spatial information processing, in addition to having a background in mathematics education.

Curriculum

Early exposure to common shapes, their names, and so forth appears to be beneficial for developing young children's basic geometric knowledge and skills. While reliance on manipulatives may enhance the initial acquisition of some concepts under specified conditions, students must eventually transition from concrete or visual representations to internalized abstract representations. The crucial steps in making such transitions are not clearly understood at present.

Research

Longitudinal studies are needed to assess more directly how developmental changes in spatial cognition can inform the design of instructional units in geometry. Studies are needed to demonstrate whether and to what extent knowledge about similar triangles enhances the understanding that the slope of a straight line is the same regardless of the two points chosen, thus leading to a more thorough understanding of linearity.

More research is needed that specifically links cognitive, theory-driven research to classroom contexts. At the same time, cognitive theorizing pertaining to geometry learning needs to take into account more facets of classroom settings if it is to eventually have a large impact on the design of instructional approaches.

Algebra

Because it is not known if the early algebra achievement of elementary school children reflects an actual implicit understanding of aspects of algebra, the Task Group focuses on explicit algebra content typically encountered in middle school to high school algebra courses.

Studies of skilled adults and high school students who have taken several mathematics courses reveal that the processing of algebraic expressions is guided by an underlying syntax or system of implicit rules that guides the parsing and processing of the expressions.

Research shows that skilled problem solvers scan and process basic subexpressions in these equations in a fraction of a second, or have automaticity. There are substantial benefits to cumulative practice, which results in better short-term and long-term retention of individual rules and a better ability to apply rules to solve problems that involve the integration of multiple rules and to discriminate between rules that might otherwise be used inappropriately.

Students who are first learning algebra and adults who are not skilled in mathematics do not have long-term memory representations of basic forms of linear equations, but this does not prevent the solving of linear equations as long as they understand the general arithmetical and algebraic concepts and rules. Research shows that diagnostic tests in which individual problems varied systematically in terms of the knowledge needed for correct solution can identify sources of common errors, such as those that reflect a poor conceptual understanding of the syntax of algebraic expressions.

A poor understanding of the concept of mathematical equality and the meaning of the "=" is common for elementary school children in the United States, and continues for many children into the learning of algebra.

Errors in the solving of algebraic equations are sometimes classified as procedural bugs. These errors can occur due to overgeneralized use of procedures that are correct for some problems or from a misunderstanding of the procedure itself. Preliminary studies suggest that remediation that focuses on these specific bugs can reduce their frequency.

Research shows that the solution of algebraic word problems requires two general sets of processes: problem translation and problem solution. Problem translation requires an understanding of the meaning and implications of the text within which the problem is embedded. The same potential sources of error described for solving of linear equations can occur during the problem solution stage of word problems.

An analysis of word problems presented in algebra textbooks found that most problems included four types of statements: assignment statements, relational statements, questions, and relevant facts. Problem translation involves taking each of these forms of information and using them to develop corresponding algebraic equations. Translation errors most frequently occur during the processing of relational statements, which specify a single relationship between two variables. Errors also occur in problems where the statement could be *directly* translated into an equation, but the direct translation is incorrect for the problem as a whole (e.g., “z is equal to the sum of 3 and y”). In addition, relational information can sometimes aid problem solving if the information is consistent with students’ previous out-of-classroom experiences and if these experiences can be used to create non-algebraic solution strategies.

Abstract problems are more difficult to solve than concrete problems, but the largest effect on students’ problem-solving skill is their familiarity with solving the class of word-problem (e.g., interest, rate).

Successful translation of algebraic word problems, as well as the solution of algebraic equations and many other problem types, is guided by schemas including the syntax of equations. Research on children’s conceptual knowledge, which was inferred based on how they sorted word problems into categories, shows that the ability to categorize word problems based on the underlying concept and the corresponding reduction in problem solving errors is consistent with development of category-specific schemas.

Researchers have demonstrated that one way in which schema development can occur is the use of worked examples. These provide students with a sequence of steps that can be used to solve problems. The students then solve a series of related problems that are in the same category and involve a very similar series of problem-solving steps. Worked examples are more effective than simply providing students with the procedural steps, as they may promote the automatization and transfer of procedures used across classes of problems.

The best predictors of the ability to solve word problems are computational skills and knowledge of mathematical concepts, as well as intelligence, reading ability, and vocabulary. Students who struggle with algebraic equations also make factoring errors and use algebraic procedures incorrectly. At a cognitive level, problem-solving errors and learning the syntax of algebraic expressions and algebraic schemas are influenced by working memory. Accuracy at solving various forms of mathematics word problems is also related to spatial abilities. It is also very likely that other factors, including motivation, self-efficacy, and anxiety, contribute to skill development in algebra.

Research on learning in general indicates a benefit for practice that is distributed across time, as contrasted with the same amount of practice massed in a single session. Algebraic skills decline steadily over time, and the best predictor of long-term retention of competencies in algebra is the number of mathematics courses taken beyond Algebra I.

Recommendations

Classroom

Teachers should not assume that all students understand even basic concepts, such as equality. Many students will not have a sufficient understanding of the commutative and distributive properties, exponents, and so forth to take full advantage of instruction in algebra.

Many students will likely need extensive practice at basic transformations of algebraic equations and explanation as to why the transformations are done the way they are. The combination of explanation of problem-solving steps combined with associated concepts is critically important for students to effectively solve word problems. For both equations and word problems, it is important that students correctly solve problems before given seatwork or homework.

Training

Teachers should understand how students learn to solve equations and word problems, and causes of common errors and conceptual misunderstandings. This training will better prepare them for dealing with the deficiencies students bring to the classroom, and for anticipating and responding to procedural and conceptual errors during instruction.

The next generation of researchers to study algebra learning will need multi-disciplinary training in mathematics, experimental cognitive psychology, and education. This can be achieved through interdisciplinary doctoral programs or, at a federal level, postdoctoral fellowships that involve work across these disciplines.

Curriculum

There are aspects of many current textbook series in the United States that contribute to the poor preparation and background of algebra students. Presenting operations on both sides of the equation; and showing worked-out examples that include conceptual explanation, procedural steps, and multiple examples are ways in which textbooks can be improved.

Distributed practice should naturally occur as students progress to more complex topics. However, if basic skills are not well learned and understood, the natural progression to complex topics is impeded.

Research

The development of assessment measures that teachers can use to identify core deficiencies in arithmetic (whole number, fractions, and decimals) and likely sources of procedural and conceptual errors in algebra are needed.

Research that explicitly explores the relation between conceptual understanding and procedural skills in solving algebraic equations is needed. Research on how students solve linear equations, and where and why they make mistakes needs to be extended to more complex equations and other key topic areas of Algebra identified by the Conceptual Knowledge and Skills Task Group.

The issue of transfer needs considerable attention, particularly determining the parameters that impede or facilitate transfer. Research on instructional methods that will reduce the working memory demands associated with learning algebra is needed. Longitudinal research is needed to identify the early predictors of later success in algebra.

A mechanism is needed for fostering translation of basic research findings into potential classroom practices and for scientifically assessing their effectiveness in the classroom. Equally important, mechanisms need to be developed for reducing the lag time between basic findings and assessment in classroom settings.

Differences Among Individuals and Groups

For large, nationally representative samples, the average mathematics scores of boys and girls are very similar; when differences are found they are small and typically favor boys.

From preschool to college, there is a mathematics performance gap between black and Hispanic students to their white and Asian counterparts. It is often proposed that socioeconomic status differences account for these disparities, but the research indicates that this is not a sufficient explanation. Other factors include attitudes, beliefs, motivation, and school-based factors such as features of teaching and learning contexts.

Stereotype threat, cognitive load, and strategy use are all potential mechanisms contributing to existing differences, and work in those areas holds promise as a means to improve the mathematics performance of black and Hispanic students. There is not, however, sufficient research to fully evaluate this promise.

There is strong support for a relation between motivational and attitudinal factors, especially task engagement and self-efficacy, and the mathematics outcomes for black and Hispanic students. Recent research also documents that social and intellectual support from peers and teachers is associated with higher mathematics performance for all students and that such support is especially important for black and Hispanic students.

At least 5% of students will experience a significant learning disability in mathematics before completing high school, and many more children will show learning difficulties in specific mathematical content areas.

There are only a few cognitive studies of the sources of the accelerated learning of mathematically gifted students, but those that have been conducted suggest an enhanced ability to remember and process numerical and spatial information. Quasi-experimental and longitudinal studies consistently reveal that accelerated and demanding instruction is needed for these students to reach their full potential in mathematics.

Recommendations

Research efforts are needed in areas that assess the effectiveness of interventions designed to: 1) reduce the vulnerability of black and Hispanic students to negative stereotypes about their academic abilities, 2) functionally improve working memory capacity, and 3) provide explicit instruction on how to use strategies for effective and efficient problem solving.

More experimental work is needed to specify the underlying processes that link task engagement and self-efficacy, and the mathematics outcomes for black and Hispanic students. Urgently needed are a scaling-up and experimental evaluation of the interventions that have been found to be effective in enhancing engagement and self-efficacy for black and Hispanic students.

Intervention studies of students with a mathematics learning disability (MLD) are in the early stages and should be a focus of future research efforts. Further research also is needed to identify the sources of MLD and learning difficulties in the areas of fractions, geometry, and algebra.

Brain Sciences and Mathematics Learning

Brain sciences research has the potential to contribute to knowledge of mathematical learning and eventually educational practices, yet attempts to make these connections to the classroom are premature. Instructional programs in mathematics that claim to be based on brain sciences research remain to be validated. Yet, promising research emerging from the field of cognitive neuroscience is permitting investigators to begin forging links between neurobiological functions and mathematical cognition.

Most research making use of brain imaging and related techniques has focused on basic mental representations of number and quantity, with a few studies exploring problem solving in arithmetic and simple algebra. In most of these studies, researchers have contrasted, mapped, and differentiated the brain regions active during mathematical activities. It has been repeatedly found that comparisons of number magnitudes, quantitative estimation, use of a mental number line, and problem solving in arithmetic and algebra activate several areas of the parietal cortex. The intraparietal sulcus is also active when nonhuman animals engage in numerical activities, and it has been proposed that a segment of this sulcus, particularly in the left hemisphere, may support an inherent number representational system.

Research also shows that the hippocampus, which supports the formation of declarative memories, is active when involved in the learning of basic arithmetic facts. Other studies suggest the parietal cortex in the adolescent brain may be more responsive than the same regions in the adult brain when individuals are learning to solve simple algebraic equations. Another study suggests differences in the brain regions that contribute to success at solving algebraic word problems and algebraic equations. In addition, research shows there may be differences in the network of posterior brain regions engaged during the learning of different arithmetical operations.

In coming years, brain imaging and related methodologies will almost certainly help answer core questions associated with mathematical learning, such as the sources of learning disabilities and the effects of different forms of instruction on the acquisition of declarative, conceptual, and procedural competencies.

Recommendations

Brain sciences research has a unique potential for contributing to knowledge of mathematical learning and cognition and eventually educational practices. Nevertheless, attempts to connect research in the brain sciences to classroom teaching and student learning in mathematics should not be made until instructional programs in mathematics based on brain sciences research are created and validated.

I. Introduction

This report reflects the work of the Task Group on Learning Processes and addresses what is known about how children learn mathematical concepts and skills. The discussion begins with an introduction to the basic principles of learning and cognition, as well as the social and motivational factors that are relevant to educational practices, and to skill development in the focal mathematical domains addressed. The focus then moves to a review of the mathematical competencies that many children bring to school, followed by reviews of research on conceptual and procedural learning in the core content areas of whole number arithmetic, fractions, estimation, geometry, and algebra. These reviews summarize the scientific literature on what is known about learning within each of these areas and identify areas in which future study is needed before definitive conclusions can be made. Next, the report addresses individual and group differences in achievement in these core domains or in mathematics achievement up to and including algebra; the Task Group addresses mathematics achievement as related to race and ethnicity, gender, learning disabilities, and giftedness. The Task Group report closes with a discussion of future directions, specifically the implications of recent advances in the brain sciences for understanding mathematical learning.

II. Methodology

For all areas and to the extent that high-quality literature was available, the reviews and conclusions of the Task Group are based primarily on studies that test explicit hypotheses about the mechanisms promoting the learning of declarative knowledge (arithmetic facts), procedural knowledge, and conceptual knowledge. The evidence regarded as strongest for this purpose is that which shows convergent results across procedures and study types. When the evidence is not as strong, conclusions are qualified and suggestions are provided for research that will strengthen the ability to draw conclusions.

The multiple approaches, procedures, and study types reviewed and assessed with regard to convergent results include the following:

- Verbal report (e.g., of problem solving approaches).
- Reaction time and error patterns.
- Priming and implicit measures.
- Experimental manipulation of process mechanisms (e.g., random assignment to dual task, or practice conditions).
- Computer simulations of learning and cognition.
- Studies using brain imaging and related technologies.
- Large-scale longitudinal studies.
- International comparisons of math achievement.
- Process-oriented intervention studies.

A. Procedures

1. Literature Search and Study Inclusion

Literature searches were based on key terms linking mathematical content, and learning and cognitive processes (Appendix B). The first search focused on core peer-reviewed learning, cognition, and developmental journals (see Appendix A). A second search supplemented the first and included other empirical journals indexed in PsychInfo and the Web of Science.

2. Criteria for Inclusion

- Published in English.
- Participants are age 3 years to young adult.
- Published in a peer-reviewed empirical journal, or a review of empirical research in books or annual reviews.
- Experimental, quasi-experimental, or correlational methods.

III. Reviews and Findings

A. General Principles: From Cognitive Processes to Learning Outcomes

Cognitive science is the basic discipline that underlies studies of human learning, including learning of academic material, just as biology is the basic discipline that underlies medical practice and physics is the basic discipline that underlies engineering. In all three cases, the basic science identifies the causal pathways to successful outcomes. The next few pages describe the key cognitive processes that control learning: information processing operations (attention, working memory, retrieval, transfer, and retention; Section 1), and mental representations (declarative, procedural, and conceptual knowledge; verbatim and gist memories; Section 2). Students also engage in metacognitive processes, which are processes that control cognitive operations, such as explicitly selecting and monitoring strategies for effective problem solving (Section 1). Students' ability to orchestrate these various cognitive and metacognitive operations depends on the maturity of their prefrontal cortex, which controls attention and working memory, as well as on specific brain regions engaged in the representation of concepts or procedures. Examples of how these cognitive and metacognitive processes affect mathematics learning are presented, as are research-based methods of enhancing each process and thereby potentially improving mathematics learning. These examples and others in the sections that follow illustrate the utility of cognitive research for understanding learning, and suggest that teachers, superintendents, policy makers, curriculum developers, and anyone else whose goal is to increase student achievement, would advance that goal by having at least a rudimentary knowledge of the basic science of cognition.

There is a great deal of scientific knowledge that could be applied today to improve learning and student achievement. Much of that knowledge is currently not used in the nation's classrooms.

The concepts, principles, and processes of cognition presented here are supported by high-quality scientific research. This research provides insights, and sometimes immediate applications, to how student learning can be improved. There is much scientific knowledge that could be applied today to improve learning and student achievement (e.g., Cepeda et al., 2006). However, much of that knowledge is not currently being applied in the nation's classrooms. The following sections provide a review of scientific evidence about topics ranging from simple information processing to complex problem solving. Even creativity has been studied scientifically; excellent work on this topic was conducted in the 1950s and continues to the present day (e.g., Holyoak & Thagard, 1995; Sternberg, 1999). Therefore, this report proceeds through each of the cognitive building blocks to student achievement, to informed citizenship, and to career development in fields that require mathematical proficiency.

Basic research in cognitive science, especially research on the factors that promote learning, provides an essential grounding for the development and evaluation of effective educational practices.

What is cognition? Cognition encompasses attention, learning, memory, conceptual understanding, and problem solving, among other "higher" mental processes. General principles of cognition underlie learning and achievement in mathematics, and other academic domains. Test performance in mathematics, for example, is the end product of cognitive processes that include encoding and storing what has been taught, and retrieving it in response to test questions. Because achievement outcomes are critically dependent on the proper sequencing and execution of multiple cognitive operations, obtaining appropriate outcomes requires instruction to be based on a sound scientific foundation. The analogy to medicine is direct: Understanding the causal pathways that produce healthy outcomes (or go awry and result in disease) allows medical researchers to fashion drugs and therapies to achieve better outcomes. Understanding causal pathways in education works the same way as in medicine as it identifies the steps in the learning process that lead to successful outcomes, as well as missteps in the process and how these can be fixed. Just as in medicine, however, interventions derived from basic science about causal pathways must be tested for practical efficacy in educational settings (much like Phase III clinical trials in medicine).

Cognitive factors are not the only causal factors that have been linked to achievement outcomes. Nevertheless, all factors eventually have their effect via cognition.

Cognitive factors are not the only causal factors that have been linked to achievement outcomes; motivation, anxiety, nutrition, stereotypes, brain functioning, and tangible resources, such as availability of quality teachers and textbooks, are among other factors also relevant to achievement (e.g., Ashcraft, 2002; Cadinu et al., 2005; see following sections in this report). Nevertheless, these factors influence learning outcomes by virtue of their effects on cognitive processing. As an illustration, individuals who are anxious about mathematics perform worse on mathematics tests and on other mathematics tasks than their less anxious peers. The finding that interventions, such as cognitive behavioral therapy, can substantially improve the mathematical performance of many of these individuals indicates that their initial deficit is not related to the ability to learn mathematics (Hembree, 1990). Cognitive studies have identified working memory as one source of the lower mathematics achievement of individuals with mathematics anxiety; while performing math tasks, these anxious individuals have thoughts related to their competence intrude into working memory (described below), which disrupts their problem solving (Ashcraft & Kirk, 2001; Ashcraft, & Krause, 2007). Beilock, Kulp, Holt, & Carr (2004) demonstrated that similar intrusions into working memory can disrupt arithmetical problem solving in high-pressure testing situations but only when the procedures are not well learned; the execution of procedures committed to long-term memory was not disrupted by high-pressure testing. Hence, anxiety disrupts performance by affecting cognitive processing (i.e., by overloading working memory) and interventions to reduce that disruption have been shown to be effective. A goal of the present report is to identify such relevant findings and principles that have emerged from cognitive research and to suggest how they could be used to improve educational practice.

1. Information Processing

Attention is the gateway to the mind and, thus, to learning.

Information processing begins when the student first encounters information and extends until that information is operated on (or transformed) and a response is made, such as when a solution to a problem is produced. The first step in information processing is attention (e.g., Cowan, 1995; Pashler, 1999). Attention is a limited capacity faculty, often described as a bottleneck in information processing. Thus, only a portion of information in the environment can be attended to at any one time. Attention is crucial to learning; information that is unattended is lost to the learner. Distractions, such as noise, further limit the ability to pay attention. In addition, attention changes developmentally: Younger children are less attentive than older children (and adults), and distractions are more costly to younger children.

The ability to pay attention should not be confused with the motivation or desire to pay attention. No matter how much younger children may wish to pay attention, their ability to do so is lower than that of older children (Cowan, Saults, & Elliott, 2002). However, specific practices and environmental supports can enhance younger children’s ability to attend (described below).

Because attention is the first step in information processing on which all subsequent steps depend, deficits in attention necessarily influence learning. Educational practices and environmental accommodations can improve children’s ability to pay attention, such as by limiting irrelevant distractions (especially in the early phases of learning). For example, guiding children’s attention to where the 0 is in comparing .03 to .30 has been shown to be effective in improving performance on judgments of relative magnitude (Rittle-Johnson et al., 2001). Recent evidence also suggests that self-regulation—intentional efforts to control attention and behavior—can be improved with practice (Baumeister, 2005; Gailliot, Plant, Butz, & Baumeister, 2007; Muraven, Baumeister, & Tice, 1999).

Working-memory capacity limits mathematical performance, but practice can overcome this limitation by achieving automaticity.

Once information is attended to, it can be encoded into working memory. Working memory is the ability to hold a mental representation of information in mind while simultaneously engaging in other mental processes. Working memory is composed of a central executive that is expressed as attention-driven control of information represented in one of three content-specific systems (Baddeley, 1986, 2000; Engle, Conway, Tuholski, & Shisler, 1995). These systems are a language-based phonetic buffer, a visuospatial sketch pad, and an episodic buffer (i.e., memories of personal experiences). The workings of these systems can be illustrated in a simple arithmetic context. Students initially solve simple addition problems, such as 3 + 4, by means of counting fingers or manipulatives. The child's ability to control the counting process is influenced by the central executive; if the counting process is not well controlled by the central executive, the child may skip a finger or manipulative, or count a single object twice. The representation of the spoken numbers is in the phonetic buffer; if the phonetic buffer is insufficient, the child may need to repeat a number that has already been stated or may skip a number. The visuospatial sketch pad would come into play if the child were counting imagined objects; insufficiencies here might lead to too few or too many objects being imagined, and therefore to inaccurate counts.

With practice, the addends and answers on problems that have been solved are transferred from working memory into more permanent long-term memory. As illustrated in later sections, deficient working memory is a major contributor to the learning problems encountered by children with mathematical learning disabilities and superior working memory is a major contributor to the accelerated learning shown by gifted children.

Working memory capacity increases as children grow older, due to improvements in their ability to control attention and to increases in the fundamental capacity of the content-specific systems (Cowan et al., 2002). At all ages, there are several ways to improve the functional capacity of working memory. The most central of these is the achievement of automaticity, that is, the fast, implicit, and automatic retrieval of a fact or a procedure from long-term memory (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). Some types of information, such as facial features, are processed automatically and without the need for any type of instruction (Schyns, Bonnar, & Gosselin, 2002). For other types of information, including much of mathematics that is taught in school, automaticity is achieved only with specific types of experiences, including practice that is distributed across time (e.g., Cooper & Sweller, 1987).

For example, repeated practice with addition facts, such as 3 + 4 = 7, eventually transforms addition from a conscious resource-demanding process (e.g., counting on one's fingers) to an automatic process, freeing up much-needed mental resources for other aspects of problem solving (Groen & Parkman, 1972; Siegler & Shrager, 1984). The ability to efficiently retrieve basic arithmetic facts has been shown to be integral to more complex, conceptual mathematical thinking and problem solving (Geary & Widaman, 1992). As Gersten and Chard (2001) state, "if too much energy goes into figuring out what 9 plus 8 equals, little is left over to understand the concepts underlying multi-digit subtraction, division, or complex multiplication." As discussed below research has demonstrated that declarative knowledge (e.g., memory for addition facts), procedural knowledge (or skills), and conceptual knowledge are mutually reinforcing, as opposed to being pedagogical alternatives. As discussed in greater detail in later sections, to obtain the maximal benefits of automaticity in support of complex problem solving, arithmetic facts and fundamental algorithms should be thoroughly mastered, and indeed, over-learned, rather than merely learned to a moderate degree of proficiency.

Young children are capable of far greater learning of mathematics than those in the United States typically attain.

Learning and development are incremental processes that occur gradually and continuously over many years (Siegler, 1996). Even during the preschool period, children have considerably greater reasoning and problem solving ability than was suspected until recently (Gelman, 2003; Gopnik, Meltzoff, & Kuhl, 1999). As stated in a recent report on the teaching and learning of science, "What children are capable of at a particular age is the result of a complex interplay among maturation, experience, and instruction. What is developmentally appropriate is not a simple function of age or grade, but rather is largely contingent on prior opportunities to learn" (Duschl, Schweingruber, & Shouse, 2007, p. 2). Claims based, in part, on Piaget's highly influential theory that children of particular ages cannot learn certain content because they are "too young," "not in the appropriate stage," or "not ready" have consistently been shown to be wrong (Gelman & Williams, 1998). Nor are claims justified that children cannot learn particular ideas because their brains are insufficiently developed, even if they possess the prerequisite knowledge for learning the ideas. As noted by Bruer (2002), research on brain development simply does not support such claims.

These findings have special relevance to mathematics learning. Research on students in East Asia and Europe show that children are capable of learning far more advanced math than those in the United States typically are taught (Geary, 2006). There is no reason to think that children in the United States are less capable of learning relatively advanced mathematical concepts and procedures than are their peers in other countries.

Practice retrieving information from memory can improve learning more than another opportunity to study.

Attending to information, encoding it into working memory, and eventually transferring it into long-term memory are only the initial steps in learning. The learner must also be able to retain the information in long-term memory storage until needed (e.g., on tests or on the job), sometimes over long periods of time, and be able to retrieve it from storage. One counterintuitive finding from these studies is that testing, which allows the learner to practice retrieving information from storage, has been found to improve performance more than the opportunity to study the material again. Such testing enhances both initial acquisition and long-term retention (Halff, 1977; Kinstch, 1968; Roediger & Karpicke, 2006a, 2006b; Runquist, 1983; Underwood, 1964). A key aspect of retrieval is the overlap between cues present at study and at test (the encoding specificity principle; Tulving & Thomson, 1973). For example, if variables within algebra problems are always stated in textbooks using *x* and *y*, but are then tested using other labels such as *j* and *k*, test performance will be reduced. Once information is stored, students must learn to recognize sometimes subtle cues (and to ignore irrelevant cues) in order to draw on the right knowledge in the right context.

Conceptual understanding promotes transfer of learning to new problems and better long-term retention.

Research has demonstrated that factors that enhance initial acquisition are not necessarily the same as those that maximize long-term retention (i.e., that minimize forgetting). For example, material that is too easy to understand can promote initial acquisition or learning, but it leads to lower retention than material that is harder to understand initially (e.g., Bjork, 1994). Challenging material causes the learner to exert more attentional effort and to actively process information, leading to superior retention. Similarly, transfer of learning is promoted by deeper conceptual understanding of learned material. Although this phenomenon was demonstrated in the early work of Gestalt psychologists (e.g., Wertheimer, 1959), it has since been verified repeatedly (for illustrative empirical studies on transfer and reviews of such studies, see Bassok & Holyoak, 1989; Reed, 1993; Wolfe, Reyna, & Brainerd, 2005). Transfer of learning refers to the ability to correctly apply one's learning beyond the exact examples studied to superficially similar problems (near transfer) or to superficially dissimilar problems (far transfer). Surprisingly, instruction using more abstract representations has been shown in some instances to benefit learning and transfer more than concrete examples (e.g., physical representations, such as manipulatives) (e.g., Sloutsky, Kaminski, & Heckler, 2005; Uttal, 2003). Thus, the cognitive processes that facilitate rote retention (e.g., of over-learned arithmetic facts), such as repeated practice, can differ from the processes that facilitate transfer and long-term retention, such as conceptual understanding. People's knowledge of how such factors affect cognition and thus how they can better monitor and control their learning—metacognition—also has been the subject of extensive research (e.g., Koriat & Goldsmith, 1996; Metcalfe, 2002; Nelson & Narens, 1990; Reder, 1987). Research has shown that there is much room for improvement in students' metacognitive judgments because they rely on misleading assumptions about their learning (e.g., using misleading cues such as retrieval fluency and familiarity, which are not perfectly correlated with strength of learning; see Benjamin, Bjork, & Schwartz, 1998).

2. Mental Representations

Although laws of memory apply to different kinds of content, just as the laws of physics apply to different kinds of objects, memories take different forms depending on their content. Declarative knowledge is explicit memory for specific events and information; procedural knowledge refers to implicit memory for cognitive (e.g., algorithms) and motor sequences and skills; and conceptual knowledge refers to general knowledge and understanding stored in long-term memory (see Hunt & Ellis, 2004, for further distinctions). Declarative, procedural, and conceptual knowledge seem to be represented in different ways in the brain (e.g., Schacter, Wagner, & Buckner, 2000). For example, a patient with brain damage can have amnesia for declarative knowledge, failing to remember his name and not recognizing his loved ones, but retain procedural skills such as piano playing or mathematical computation.

Using the mental number line: Counting supports learning addition.

One important mental representation in mathematics learning is the number line. The acquisition of counting, which forms the basis for arithmetic learning, is eventually mapped for successful learners onto an internal number line. One key use of this internal representation is for understanding the meaning of basic arithmetic operations. Counting, conceived as proceeding in steps up and down such an internal linear representation, provides a transition to learning arithmetic. Addition and subtraction can then be analogously conceived as proceeding in steps up and down that internal number line. The mental number line also plays a role in estimating the magnitudes of numbers in situations in which precise calculation is impossible (Siegler & Booth, 2005). For example, providing low-income children who attend Head Start centers an hour of practice playing numerical board games using consecutively numbered, linearly arrayed squares, dramatically improves their understanding of the mental number line and their estimation of numerical magnitudes (Siegler & Ramani, in press). In addition, explicitly instructing children from low-income backgrounds in number line skills using linearly organized board games (i.e., practicing with increments of only one step up or down) improves their procedural and conceptual arithmetic skills more than a year after this instruction, demonstrating both near and far transfer (e.g., Griffin, Case, & Siegler, 1994).

Mental models guide the acquisition of cognitive skills and the development of strategies, improving mathematics performance.

Mental models are ways of internally representing problems, often in the form of specific images. A mental number line is an example of a mental model (Case & Okamoto, 1996). The application of these mental models can be illustrated by thinking about fractions. A physical model, which can be internalized with practice and be used to think about fractions, is the familiar pie diagram; for example, $\frac{3}{4}$ might be represented by thinking of a pie cut into four equal pieces, three of which are highlighted. Analogous physical models can be constructed using folded paper, chips, and other physical objects. An alternative mental model for thinking about $\frac{3}{4}$ would be imagining two children who wanted to build a tower from their collection of Legos®. If one child supplied three of the Legos® and the other child a single Lego®, the first child would have supplied $\frac{3}{4}$ of the Legos®. Similarly, the ways in which children physically, and then mentally, represent the relation between divisors and quotients influence their skill at solving simple division problems (Squire & Bryant, 2003). For example, mentally picturing two sets of six objects helps children solve such problems as $\frac{12}{2}=6$. As Halford (1993) has pointed out, appropriate mental models—a mental picture of the concepts underlying the problem—provide a framework for problem solving that improves performance.

Verbatim memories of problem details are encoded separately from gist memories of the meaning of problem information; thinking in terms of gist often produces superior reasoning.

Experimentation on the relations between memory and reasoning has addressed how memory controls reasoning, why some forms of such reasoning are easier and more accurate than others, and what sorts of instruction benefit reasoning (e.g., Reyna & Brainerd, 1991). The most basic finding from this research is that there are two main types of memory, namely verbatim memory and gist memory (Brainerd & Reyna, 1993; Reyna & Brainerd, 1993). The importance of this distinction can be illustrated by a study of children's memory for numerical information within stories (Brainerd & Gordon, 1994). The verbatim level consisted of the actual numbers within the stories; the gist level consisted of various numerical relations, such as "more," "less," "most," "least," and "between." When told, for example, that Farmer Brown owned 3 dogs, 5 sheep, 7 chickens, 9 horses, and 11 cows, children accurately remembered that he had fewer dogs and more cows than any other animal. They were considerably less accurate in remembering how many of each type of animal he had (Brainerd & Gordon).

In some contexts, less precise gist memories are more important to performance than verbatim memories of the actual numbers and operations (Reyna & Brainerd, 1993). Many other features of these problems can be answered accurately and effortlessly by one or another type of gist knowledge. Estimation provides one such context. The late physicist Richard Feynman, for instance, argued that solving complex problems depends on seeing where solutions must lie–getting the gist of problems–more than on verbatim calculation (Leighton, 2006). Thus, being able to estimate that 74×97 must equal a little less than 7400 and thus cannot equal either 718 or 71,780, can help children recognize that they have made a mistake if they obtain either of those answers.

Psychological theory explains why ratio concepts, such as fractions, probabilities, and proportions, are especially difficult; this theory also provides straightforward ways to improve performance.

The importance of memory for gist extends to more complex mathematical relations as well, such as ratios, fractions, and probabilities (e.g., Hecht, Close, & Santisi, 2003; Reyna, 2004). For instance, in probability judgments, making accurate forecasts about the relative likelihood of occurrence of a set of events is usually quite difficult (Reyna & Brainerd, 1994), but it becomes much easier when gists are used (e.g., expressing the probabilities of the individual events in terms such as *more than half* or *less than half* (e.g., Brainerd & Reyna, 1995; Spinillo & Bryant, 1991). Based on these findings, interventions have been designed and tested with students ranging from young children to medical residents, and found to virtually eliminate common errors (e.g., Brainerd & Reyna, 1990, 1995; Lloyd & Reyna, 2001).

One reason why mathematics is so difficult to master is that it requires the accumulation of considerable verbatim knowledge, which often requires more effort to learn than the gist. Nonetheless, verbatim recall of facts, concepts, postulates, and other knowledge is an essential feature of a strong mathematics education, despite its often requiring a great deal of time, effort, and practice. Gist memory, or less precise, conceptual memory traces, has broad implications for learning because it is the form of memory that is typically relied on in reasoning. In short, a strong mathematics background requires a combination of gist and verbatim representations, with the importance of one or the other dependent on the goal at hand.

B. Social, Motivational, and Affective Influences on Learning

Research has shown that motivation enhances learning—and that some kinds of motivation are more effective than others. Motivation to persevere when intrinsic enjoyment is low should be distinguished from making learning enjoyable; the former may be especially important in sustaining the effortful learning needed to master difficult content. Perceived utility and willingness to engage in difficult learning is influenced by beliefs about the contributions of ability versus effort in learning, self efficacy (i.e., the belief that one has the specific skills needed to be successful, which differs from self esteem), and an array of other intrapersonal and social factors. In the following sections, the Task Group reviews major theories and findings related to these factors and how they influence learning mathematics and student achievement. Theoretical frameworks are reviewed that focus on learning goals, motivation to learn, attributions and beliefs about learning outcomes, mathematics anxiety, and sociocultural considerations. For more comprehensive coverage of theories and empirical data in this area, see Ames and Archer (1988), Barron and Harackiewicz (2001), Eccles and Wigfield (2002), Grant and Dweck (2003), Meece, Anderman, and Anderman (2006), Bandura (1993), Ellis, Varner, and Becker (1993), and Rieber and Carton (1987).

1. Goals and Beliefs About Learning

Children's goals and beliefs about learning are related to their mathematics performance. Children who adopt mastery-oriented goals show better long-term academic development in mathematics than do their peers whose main goals are to get good grades or outperform other children. They also are more likely to pursue difficult academic tasks. Students who believe that learning mathematics is strongly related to innate ability show less persistence on complex tasks than peers who believe that effort is more important. Experimental studies have demonstrated that children's beliefs about the relative importance of effort and ability can be changed, and that increased emphasis on the importance of effort is related to improved mathematics grades. The Task Group recommends extension of these types of studies.

Children's learning goals vary along several dimensions. One important dimension is whether the goals emphasize accomplishing a task or enhancing one's ego (Nicholls, 1984). Another important distinction is whether the goals emphasize mastery of the material or outperforming other students (Ames, 1990; Dweck & Leggett, 1988). Yet another important distinction is between performance approach goals (i.e., striving to surpass the performance of others) and performance avoidance goals (i.e., trying to avoid looking less knowledgeable or inferior) (Elliott & Harackiwiez, 1996; Midgley, Kaplan, Middleton, Maehr, & Urban, 1998).

Mastery and performance goals have received the most empirical attention. When pursuing mastery goals, students tend to choose tasks that are challenging and concern themselves more with their own progress than with outperforming peers. Mastery goal orientation should not be confused with Bloom's (1971; 1981) notion of *mastery learning.* The latter refers to an instructional approach whereby teachers lead students through a discrete set of stair step learning units with the progression predicated on pre-established criteria for proficiency at each step. When pursuing performance goals, students focus on outperforming others and thus prefer and seek tasks in which they are already competent (i.e., "easy," less challenging tasks). In the face of failure or incorrect performance, mastery-oriented students are likely to attribute the result to their own lack of effort or insufficient opportunities for mastery rather than to lack of ability; children who emphasize their lack of effort or opportunities are more likely to redouble their levels of effort when faced with later challenging problems (Ames, 1992; Ames & Archer, 1988). In contrast, when faced with demanding problems, performance-oriented students often conclude that they do not have the ability to do well in the domain, and thus tend to avoid challenging material when they begin to experience failure.

With respect to math outcomes, Wolters (2004) for example has shown that among middle school students a mastery orientation was positively related to engagement in learning and math grades, but this was not the case for a performance goal orientation. Elsewhere, Linnenbrink (2005) found that among fifth- and sixth-grade students working on a five-week math unit on statistics and graphing, those pre-tested as high in mastery orientation reported greater self-efficacy, personal interest in math, and more adaptive help seeking. These students performed significantly better on the math unit exam than those who were pretested as high in performance goal orientation.

In high school, children who tend to have mastery goals also tend to be high in self-efficacy. Such children also tend to obtain high grades in mathematics courses (Gutman, 2006). Moreover, parents' mastery goals are associated with better grades in mathematics courses by their children. Graham and Golan (1991) have shown that instructions that prompt a mastery orientation lead to higher academic outcomes than do performance-based instructions, when the task calls for deep processing of complex concepts.

Ames (1992) reviewed several types of academic contexts likely to foster mastery goal orientations in school. These include contexts that 1) provide meaningful reasons (e.g., personal relevance) for task engagement or developing understanding of content; 2) promote high interest and intermediate challenge; 3) emphasize gradual skill improvement; and 4) promote novelty, variety, and diversity.

Beliefs about learning and intelligence also influence mathematics performance. When faced with challenging problems, children who believe that intelligence is in large part created by their efforts to learn tend to do better than children who believe that intelligence is a fixed quality that cannot be changed (Dweck, 1999). Looking more specifically at mathematics achievement, Dweck and her colleagues recently showed that students who viewed their intelligence as a fixed trait fared more poorly across the transition to junior high than did their peers who believed that their intelligence was malleable and could be developed (Blackwell, Trzesniewski, & Dweck, 2007). Although both groups began junior high with equivalent mathematics achievement, their mathematics grades diverged by the end of the first semester of seventh grade and continued to move apart over the next two years. The superior performance of students who believed that intelligence is malleable was mediated by their greater emphasis on learning, their greater belief in the importance of effort, and their more mastery-oriented reactions to setbacks.

Blackwell et al. (2007) also conducted an intervention with a different group of seventh-graders with declining mathematics grades. Both the experimental and control groups received an eight-session workshop that taught them useful study skills. However, for the experimental group, several of the sessions also taught them a malleable theory of intelligence. (These sessions began with the article *You Can Grow Your Intelligence*, which likened the brain to a muscle; the article also described how neurons in the brain were transformed through learning. Students then learned how to apply this idea to their schoolwork.) Whereas the control group continued its downward grade trajectory, the experimental group showed a significant rebound in mathematics grades. Moreover, teachers (blind to condition) singled out three times as many students in the experimental group as having shown marked changed in motivation to learn mathematics.

2. Intrinsic and Extrinsic Motivation

Young children's intrinsic motivation to learn (i.e., desire to learn for its own sake) is positively correlated with academic outcomes in mathematics and other domains. However, intrinsic motivation declines across grades, especially in mathematics and the sciences, as material becomes increasingly complex and as instructional formats change. The complexity of the material being learned reflects demands of the modern workforce that may not be fully reconcilable with intrinsic motivation—the latter should not be used as the sole gauge of what is appropriate academic content. At the same time, correlational evidence suggests that the educational environment can influence students' intrinsic motivation to learn in later grades. The Task Group recommends studies that experimentally assess the implications of these correlational results, that is, studies aimed at more fully understanding the relation between intrinsic motivation and mathematics learning.

Intrinsic motivation to learn is the desire to learn for no reason other than the sheer enjoyment, challenge, pleasure, or interest of the activity (Berlyne, 1960; Hunt, 1965; Lepper et al., 2005; Walker, 1980). It is often contrasted to extrinsic motivation, in which the motivation to learn is to gain an external reward, such as the approval of parents and others, or the respect of peers. Thus, intrinsic motivation is related to mastery goals and extrinsic motivation to performance goals.

Several studies have shown that learning and academic achievement are positively correlated with intrinsic motivation (Lepper et al., 2005). For example, in a recent study by Lepper et al., it was found that across a sample of third- to eighth-grade students, an intrinsic motivation orientation was positively correlated with mathematics grade point average (GPA) and with performance on a mathematics achievement test, whereas an extrinsic motivation orientation was negatively correlated to these outcomes. However, there is evidence that intrinsic motivation declines as children progress through school and as material becomes more challenging (Gottfried et al., 2001). For example, Gottfried et al. found that from the ages of 9 to 16 years (although there was a slight increase for 17-year-olds), children's overall intrinsic motivation for academic learning declined, with particularly marked decreases in mathematics and the sciences. Findings like these have led some to propose that such a reduction in intrinsic motivation over time may compromise engagement in mathematics learning in the upper grades.

3. Attributions

Student beliefs about the causes of their success and failure have been repeatedly linked to their engaging and persisting in learning activities. *Self-efficacy*—a central concept in attributional theories—has emerged as a significant correlate of academic outcomes. However, the cause-effect relation between self-efficacy and mathematics learning remains to be fully determined, as does the relative importance of self-efficacy versus ability in moderating these outcomes. The Task Group recommends experimental and longitudinal studies that assess the relative contributions of these factors to mathematics learning.

Students can attribute their successes and failures to ability in (e.g., I'm just good/bad at) mathematics, effort (e.g., I worked/did-not-work hard enough), luck, or powerful people (e.g., the teacher loves/hates me). These attributions influence students' subsequent engagement in learning.

Self-efficacy can be defined as beliefs about one's ability to succeed at difficult tasks (Bandura, 1997). Mathematics self-efficacy moderates the effect of ability on performance. In other words, ability is important for mathematics learning but is not sufficient; self-efficacy or confidence in one's mathematics ability is also crucial for high levels of achievement. At times, self-efficacy is more influential than general mental ability in predicting high school mathematics performance (Stevens, Olivarez, & Hamman, 2006), although other studies suggest that ability may be more important than motivational influences in general (Gagné & St Père, 2002). Studies that simultaneously assess ability, prior content knowledge, motivation, and efficacy beliefs are needed to more firmly establish the relative contributions of these factors to mathematics learning and achievement.

4. Self-Regulation

Self-regulation is a mix of motivational and cognitive processes. It includes setting goals, planning, monitoring, evaluating, making necessary adjustments in one's own learning process, and choosing appropriate strategies. Self-regulation has emerged as a significant influence on some aspects of mathematics learning. Although the concept appears promising, research is needed to establish the relation for a wider range of mathematics knowledge and skills.

The concept of self-regulation includes aspects of both motivation and cognition. Among the processes that are associated with self-regulation are monitoring one's own actions, evaluating one's success, and reacting to discrepancies between one's outcomes and one's goals.

Fuchs et al. (2003), deploying an experimental design, provided evidence for the effectiveness of self-regulated learning strategies in enhancing mathematics problem solving. They focused on elementary school students' application of knowledge, skills, and strategies to novel mathematics problems. In the self-regulation condition, students were prompted to engage in self-regulated strategies by being required to check their answers, set goals of improvement, and chart their daily progress. These efforts to improve self-regulation improved the children's mathematics learning.

Another form of self-regulation involves choosing which strategy to use to solve problems. Children know and use a variety of strategies for solving mathematics problems. For example, to solve simple addition problems, elementary school children sometimes count from one, count from the larger addend, decompose the problem into two simpler problems, or retrieve the answer from memory. Individual differences in children's arithmetic strategy choices reflect differences in knowledge of answers to problems and also in degree of perfectionism. One group of children—labeled *good students* by Siegler (1988a)—has high knowledge and usually retrieves answers to problems. Another group of children—labeled *perfectionists*—has comparable knowledge but prefers to double-check their retrieved answers via counting strategies. A third group of children—labeled *not-so-good students*—has poor knowledge and often guesses at the answer. Perfectionists are toward the high end of self-regulation and not-so-good students are toward the low end. Children who fall into the *not-so-good student* group are more likely than the others to subsequently be labeled as mathematics disabled or not promoted to the next grade (Kerkman & Siegler, 1993).

5. Mathematics Anxiety

Anxiety about mathematics performance is related to low mathematics grades, failure to enroll in advanced mathematics courses, and poor scores on standardized tests of mathematics achievement. It also may be related to failure to graduate from high school. At present, however, little is known about its onset or the factors responsible for it. Potential risk factors include low mathematics aptitude, low working memory capacity, vulnerability to public embarrassment, and negative teacher and parent attitudes. The Task Group recommends research that assesses these potential risk factors; it also recommends development of promising interventions for reducing debilitating mathematics anxiety.

Mathematics anxiety refers to an emotional reaction, ranging from mild apprehension up through genuine fear or dread, in academic and everyday situations that deal with numbers, for instance taking a standardized achievement test, or figuring out a restaurant bill or change. Considerable research was done in the 1970s and 1980s on the relationships between mathematics anxiety, personality characteristics, and aspects of academic achievement, yielding a rather bleak picture (see Hembree, 1990). In brief, individuals with high mathematics anxiety perform poorly in school math, earn poor grades in math classes, take fewer elective mathematics courses in high school and college, and avoid college majors that rely on mathematics (e.g., mathematics, science, and engineering fields). There is a tendency, although weak, for women to exhibit higher levels of mathematics anxiety.

Mathematics anxiety research beginning in the 1990s has taken a more process-oriented approach to understanding the phenomenon, asking, "What are the cognitive consequences of mathematics anxiety?" The major discovery from this work is that during mathematics performance, those with mathematics anxiety focus many of their limited working memory resources on managing their anxiety reaction rather than on the execution of the mathematics procedures and processes necessary for successful performance (Ashcraft & Kirk, 2001). Difficult mathematics problems require considerable working memory resources for keeping track of intermediate solutions, retrieving facts and procedures, and so forth (LeFevre, DeStefano, Coleman, & Shanahan, 2005). These resources are limited to begin with, and thus are seriously compromised when the individual devotes substantial portions of them to the worry and negative thoughts associated with mathematics anxiety. This research aligns with other contemporary research on factors such as stress and stereotype threat (e.g., Beilock, Rydell, & McConnell, 2007), and their negative effects on high-stakes testing outcomes. As such, mathematics anxiety may be yet another factor leading to poorer-than-expected performance on proficiency and standardized tests of mathematics achievement (Ashcraft, Krause, & Hopko, 2007).

Interestingly, the research shows that highly math anxious individuals who undergo successful therapeutic interventions, especially cognitive-behavioral therapies, then show math achievement scores approaching the normal range. This suggests that their original math learning was not as deficient as originally believed, but instead that their math achievement scores had been depressed by their math anxiety during achievement testing itself. More precisely, in a meta-analysis, Hembree (1990) found that reductions in mathematics anxiety can result in significant (~ .5 standard deviations) improvements in mathematical test scores and in grade point average in mathematics courses. However, not all treatments are equally effective. Traditional individual or group counseling techniques appear to be relatively ineffective in reducing mathematics anxiety or improving mathematical performance. Similarly, changes in classroom mathematics curriculum, such as providing calculators or microcomputers to aid in problem solving, appear to be largely ineffective in reducing mathematics anxiety. A promising exception appears to be curricular changes that increase student's mathematical competence. Hutton and Levitt (1987) improved feelings of competence, or self-efficacy, by focusing on the relation between mathematical performance and good study habits, and by improving basic skills. These goals were achieved, in the context of an algebra class, through the use of a specially designed textbook. For each algebraic topic, the textbook presented a review of the basic arithmetic skills needed to solve the associated algebra problems. These basic skills were then practiced. Lectures and the text material were synchronized, such that the basic foundation of each lecture was presented as "skeletal notes" in the textbook. This feature was designed to improve students' note taking, and to focus them on essential features of the lecture. The intervention resulted in significant reductions in mathematics anxiety and was associated with algebraic skills that did not differ from those of children without mathematics anxiety; as noted, these two groups typically differ and thus no difference suggests a gain on the part of the students with mathematics anxiety.

Cognitive therapies that focus on the worry component of mathematics anxiety are also promising (Ellis et al., 1993). They are associated with moderate declines (~ .5 standard deviations) in mathematics anxiety, as well as modest (~ .3 standard deviations) improvements in mathematical performance (Hembree, 1990). These therapies focus on reducing the frequency of intrusive thoughts during mathematical activities, and on changing the individual's attributions about their performance. Poor performance that is attributed to a lack of ability will often result in avoidance of and lack of persistence on difficult mathematical tasks, as noted above. Changing attributions so that they focus on more controllable factors, such as preparation and hard work, often results in more persistent task-related behaviors and improvements in performance (Dweck, 1975). Ellis et al. argued that the treatment of mathematics anxiety should include building the student's basic competencies, knowledge, and skills. Increasing the competencies of students appears to reduce both the emotionality and worry components of mathematics anxiety, in addition to being an important goal in and of itself (e.g., Randhawa, Beamer, & Lundberg, 1993).

6. Vygotsky’s Sociocultural Perspective

The sociocultural perspective of Vygotsky has been influential in education and characterizes learning as a social induction process through which learners become increasingly able to function independently through the guidance of more knowledgeable peers and adults. Aspects of this approach may add to the understanding of mathematics learning. However, a shortage of controlled experiments makes the usefulness of this approach for improving mathematics learning difficult to evaluate, and thus its utility in mathematics classrooms and mathematics curricula remains to be scientifically tested.

Vygotsky’s sociocultural perspective posits that knowledge is first acquired in the interaction between the learner and other people, and that the knowledge is later internalized so that the learner can act on the knowledge in increasingly independent ways (see Rieber & Carton, 1987). Through this process of internalization, learners gain the knowledge and skills necessary for adequate functioning in their society (Wertsch, 1985).

From a sociocultural perspective, the most useful unit of analysis is not the child per se, but rather the child performing an activity in context. From this analytical frame, it is undesirable if not impossible to separate who the child is, from what the child does, from where the child does it. This framework calls for distinctive ways of construing learning processes, where notions such as zone of proximal development (the gap between what a learner can achieve independently and what the learner can achieve under the guidance of others), scaffolding (support from other people for problem solving activity), intersubjectivity (establishing a shared focus of attention), and apprenticeship (learning from more knowledgeable others) hold sway. Knowledge is not viewed as residing inside the child’s head but rather as being distributed across the collectively held understandings of groups of people interacting with books, computers, worksheets, and other cultural tools. Knowledge acquisition is viewed as arising from participation in successful practices within a community of practice.

The evidence presented by socioculturalist scholars for their educational claims is not typically in the form of experiments or systematic empirical studies. Instead, detailed descriptions of people's everyday experiences in various contexts are provided and used to argue for particular educational arrangements. For example, Gonzales, Andrade, Civil, and Moll (2001) examined the informal mathematics knowledge shared among a group of community participants that was embedded in the everyday practical activity of sewing. It was demonstrated that through guided group discussions, the participants were able to link their informal mathematics knowledge to formal mathematics concepts and processes such as measurement, symmetry, geometric shapes, and angles, and to addition, subtraction, percentages, and proportions. Likewise, Kahn and Civil (2001) described how fourth- and fifth-grade children gained insight into domains like measurement (in this case the relation between area and perimeter) and graphing in the course of participating in a class-wide gardening project. The students' insights were said to be mediated by factors akin to guided participation and scaffolding. In addition, Gauvain (1993) has written extensively concerning how spatial thinking grows through participation in everyday practices. Although these descriptions are intriguing, lack of experimental studies makes it impossible to evaluate at present whether widespread adoption of such approaches would help or hinder mathematics learning and, if helpful, what specific areas of mathematics.

All told, concepts and processes such as zone of proximal development, scaffolding, and guided participation reflect core aspects of Vygotsky's sociocultural theory as related to instruction, and may hold important heuristic value. Yet to date, they have eluded measurement specificity and proven difficult to reliably quantify. Their ultimate utility in promoting effective evidence-based mathematics learning must await such specification and experimental validation.

7. Conclusions and Recommendations

Children's goals and beliefs about learning are related to their mathematics performance. Children who adopt mastery-oriented goals show better long-term academic development in mathematics than do their peers whose main goals are to get good grades or outperform other children. They also are more likely to pursue difficult academic tasks. Students who believe that learning mathematics is strongly related to innate ability show less persistence on complex tasks than peers who believe that effort is more important. Experimental studies have demonstrated that children's beliefs about the relative importance of effort and ability can be changed, and that increased emphasis on the importance of effort is related to improved mathematics grades. The Task Group recommends extension of these types of studies.

Young children's intrinsic motivation to learn (desire to learn for its own sake) is positively correlated with academic outcomes in mathematics and other domains. However, intrinsic motivation declines across grades, especially in mathematics and the sciences, as material becomes increasingly complex and as instructional formats change. The complexity of the material being learned reflects demands of a modern workforce that may not be fully reconcilable with intrinsic motivation. The latter should not be used as the sole gauge of what is appropriate academic content. At the same time, correlational evidence suggests that the educational environment can influence students' intrinsic motivation to learn in later grades.

The Task Group recommends studies that experimentally assess the implications of these correlational results, that is, studies aimed at more fully understanding the relation between intrinsic motivation and mathematics learning.

Student beliefs about the causes of their success and failure have been repeatedly linked to their engaging and persisting in learning activities. Self-efficacy has emerged as a significant correlate of academic outcomes. However, the cause and effect relation between self-efficacy and mathematics learning remains to be fully determined, as does the relative importance of self-efficacy versus ability in moderating these outcomes. The Task Group recommends experimental and longitudinal studies that assess the relative contributions of these factors to mathematics learning.

Self-regulation is a mix of motivational and cognitive processes. It includes setting goals, planning, monitoring, evaluating, and making necessary adjustments in one's own learning process; and choosing appropriate strategies. Self-regulation has emerged as a significant influence on some aspects of mathematics learning. Although the concept appears promising, research is needed to establish the relation for a wider range of mathematics knowledge and skills.

Anxiety about mathematics performance is related to low mathematics grades, failure to enroll in advanced mathematics courses, and poor scores on standardized tests of mathematics achievement. It also may be related to failure to graduate from high school. At present, however, little is known about its onset or the factors responsible for it. Potential risk factors include low mathematics aptitude, low working memory capacity, vulnerability to public embarrassment, and negative teacher and parent attitudes. The Task Group recommends research that assesses these potential risk factors; it also recommends development of promising interventions for reducing debilitating mathematics anxiety.

The socio-cultural perspective of Vygotsky has been influential in education and places learning as a social induction process through which learners become increasingly able to function independently through the guidance of more knowledgeable peers and adults. Aspects of this approach may add to our understanding of mathematics learning. However, a shortage of controlled experiments makes the usefulness of this approach for improving mathematics learning difficult to evaluate, and thus its utility in mathematics classrooms and mathematics curricula remains to be scientifically tested.

Despite all that has been learned about the relation between these social/motivational goal orientations, attitudes, and beliefs and mathematics grades and achievement, too little is known about whether these influences reflect stable dispositions of students, or reflect teacher or peer influences in certain learning settings (Meece et al., 2006). The question of whether students in classroom settings have multiple goals or beliefs related to academic goals remains to be fully answered (Harackiewicz, Barron, Pintrich, Elliott, & Thrash, 2002; Brophy, 2005). In any case, the Blackwell et al. (2007) investigation, among others, indicates that beliefs about mathematics learning can be adaptively changed through targeted interventions. The Task Group recommends development and elaboration of these forms of intervention and assessment of ease with which they can be implemented by classroom teachers.

C. What Children Bring to School

Mathematical development begins in infancy, long before children go to school, and continues through the toddler and preschool years. The amount of mathematical knowledge that children bring with them when they begin school has large, long-term consequences for their further learning in this area. Thus, it is important to understand how mathematical knowledge typically develops before children start school, how children from different backgrounds and cultures vary in this knowledge, and how early mathematical learning can be improved.

1. Roots of Numerical Understanding

Mathematical development starts in infancy. Even infants between 1 and 4 months of age form nonverbal representations of the number of objects in very small sets. For example, when repeatedly shown two objects—two dots, two stars, two triangles—infants of this age gradually lose interest, and look for shorter and shorter times. However, when the number of objects is switched to one or three, they look longer, thus indicating that they noted the difference between sets with two objects and sets with one or three (e.g., Antell & Keating, 1983). This evidence suggests that sensitivity to number is innate to human beings.

Infants' surprising early numerical ability extends to a kind of nonverbal arithmetic. When 5-month-olds see a doll hidden behind a screen, and then see a second doll also placed behind the screen, they seem surprised and look longer when, through a trick, lifting the screen reveals one or three objects rather than two (Wynn, 1992). Presumably, they expected 1 + 1 to equal 2, and were surprised when it did not. A similar nonverbal form of subtraction is evident at the same age; when two objects are placed behind a screen and one object is removed, 5-month-olds look longer when lifting the screen reveals two objects rather than one. Whether these competencies are inherently numerical or not is debated (Cohen & Marks, 2002), but the basic finding has been replicated many times (e.g., Kobayashi, Hiraki, Mugitani, & Hasegawa, 2004).

In addition to these relatively precise nonverbal representations of very small numbers of objects, infants also display rudimentary estimation skills that allow them to discriminate between more and less numerous sets when the more numerous set has at least twice as many objects as the less numerous one. For example, they discriminate between sets of 16 and 8 objects, seeming to know that the set of 16 has more objects (Brannon, 2002; Xu & Spelke, 2000). These remarkable early nonverbal numerical abilities provide the foundation for learning about the verbal number system, including the number words, counting, numerical comparison, and more formal addition and subtraction.

2. Mathematical Understanding in the Preschool Period

a. Acquisition of Number Words and Counting

Many 2-year-olds in the United States know some number words, by age 3 or 4 years of age, many children can count (in the sense of accurately reciting the number words) from 1 to 10, and by the time they enter school, many children can count to 100 (U.S. Department of Education, NCES, 2001; Fuson, 1988; Miller, Smith, Zhu, & Zhang, 1995; Siegler & Robinson, 1982). Children also begin to learn to count objects at age 2 and a half or 3 years. At first, however, they acquire the superficial form of counting objects without understanding its purpose. Thus, when presented five objects and asked, “How many are there,” many 2 and a half- and 3-year-olds will count to five but not answer the question. When again asked, “So how many are there,” they will either again count to 5 without saying “There are 5” or say, “I don’t know” (Le Corre, Van de Walle, Brannon, & Carey, 2006; Schaeffer, Eggleston, & Scott, 1974). This difficulty in connecting procedures with their goals and underlying principles is a persistent problem at all ages.

By age 4 or 5, when most children have had a reasonable amount of counting experience, they also come to understand the principles underlying the counting procedure: that each object must be labeled by one and only one number word, that counting requires the numbers to be recited in a constant order, and that the final word in the count indicates the number of objects in the set that has been counted (Gelman & Gallistel, 1978). Understanding these principles allows children to count in flexible ways, including, for example, starting the counting in the middle or at the right end of a row of objects if asked to do so, and to reject counts that skip an object or count an object twice (Frye, Braisby, Love, Maroudas, & Nicholls, 1989).

b. Ordering Numbers

Although it may seem surprising, being able to count from 1 to10 does not guarantee knowledge of the relative magnitudes of the numbers. Many 3- and 4-year-olds can count flawlessly to 10, but do not know that 8 is larger than 7 or that 7 is larger than 6 (Siegler & Robinson, 1982). By the time they enter kindergarten, however, most children know the relative magnitudes of numbers in this 1 to 10 range very well. Most children from middle-income backgrounds also have some knowledge of the order of numbers up to 100 when they enter school. When kindergartners are presented a number line with 0 at one end and 100 at the other end and asked to estimate the locations of numbers between 0 and 100 on the line, their estimates reflect the ordering of the numbers quite well, though not perfectly (Siegler & Booth, 2004).

c. Arithmetic

As with other numerical skills, children first show competence on addition problems with one to three objects. For example, if the experimenter asks a child to put three balls in an opaque tube, removes one of them, and then asks the child to remove the remaining balls, most 2 and a half- and 3-year-olds will reach into the tube exactly twice to pull out the remaining balls (Starkey, 1992). Children of this age usually fail, however, if the experimenter has the child put in four balls, removes one, and then makes the same request. The difficulty appears to involve limited ability to represent numbers precisely. Very young children show much greater ability to represent numbers approximately (Huttenlocher, Jordan, & Levine, 1994).

Most 4- and 5-year-olds can retrieve from memory the answers to at least a few basic addition and subtraction facts, such as 2 + 2 = 4, and also know a variety of other procedures for solving simple addition and subtraction problems. These include using fingers or objects to represent each addend and then counting them from one, representing the problem with fingers or objects and then recognizing how many of these are present, and counting from one without using objects or putting up fingers (Siegler & Shrager, 1984).

Even in the preschool period, children use these strategies in surprisingly adaptive ways. The harder the problem, the more likely 4- to 6-year-olds are to rely on counting or finger recognition strategies (Siegler & Shrager, 1984). This approach allows children to solve the easiest problems, such as 2 + 2, by using the fast approach of retrieving the answer from memory, and to solve problems that are too difficult to retrieve from memory via the slower but accurate alternative approaches of counting fingers or objects. The use of counting strategies on hard problems helps children generate the correct answer on those problems, which improves their likelihood of remembering it when the problem is presented later (Siegler, 1996).

Preschoolers also show some understanding of arithmetic concepts. For example, many 4- and 5-year-olds recognize that addition and subtraction are inverse operations. Thus, if presented problems of the form $A + B - B$, many preschoolers quickly answer "*A*" (Rasmussen, Ho, & Bisanz, 2003).

d. Measurement

During the preschool period, children acquire measurement strategies that are greatly oversimplified but that nonetheless reflect basic understanding of relations of equality, more than, and less than (Geary, 1994). When asked to divide up candies among friends, 2- and 3-year-olds typically give everyone some, without regard for whether each child receives the same number. In contrast, most 5-year-olds maintain exact numerical equivalence by using a "one for you, one for me, one for him, one for her" approach. They take this counting strategy too far, however, and use it even if one pile includes more large pieces of food than the other (Miller, 1984). Even 7- and 9-year-olds often use this strategy. Learning to restrict procedures to situations where they fit is another persistent challenge in mathematics learning.

e. Geometric Knowledge

During the preschool period, children also acquire rudimentary geometric knowledge. The large majority of 4- and 5-year-olds accurately identify circles and squares, and many also can identify triangles; by age 5, most also discriminate between squares and rectangles, and can describe some geometric attributes of those shapes (Clements, Swaminathan, Hannibal, & Sarama, 1999). Most children of these ages also have some skill in judging whether these basic figures are congruent; they usually adopt a strategy of comparing corresponding edges to do so (Clements, 2004).

Children's spatial knowledge also develops considerably in the preschool years. Most 5-year-olds can represent a location in terms of multiple landmarks, and from 5 to 7 years of age develop in their ability to maintain locations in challenging circumstances such as open areas (Newcombe & Huttenlocher, 2000). They use, implicitly, two coordinates in remembering direction, either polar or Cartesian, and can use simple external representation systems such as maps (Clements & Sarama, 2007b).

f. Number Sense

Through engaging in a variety of numerical activities, preschoolers, at least those from middle-income backgrounds, begin to develop number sense (Berch, 2005; Case & Sowder, 1990; Gersten & Chard, 1999; Jordan, Kaplan, Olah, & Locuniak, 2006). Number sense is the ability to approximate numerical magnitudes. The approximations can involve the numerical magnitude of specific dimensions of objects, events, or sets (e.g., "About how long is this line?" "About how many times have you been to New York?" "About how many people were at the play?"), or they can involve the results of numerical operations ("About how much is 24×94?"). Number line estimation tasks have proved particularly useful for investigating number sense. Such tasks involve presenting lines with a number at each end (e.g., 0 and 10) and no other numbers or marks in-between, and asking participants to locate a third number on the line (e. g., "Where does 7 go?").

Performance on this and other tasks used to measure number sense show that even before children enter school, children from middle-income backgrounds are developing a good sense of numerical magnitudes, whereas children from lower-income backgrounds have little sense of the numbers' magnitudes (Ramani & Siegler, 2008). This difference is important, because early number sense predicts subsequent ability to learn arithmetic in elementary school, above and beyond other important characteristics such as working memory (Locuniak & Jordan, in press). Measures of number sense also are strongly related to overall mathematics achievement (Booth & Siegler, 2006; Siegler & Booth, 2004). Although the number sense of children from low-income backgrounds typically lags behind that of peers from more affluent families, low-income children's number sense can be improved through playing linearly arranged numerical board games (Ramani & Siegler, 2008; Siegler & Ramani, in press).

3. Differences Among Individuals and Groups

Clear and systematic differences in children's mathematical competence emerge in the preschool period. The differences are present in counting, comparing magnitudes, adding, subtracting, and other aspects of numerical knowledge. These early-emerging differences among children appear to have important long-term consequences. A study that followed over many years large, nationally representative samples of U.S. children, as well as children from Canada and Great Britain, showed that mathematical knowledge during preschool and kindergarten is strongly predictive of mathematical knowledge in third grade, fifth grade, and eighth grade (Duncan et al., 2007). The relation is similarly strong for boys and girls and for children from low-income and middle-income backgrounds. It also is apparent in both math achievement test scores and teacher ratings of children's mathematical competence. Thus, children's mathematical knowledge differs substantially by the time they enter school and in ways that predict their mathematics achievement at least through middle school.

Differences in mathematical knowledge of U.S. children at the beginning of kindergarten reflect many aspects of the children's background. The Early Childhood Longitudinal Study (ECLS), which examined a large, representative sample of U.S. children, revealed several factors that predict children's mathematical knowledge when they enter kindergarten (U.S. Department of Education, NCES, 2001). One predictor is a mother's education; children of mothers with at least some college education usually have more knowledge of numbers and shapes than children whose mothers did not graduate high school. Another group of predictors involves risk factors such as single-parent families, families in which English is not the primary language spoken in the home, and families living in poverty. Children from families with fewer risk factors usually enter kindergarten with greater knowledge of numbers and shapes than children from families with more risk factors. A third predictor is race and ethnicity: white, non-Hispanic children and Asian children usually enter kindergarten with greater mathematical knowledge than black and Hispanic children.

The mathematical knowledge that children bring to school also varies with the country in which the child was raised. Children from East Asia generally have more mathematical knowledge when they enter school than do children in the United States. This superior knowledge seems to reflect the greater cultural emphasis on math learning within East Asian cultures. As the Japanese psychologist Giyoo Hatano commented, "Asian culture emphasizes and gives priority to mathematical learning; high achievement in mathematics is taken by mature members of the culture to be an important goal for its less mature members" (1990, pp. 110–111). Consistent with this observation, mothers in China rate doing well at math as being just as important for their children as doing well at reading, whereas mothers in the United States rate learning math as considerably less important (Miller, Kelly, & Zhou, 2005). Also reflecting the greater East Asian emphasis on math, in one study that compared Chinese and U.S. children from similar backgrounds who were just beginning kindergarten, the Chinese children generated three times as many correct answers to addition problems (Geary, Bow-Thomas, Fan, & Siegler, 1993). The difference was due to the Chinese children having memorized more correct answers to problems and to their using more advanced strategies when they could not retrieve the answer from memory. Preschoolers in China also count much higher, aided by the greater regularity of number words in their language (Miller et al.). Knowledge of shapes and other geometric information, memory for numbers, and other mathematical skills are also more advanced for Chinese than for U.S. preschoolers (Starkey et al., 1999). Although almost all studies show this pattern, a few have not; for example, Song and Ginsburg (1987) found that U.S. preschoolers outperformed Korean preschoolers in informal math knowledge.

4. Improving Early Mathematical Knowledge

A variety of instructional programs have been developed to improve the mathematical knowledge of U.S. preschoolers, especially preschoolers from low-income backgrounds. Several of these programs have met with considerable success. Project Rightstart and its successor Number Worlds (Griffin, 2004) focus on helping young children form an appropriate mental representation of numbers, akin to a mental number line; on using this mental representation to think about sets of real-world objects and arithmetic operations on those sets; and on familiarizing children with the language of numbers and mathematics. The

Berkeley Math Readiness Project (Klein & Starkey, 2004; Starkey, Klein, & Wakeley, 2004) provides preschool children with experience in counting and numerical estimation; arithmetic, spatial, geometric, and logical reasoning; measurement; and other aspects of mathematics. The Building Blocks program (Clements, 2002; Sarama, 2004; Sarama & Clements, 2004) uses computer software tools to help preschoolers acquire geometric and numerical ideas and skills. All of these produce substantial positive effects on children's mathematical knowledge. For example, in one study, Griffin's Number Worlds curriculum produced median effect sizes (Cox Index for standardized mean differences between experimental and control group) of 1.79 for 6 measures on the posttest and 1.40 for 13 measures on a later follow-up. The Berkeley Math Readiness curriculum produced an overall effect size of .96 (Hedges' g) among low-income children, and the Building Blocks program produced an overall effect size of .77 (Hedges' g) on 9 measures of numerical understanding and 1.44 on 8 measures of geometrical understanding. These are not the only programs that have been shown to increase preschoolers' mathematical competence, but they are good examples of the types of promising efforts that are being made in this direction (for a more comprehensive review of these and other programs aimed at enhancing preschoolers' mathematical competence, see Sarama & Clements). Research is needed to establish the longer-term effects of these programs.

5. Conclusions and Recommendations

Most children develop considerable knowledge of numbers and other aspects of mathematics before they begin kindergarten. Even in kindergarten, children from single-parent families with low-parental education levels and low incomes have less mathematical knowledge than do children from more advantaged backgrounds. The mathematical knowledge that children from both low- and middle-income families bring to school influences their learning for many years thereafter, probably throughout their education. A variety of promising instructional programs have been developed to improve the mathematical knowledge of preschoolers' and kindergartners, especially those from at-risk backgrounds. Research that scales up these interventions and evaluates their utility in preschool and early kindergarten settings is urgently needed, with a particular focus on at-risk children.

D. Mathematical Development in Content Areas

This section provides a review of the cognition literature as related to learning in the core mathematical content areas identified in the *Report of the Task Group on Conceptual Knowledge and Skills*. At the most general level, these include whole number arithmetic, fractions, estimation, geometry, and algebra. The quantity and quality of research on this learning differs considerably across the mathematical content areas. The Task Group notes areas in which substantive conclusions about learning or obstacles to learning can be drawn, and key mathematical areas in which a better understanding of learning is needed but for which the research base does not allow strong conclusions to be drawn. At the end of the review for each content area, the Task Group presents Conclusions and Recommendations.

Due to limited time, space, and resources, the coverage in this review is far from exhaustive. Nonetheless, the literature was thoroughly reviewed across all theoretical perspectives on mathematics learning. The studies included in the review were the ones that met the highest criteria of methodological rigor, as documented in Section IV, Methodology, in this report.

1. Whole Number Arithmetic

Children's learning of whole number arithmetic is a critical step in their mathematics education and a complex undertaking that extends for many years and engages multiple memory and cognitive systems. Core areas of competency include knowledge of basic arithmetic facts, skill at using standard procedures or algorithms for solving complex problems, estimating answers, and knowledge of key concepts (National Council of Teachers of Mathematics, 2006). For some of these core areas, such as simple addition, there is a substantive research base from which reliable descriptions of skill development can be provided, and inferences regarding at least some of the factors that facilitate or impede this development can be drawn. At the same time, there are other core areas, such as division algorithms, for which there is comparatively little empirical research, and thus the Task Group cannot make strong statements regarding the progression of skill development or the factors that influence this development.

Debate is common regarding whether mathematics education and related research studies should focus on conceptual knowledge or procedural skills (Baroody, Feil, & Johnson, 2007; Star, 2005, 2007). Empirical studies that have simultaneously assessed both of these aspects of mathematical competency reveal interdependence in children's development of declarative knowledge (e.g., addition facts), procedural knowledge (e.g., arithmetical algorithms), and conceptual knowledge (e.g., understanding the base-10 system). Aspects of skill development for each of these different types of competencies may require different prior knowledge, different instructional techniques, and different patterns of practice for mastery (Cooper & Sweller, 1987; Kalyuga, Chandler, Tuovinen, & Sweller, 2001; Sweller, Mawer, & Ward, 1983), yet their development is often interrelated (Rittle-Johnson et al., 2001). Children's use of one algorithm or another, or the detection of a computational error can be influenced by their understanding of related concepts, and the execution of algorithms can provide a context for their conceptual learning (Geary, Bow-Thomas, & Yao, 1992; Fuson & Kwon, 1992b). Children's skill at estimating is firmly linked to their computational skills (Dowker, 2003), and their ability to solve different types of complex word problems is dependent on different mixes of declarative, procedural, and conceptual competencies (Fuchs et al., 2006; Hecht et al., 2003).

For ease of presentation, the Task Group covers skill progression separately for these different competencies; nonetheless, it includes a few explicit examples of their interrelationships. The associated cognitive studies involve a detailed and time-intensive assessment of children's problem solving and learning and thus do not typically include large, nationally representative samples. The smaller-scale cognitive studies have, nevertheless, produced findings that have been replicated by many research groups and oftentimes in many nations. The Task Group's focus is on these replicated outcomes.

a. Acquisition of Arithmetic Facts

Addition and subtraction

One of the most thoroughly studied areas in children's mathematical learning involves descriptive assessments of developmental and schooling-based changes in the ways children solve simple addition and subtraction problems (Ashcraft, 1982; Carpenter & Moser, 1984; Geary, 2006; Geary, Bow-Thomas et al., 1996), as well as theoretical (e.g., computer simulations) and quantitative studies of the cognitive mechanisms underlying these changes (Shrager, & Siegler, 1998; Siegler, 1987; Siegler, 1988a). These studies and studies in other domains have clearly indicated that children's problem solving does not involve a step-by-step progression from use of one procedure to the next, but rather involves a mix of procedures and memory-based processes (e.g., direct retrieval of a fact) at most ages (Siegler, 1996). Learning involves a change in the mix of strategies used during problem solving, as well as improvement in the speed and accuracy with which individual procedures and memory-based processes are executed (Delaney, Reder, Staszewski, & Ritter, 1998; Geary, Bow-Thomas et al.). The focus here is on the mix of procedures and processes children use when they solve simple addition and subtraction problems and on their progression toward the learning of basic facts.

The Task Group notes that the learning and subsequent retrieval of basic facts does not involve the representation of isolated problem-answer combinations in long-term memory. Rather, this knowledge is embedded in a network of number- and arithmetic-related information. The use of the term *fact retrieval* simply refers to the goal of remembering the correct answer; it does not imply that associated problems, numbers, and answers are unrelated to other forms of knowledge, such as knowledge of general magnitude of the answer.

Paths of acquisition

Concepts. Young children's ability to solve formal addition and subtraction problems, such as $5 + 3 =$; or $7 – 2 =$, requires an integration of their emerging knowledge of the properties of associativity and commutativity (described below) with their counting knowledge and counting procedures (Ohlsson & Rees, 1991; Rittle-Johnson, & Siegler, 1998). Although there is some evidence for such an integration, the relation between these conceptual and procedural aspects of children's arithmetical learning has not been as thoroughly studied as the independent development of these competencies. For instance, there are many studies of children's emerging counting procedures and concepts (e.g., Briars, & Siegler, 1984; Fuson, 1988; Gelman, & Meck, 1983; LeFevre et al., 2006) and many studies of children's procedural development in addition and subtraction (described below), but only a few studies that have explicitly attempted to examine the link between these competencies (e.g., Geary et al., 1992).

Procedures. By the time children in the United States enter kindergarten, the most common procedures used to solve simple addition problems involve finger counting; some problems will be solved by counting out loud or mentally, and some children will know a few basic facts (Siegler, & Shrager, 1984). Counting procedures vary in sophistication—in terms of supporting conceptual knowledge and working memory demands—and kindergarten children typically rely on the least sophisticated of these procedures, referred to as *counting-all*, whereby children count both addends starting from 1. With the more sophisticated procedure called *counting-on*, children state the value of one addend (suggesting they understand the cardinality principle) and then count a number of times equal to the value of the other addend, counting 5, 6, 7, 8 to solve 5 + 3 = (Fuson, 1982; Groen & Parkman, 1972). Preliminary studies suggest that children's shift from counting-all to counting-on is related, in part, to improvements in their understanding of counting concepts (Fuson; Geary et al., 1992; Geary et al., 2004).

The frequent use of counting procedures results in the development of memory representations of basic facts (Siegler & Shrager, 1984); the act of counting 5, 6, 7, 8 to solve 5 + 3 facilitates the formation of an association in declarative memory between the addends and the answer generated by the counting. Once formed, these representations support the use of memory-based problem-solving processes. The most common of these are direct retrieval of arithmetic facts and decomposition. The latter involves reconstructing the answer based on the retrieval of a partial sum; for example, the problem 6 + 7 might be solved by retrieving the answer to 6 + 6 and then adding 1 to this partial sum (Siegler, 1987). A similar pattern is evident with children's skill progression in subtraction (Carpenter & Moser, 1983, 1984; Siegler, 1989). As with addition, children initially use a mix of strategies but largely count, often using their fingers or physical objects (i.e., manipulatives) to help them represent the problem and keep track of the counting. Children also rely on their knowledge of addition facts to solve subtraction problems, which is called addition reference (9 – 7 = 2, because 7 + 2 = 9) or use other related information (see Thornton, 1990). The most sophisticated processes involve decomposing the problems into a series of simpler problems and directly retrieving the answer (Fuson & Kwon, 1992a).

Declarative information. The primary declarative information contributing to the fast and efficient solving of simple addition and subtraction problems is knowledge of basic facts. The representation of these facts in long-term memory enables the use of direct retrieval and decomposition to solve these problems. Cognitive studies indicate that, unlike their peers in East Asian countries (), many college students in the United States have not memorized all of the basic addition and subtraction facts and thus often resort to use of backup strategies (Campbell & Xue, 2001; Geary, 1996; Geary & Wiley, 1991; Geary, Frensch, et al., 1993).

Multiplication and division

Paths of acquisition

Concepts. The core associative, commutative, distributive, and identity concepts as related to multiplication are described in a separate section below.

Procedures. Trends in children's ability to solve simple multiplication problems mirror those described for addition and subtraction, although formal skill acquisition begins in the second or third grade, at least in the United States. The initial mix of strategies is grounded in children's knowledge of addition and counting, including use of repeated addition and counting by n (e.g., Campbell & Graham, 1985; Mabbott & Bisanz, 2003; Siegler, 1988b; Thornton, 1978, 1990). Repeated addition involves representing the multiplicand, the number of times indicated by the multiplier, and then successively adding these values; when presented with 2 × 3, the child adds 2 + 2 + 2. The counting by n strategy is based on the child's ability to count by 2s, 3s, 5s, and thus is dependent on memorization of these counting sequences. Somewhat more sophisticated strategies involve the use of rules (identity element in this example; see below), such as $n \times 0 = 0$, and decomposition (e.g., $12 \times 2 = 10 \times 2 + 2 \times 2$). As with addition and subtraction, the use of these procedures appears to result in the formation of problem and answer associations in long-term memory (Miller & Paredes, 1990).

In comparison with the other operations, considerably less research has been conducted on skill progression in division. The research that has been conducted indicates that children rely heavily on their knowledge of addition and multiplication (Ilg & Ames, 1951; Robinson, Arbuthnott, et al., 2006). Robinson, Arbuthnott, et al. found that fourth-graders solved more than half of simple division problems by means of an addition-based procedure; to solve $\frac{20}{4}$, they repeatedly added the divisor until the dividend was reached, $4 + 4 + 4 + 4 + 4 = 20$, and then counted the divisors. Fourth-graders will sometimes solve the problem through reference to the corresponding multiplication problem ($5 \times 4 = 20$ for this example) or retrieve a division fact (e.g., $\frac{6}{3} = 2$). By seventh grade, the majority of the problems are solved by multiplication reference, although retrieval and the addition-based procedure are still used to solve some problems. Unlike the three other operations, use of direct retrieval did not increase across grade level; about 15% of division problems were solved by direct retrieval in grades 4 through 7.

Declarative information. As with addition and subtraction, the primary declarative information contributing to the fast and efficient solving of simple multiplication and division problems is knowledge of basic facts, that is, the representation of these facts in long-term memory. Studies of college students in the U.S. and Canada [computational skills are similar for students from these countries (Tatsuoka, Corter, & Tatsuoka, 2004)] suggest that many of these adults have not mastered all basic multiplication facts (LeFevre et al., 1996), and may continue to rely on multiplication reference to solve larger division problems (e.g., $\frac{72}{9}$) (Campbell, 1999; LeFevre & Morris, 1999; Robinson, Arbuthnott, & Gibbons, 2002). In contrast, college students who received their primary education in China can quickly and accurately retrieve the answers to all multiplications problems—though they rely on the commutative relation between problems to facilitate retrieval of some problems (e.g., 9×6 is retrieved based on 6×9)—and most simple division problems (Campbell & Xue, 2001; LeFevre & Liu, 1997). The implication is that many, perhaps most, U.S. children have not achieved fluency with simple multiplication and division.

Obstacles to mastery

In keeping with the broader methods and literature described in the section in this report entitled General Principles: From Cognitive Processes to Learning Outcomes (e.g., Ericsson, Krampe, & Tesch-Römer, 1993; Newell & Rosenbloom, 1981), the learning of simple and complex arithmetic has been studied using a variety of speed-of-processing, behavioral, and brain imaging methods (Charness & Campbell, 1988; Frensch & Geary, 1993; Klapp, Boches, Trabert, & Logan, 1991; Rickard, Healy, & Bourne, 1994; Royer, Tronsky, Chan, Jackson, & Marchant, 1999). These studies consistently find that practice results in faster solutions to basic problems and fewer errors, as well as related reductions in the working memory resources needed for problem solving and changes in the brain regions supporting this problem solving (described in Brain Sciences and Mathematical Learning, Section F in this report). The cognitive mechanisms underlying these changes include increased use of memory-based processes, more rapid execution of problem-solving procedures, and faster retrieval of relevant information from long-term memory (e.g., Delaney et al., 1998; Rickard, 1997). Skilled use of procedures also appears to require an understanding of associated concepts (Geary et al., 1992; Ohlsson & Rees, 1991).

The experimental studies have revealed that for most individuals, the ease of learning and retrieving arithmetic facts varies by the type of operation. The learning of addition and multiplication facts occurs with less practice than the learning of subtraction and division facts (Campbell & Xue, 2001; Rickard, 2004, 2005; Rickard et al., 1994). One reason for this operation effect is that the commutative principle for addition and multiplication (i.e., $a + b = b + a$; $a \times b = b \times a$) facilitates the learning of the associated facts; the learning of one combination of addition (e.g., 3 + 4) or multiplication pairs (e.g., 7 × 2) contributes to the learning of the commuted pair (i.e., 4 + 3, 2 × 7); but this relation does not hold for subtraction and division. Graduate students educated in China directly retrieve answers to smaller-valued subtraction and division (e.g., $\frac{28}{4}$) problems, but often solve larger-valued problems (e.g., $\frac{56}{8}$) through reference to the corresponding addition or multiplication problem, respectively (Campbell & Xue). A similar pattern is found for North American (Canada and United States) college students, but often extends to smaller-valued subtraction and division problems (LeFevre & Morris, 1999; Mauro, LeFevre, & Morris, 2003). This type of "mediated" retrieval is faster and more efficient than the use of procedures but still requires more time and an additional cognitive step—thus increased opportunity to commit an error—than direct retrieval of the answer. Finally, Rickard (2005) found that skill at factoring is related to knowledge of multiplication facts and that the practice of factoring (e.g., when presented with 21, the participant produces 7, 3) speeds subsequent retrieval of multiplication facts.

In studying the cognitive bases of children's arithmetic learning, researchers have not only examined how problem-solving approaches change with practice but also how these approaches vary across grade level and follow introduction of the operation (e.g., multiplication) in the school curriculum (Geary, 1996; Geary, Bow-Thomas, et al., 1996; Lemaire & Siegler, 1995; Miller & Paredes, 1990; Royer et al., 1999; Siegler, 1988b, 1989; Siegler & Jenkins, 1989; Steel & Funnell, 2001). The results of such studies are consistent with the experimental research: Fast and efficient problem solving is achieved with shifts from frequent use of counting or other procedures to direct retrieval of basic facts or use of decomposition. As with the experimental studies, children appear to learn addition and multiplication facts more easily than they learn subtraction and division facts, although comparatively little is known about children's learning of division.

Studies of children in the United States, comparisons of these children with children from some other nations, and even cross-generational changes within the United States indicate that many contemporary U.S. children do not reach the point of fast and efficient solving of basic arithmetic problems (Fuson & Kwon, 1992a; Koshmider & Ashcraft, 1991; Geary, Salthouse, et al., 1996; Geary et al., 1997; Schaie, 1996; Stevenson et al., 1985; Stevenson, Lee, Chen, Lummis, et al., 1990; for discussion see Loveless & Coughlan, 2004).

This point is particularly evident with comparisons of U.S. children to children educated in East Asia. Stevenson, Lee, Chen, Lummis, et al. (1990) assessed the speed and accuracy with which 480 first- or fifth-grade children from the United States and 264 same-grade children from China solved simple addition problems (e.g., 5 + 9), and found first-graders from China accurately solved three times as many problems (in 1 minute) than their U.S. peers (d = 2.97). The difference was smaller (1.6:1, d = 1.69) but still substantial for fifth-graders. The pattern was replicated for the basic subtraction skills of children from China and the U.S. for 6th- (d = 2.05) and 12th- (d = 2.05) graders matched or equated on cognitive ability (Geary et al., 1997). Although cognitive studies of how the children solved the problems have not been conducted for all arithmetical operations, the available evidence for addition suggests the differences in efficiency are related to less frequent use of retrieval by U.S. children, use of less sophisticated counting procedures, and slower retrieval and procedural execution when the problems are solved the same way (Geary, Bow-Thomas, et al., 1996).

The reasons for these differences are likely to be multifaceted, including language-related differences in the structure of number words, parental involvement in mathematics learning, and curricula (Miller et al., 1995; Miura, Okamoto, Kim, Steere, & Fayol, 1993; Steel & Funnell, 2001; Stevenson, Lee, Chen, Stiegler, et al., 1990). For instance, the structure of Asian-language number words where the teen values are stated as ten one, ten two, may facilitate, with teachers' guidance, the use of decomposition strategies to solve simple addition and subtraction problems (Fuson & Kwon, 1992a). Cross-national differences in mathematics curricula have not been directly tied to the cognitive studies of children's arithmetic learning. Nonetheless, results from several smaller-scale studies suggest such a link: In a review of the frequency of presentation of simple addition problems in first- to third-grade mathematics textbooks in the United States, Hamann and Ashcraft (1986) found that easier problems (e.g., 3 + 4) were presented much more frequently than harder problems (e.g., 8 + 7). In contrast, Geary (1996) found the opposite pattern in workbooks used to learn addition in China; a similar pattern of easier mathematics problems being presented in U.S. textbooks relative to same-grade textbooks from other nations has been reported by other researchers (Fuson, Stigler, & Bartsch, 1988). For third-grade children from the United States and China, the speed with which individual addition facts were retrieved from long-term memory was correlated (r's = .34 to .49; d's = 0.74, 1.12) with the cumulative (first- to third-grade) frequency with which the problems were presented in their respective countries. Whether children are learning addition in China or the United States, fast and efficient problem solving is related to frequency of prior exposure to the problem.

b. Learning Arithmetical Algorithms

Addition, subtraction, multiplication, and division

In this section, the four arithmetic operations are considered together because so little is known about children's learning of multiplication algorithms and division, and because what is known suggests similar obstacles to mastery across operations. The learning of algorithms requires a combination of an explicit conceptual understanding of related concepts (e.g., base-10); an understanding of when the algorithm should and should not be used; and, eventually, the ability to use the algorithm quickly and efficiently.

Paths of acquisition

Concepts. A central concept related to the use of arithmetical algorithms is the base-10 system and the corresponding understanding of place value and "trading" across columns (Blöte et al., 2001; Fuson & Kwon, 1992b). Coming to understand the base-10 system and place value is highly dependent on instruction (Hiebert & Wearne, 1996). Studies conducted in the United States have repeatedly demonstrated that many elementary-school children do not fully understand the base-10 structure of multidigit written numerals (e.g., understanding the place value meaning of the numeral) or number words (Fuson, 1990). As a result, many of these children are unable to effectively use this system when attempting to solve complex arithmetic problems. It appears that many children require instructional techniques that explicitly focus on the specifics of the repeating decade structure of the base-10 system and that focus on clarifying often confusing features of the associated notational system (Fuson & Briars, 1990; Varelas & Becker, 1997). An example of the latter is that sometimes "2" represents two units; other times it represents two tens; and, still other times it represents two hundreds (Varelas & Becker). Unlike East Asian languages where the base-10 structure is transparently represented in the associated number words (e.g., 21 is stated as two ten one), the English language number word system may actually lead to confusions about this relation (Miura et al., 1993). The development of base-10 knowledge is also facilitated by understanding that basic units ("ones") can be aggregated to form higher-order ones ("tens"), and prior understanding of cardinality, min counting, (i.e., stating the value of the larger addend and counting a number of times equal to the value of the smaller addend) and skill at decomposing numbers (Saxton & Cakir, 2006).

Procedures. The solving of arithmetic problems that are more complex than the simple problems described above, such as 23 + 6 or 12 × 73, involves the application of prior arithmetical skills and knowledge, the incorporation of new knowledge, and the learning of new procedures or algorithms.

When learning complex addition problems, children initially rely on the knowledge and skills acquired for solving simple addition problems, as reviewed in Siegler (1983); problems can be solved by means of counting, decomposition, or regrouping, as well as the formally taught columnar procedure (Ginsburg, 1977; Reys, Reys, Nohda, & Emori, 1995; Siegler & Jenkins, 1989). The decomposition or regrouping strategy involves adding the tens values and the units values separately; the problem 23 + 45 would involve the steps 20 + 40, 3 + 5, and then 60 + 8 (Fuson & Kwon, 1992b). The most difficult process in terms of time needed to solve the problem and frequency of errors involves regrouping or "trading", as in the problem 46 + 58. As described in the Obstacles to Mastery section in this section, several factors contribute to the difficulty of regrouping.

A similar pattern is found when children are first learning to solve complex subtraction problems; they rely on their knowledge of simple subtraction and addition when using counting or decomposition to solve the problem (Siegler, 1989). As with complex addition, the process of regrouping, as with 33 - 17, is the most common source of difficulty (Fuson & Kwon, 1992b). There are comparatively few cognitive studies of children's learning of the algorithms for solving complex multiplication and division problems, as noted. Studies of adults reveal they solve most complex multiplication problems by using the standard algorithm and partial products (Figure 1), and that the carry operation is the most time-consuming process (Geary, Widaman, & Little, 1986; Tronsky, 2005).

Figure 1: Multiplication Algorithms

Partial Products	Standard Algorithm
32 x 53 6 90 100 1500 1696	32 x 53 96 1600 1696

Although they did not provide detailed results on the problem-solving steps children use to solve division problems (e.g., $\frac{345}{23}$), Pratt and Savoy-Levine's (1998) study of contingent tutoring (i.e., providing different levels of support ranging from hints to explicit demonstration) is insightful. In one component of this study, fourth- and fifth-grade children from Canada—recall that the computational skills of U.S. and Canadian children are comparable (Tatsuoka et al., 2004)—solved four division problems and were scored on the accuracy of executing four problem solving steps: estimating the quotient; multiplying the divisor and the quotient; subtracting the product from the dividend; and obtaining the remainder. Before tutoring, these children correctly executed less than four of 32 problem-solving steps across four problems (there were eight steps/problem); one type of tutoring substantially increased accuracy but other types did not. The overall pre-tutoring accuracy rate of 12% is substantially lower than the 25% to 72% correct found for fourth-graders from Japan for problems of similar complexity (Reys et al., 1995).

Declarative information. Adults who are skilled at using arithmetical algorithms can describe the steps they used in the execution of the algorithms (Tronsky, 2005). Mastery of algorithms, however, may involve commitment of the associated steps to procedural memory, rather than to explicit declarative memory. With mastery, it is expected—based on studies of procedural learning in other domains and the studies that have been conducted in arithmetic (Delazer et al., 2003; Pauli et al., 1994; Tronsky)—that the algorithms can be executed automatically and without need for explicit recall and representation of each problem-solving step in working memory.

Obstacles to mastery

As the complexity of the arithmetical problem increases, the number of potential obstacles to mastery increases. The learning of arithmetical algorithms and their fluent execution once learned are influenced by process constraints, conceptual knowledge, errors of induction, and current context. Process constraints include the individuals' working memory capacity (DeStefano & LeFevre, 2004; Hitch, 1978), and the fluency with which component skills embedded within the algorithm can be executed (e.g., ease of retrieving basic facts) (Fuchs et al., 2006; Royer et al., 1999; Starch, 1911). Conceptual knowledge, especially an understanding of the base-10 system and place value, influences how the individual organizes the component processes that compose the algorithm, and facilitates the flexible use of alternative algorithms and the transfer of algorithms to the solving of novel problems (Blöte et al., 2001; Fuson & Kwon, 1992b; Hiebert & Wearne, 1996). During algorithmic learning, children and adults often make errors of induction based on prior learning of related algorithms or related concepts (Ben-Zeev, 1995; VanLehn, 1990). Contextual factors vary from external factors that reduce process limitations (e.g., scratch paper) or that exacerbate (e.g., high-stakes testing) these limitations, as well as factors (e.g., teacher, worked examples) that may help the individual recall relevant concepts (Beilock et al., 2004; Cary & Carlson, 1999).

The solving of complex arithmetic problems, especially during the early phases of learning, requires the retention of intermediate results in working memory while the individual processes the next problem-solving step. These demands require attentional control and working memory resources, and are a potential source of problem solving failure (e.g., Ashcraft & Kirk, 2001; DeStefano & LeFevre, 2004; Hitch, 1978; Logie, Gilhooly, & Wynn, 1994). Experimental manipulations of problem complexity and results from the use of dual-task procedures—asking the individual to engage in an activity that occupies one component of working memory (e.g., repeating nouns) during arithmetical problem solving—suggest the central executive component of working memory is a core source of processing limitations. The phonological loop and visuospatial sketch pad also can pose limitations for some aspects of problem solving (DeStefano & LeFevre); the execution of the carry or borrow procedure is particularly time consuming, and places added demands on the central executive and phonological loop. Working memory resources improve as children mature (Cowan et al., 2002), and can be functionally improved at any age with practice of the algorithm (Beilock et al., 2004; Tronsky, 2005) and with use of external memory aids (e.g., scratch paper) (Cary & Carlson, 1999).

Practice reduces the working memory demands of the problem because it results in the formation of procedural memories, such that the algorithm can be executed without the need to explicitly recreate and represent the sequence of steps in working memory. External aids reduce these demands because intermediate steps can be noted externally (e.g., on scratch paper or with a worked example) rather than in working memory. Practice to the point of automaticity reduces the disruptive effects of anxiety on problem-solving performance. In high-stakes situations, as when performance will be evaluated by others, anxious individuals tend to have thoughts regarding their competency intrude into working memory (Ashcraft & Kirk, 2001; Beilock et al., 2004); these intrusions functionally lower working memory capacity and thus increase the likelihood of committing an error. Beilock et al. experimentally demonstrated that this "choking under pressure" occurs much more often when problem solving requires use of infrequently practiced algorithms; in their studies, errors were rare for frequently practiced algorithms in both low-pressure and high-pressure situations.

Although practicing algorithms has the benefit of eventual automatic execution and reduced working memory demands, practice without conceptual knowledge can result in reduced flexibility in use of alternative algorithms (Blöte et al., 2001; Hiebert & Wearne, 1996). In an experimental study of algorithmic learning in second-graders in the Netherlands, Blöte et al. found that the combination of direct instruction of algorithms in the context of learning associated concepts resulted in a better ability to flexibly use one algorithm or another, depending on the structure of the problem, than did direct instruction of algorithms without a conceptual context. The Task Group discusses the importance of conceptual knowledge in more depth in the next section, Core Arithmetical Concepts; it is noted here that conceptual understanding in one area of arithmetic can sometimes facilitate transfer of algorithms to related problems but at other times can interfere with algorithmic learning (Ben-Zeev, 1995; VanLehn, 1990). For instance, learning the commutative property for addition can lead to the overgeneralization of this property to subtraction; leading students to infer that since 92 – 17 = 75, 17 – 92 = 75.

At other times, errors in executing algorithms are related to a poor understanding of the base-10 system and place value (Fuson & Briars, 1990; Fuson & Kwon, 1992b). Because they do not understand the base-10 concept and place value, many children do not understand that the 1 traded from the units- to the tens-column, for instance when solving 24 + 38, actually represents 10 and not 1; in this case, they write 52 as the answer, instead of 62. Children may not execute the carry procedure at all (leading to answers such as 24 + 38 = 512), or they may ignore place holding 0 values and carry across columns (e.g., 407 + 309 = 806). A similar type of algorithmic error has been found with complex subtraction (VanLehn, 1990; Young & O'Shea, 1981), but much less is known about algorithmic development in complex multiplication and division.

In the earlier mentioned assessment of the speed and accuracy of the arithmetical problem solving of first- and fifth-grade children from the United States and China, Stevenson, Lee, Chen, Lummis, et al. (1990) found that fifth-graders from China solved more than twice as many multidigit (e.g., 34 + 86) addition problems in 1 minute as did their U.S. peers (d = 1.91). A similar pattern was found comparing multidigit subtraction skills of children from China and the United States for 6th- (d = 1.89) and 12th- (d = 1.82) graders that matched or equated on general cognitive ability (Geary et al., 1997). The latter study found a smaller gap for multidigit addition than that found by Stevenson et al., but the differences were still substantial in both 6th- (d = 1.22) and 12th- (d = 1.30) grades. The same pattern was found for multidigit multiplication problems (e.g., 23 x 6), whereby fifth-graders from China solved more than twice as many problems in 1 minute as did their U.S. peers (d = 1.57; Stevenson, Lee, Chen, Lummis, et al.). The source of these fluency differences is not entirely understood but is related at least in part to a better understanding of the base-10 system and place value in East Asian than in U.S. students. It also is likely related to differences in the grade placement, the quantity and quality of algorithmic practice, and the extent to which this practice is integrated with concept learning (Fuson & Kwon, 1992b; Fuson et al., 1988).

c. Core Arithmetical Concepts

The core arithmetical concepts that children should come to understand and apply during problem solving are the associative and commutative properties of addition and multiplication (described below), the distributive property of multiplication [e.g., $a \times (b + c) = (a \times b) + (a \times c)$], identity elements for addition ($a + 0 = a$) and multiplication ($b \times 1 = b$), and the inverse relation between addition and subtraction, and between multiplication and division. The availability of research on children's understanding and skill at using these concepts is quite variable across these topics. There is, for instance, considerable work on children's understanding of commutativity as related to addition, but comparatively little work on children's understanding of identity elements and the inverse relation between multiplication and division.

Associativity and commutativity

The commutativity property concerns the addition or multiplication of two numbers, and states that the order in which the numbers are added or multiplied does not affect the sum or product ($a + b = b + a$; $a \times b = b \times a$). The associativity property concerns the addition or multiplication of three numbers, and again states that the order in which the numbers are added or multiplied does not affect the sum or product [$(a + b) + c = a + (b + c)$; $(a \times b) \times c = a \times (b \times c)$]. Empirical research on children's understanding of these concepts has focused on the commutative property of addition (Baroody, Ginsburg, & Waxman, 1983; for review see Resnick, 1992), although some research has been conducted on associativity (Canobi et al., 1998, 2002). Different approaches have been used in this research:

A. An informal understanding is sometimes inferred when preschool children's physical manipulation of sets of objects or responses to such manipulations is consistent with these concepts (for review, see Resnick, 1992). A child might watch as sets of different objects (e.g., red candy, blue candy) are given to different dolls in different orders. Implicit knowledge of commutativity is inferred if the child indicates the dolls received the same amount.

B. A formal understanding is inferred when the child can explicitly state that answers to problems are equal (e.g., 14 + 78 = 78 + 14) and can justify his or her answer using the appropriate concept, that is, that number order does not affect the answer.

C. An intermediate level of knowledge is inferred when a child's solving of formal problems is consistent with an implicit understanding of the concept or the child provides a partial explicit justification (Baroody et al., 1983). If the problems 3 + 14 and 14 + 3 are presented one after the other, and the child counts to solve the first problem (e.g., 14, 15, 16, 17) and quickly states the same answer without counting to solve the second problem, an implicit understanding of commutativity is inferred. A partial justification might involve the child stating that the problems are the same, but does not include statements regarding number order.

Paths of acquisition

Concepts. By 4 to 5 years of age, many children understand that sets of physical objects can be decomposed and recombined into smaller and larger sets, and that the order of these manipulations is not important; that is, they implicitly understand commutativity in this context (Klein & Bisanz, 2000; Sophian, Harley, & Martin, 1995; Sophian & McCorgray, 1994; Canobi et al., 2002). This implicit knowledge is limited to two sets of objects, indicating that most children of this age do not implicitly understand associativity. Moreover, many of these children may not link commutativity, as expressed in manipulation of physical sets, to addition of specific quantities. About half of kindergartners implicitly recognize commutative relations in simple addition problems (e.g., $3 + 2 = 3 + 2$), as do the majority of first-graders (Baroody et al., 1983; Baroody & Gannon, 1984). Many second- and third-graders will begin to provide partial explicit explanations of commutative relations, but it is not well understood when and under what instructional conditions children come to explicitly understand commutativity as a formal arithmetical principle. Some kindergarten children recognize associative relations when presented with sets of physical objects, and many first- and second-graders implicitly understand associative relations when they are presented as addition problems (Canobi et al., 1998, 2002). These studies have also demonstrated that an implicit understanding of associativity does not emerge until after children implicitly understand commutativity.

In comparison to addition, much less is known about children's implicit and explicit knowledge of commutativity and associativity as related to multiplication. In a study with third-graders who were just being introduced to multiplication, Baroody (1999) found that practice at solving multiplication problems (e.g., 3×4) made the solving of unfamiliar commuted problems (e.g., 4×3) faster and less error prone than other unfamiliar problems. This type of finding is consistent with the adult studies on retrieval of multiplication facts but is not sufficient to demonstrate an explicit conceptual understanding of the commutative property as related to multiplication.

Declarative information. The core concept of commutativity and associativity is that the order in which two (commutativity) or three (associativity) numbers are added or multiplied does not affect the result. Although elementary school children's justifications for a problem-solving approach often reflect a partial understanding of this equivalence (Baroody et al., 1983), many children do not explicitly state this core concept as a justification. It is not known when and under what instructional conditions children can express these concepts algebraically (e.g., $a \times b = b \times a$). It is also important for children to come to understand that commutativity and associativity do not apply to subtraction and division; children's problem-solving errors in subtraction suggest they often draw the incorrect inference that the principles apply to these operations as well (VanLehn, 1990).

Obstacles to mastery

The relation between children's implicit and explicit knowledge of commutativity and associativity is not fully understood. Resnick (1992) proposed that children's implicit understanding of commutativity and associativity provides the foundation for their explicit understanding of these concepts, but evidence for such a relation is mixed (Baroody et al., 1983). Many children implicitly or explicitly infer that commutativity applies to subtraction and thus often make errors; since $7 - 3 = 4$, it is inferred that $3 - 7 = 4$ (Young & O'Shea, 1981); the use of these "buggy rules" (i.e., use of a procedure that is correct for one type of problem to solve another type of problem for which the procedure is not appropriate) varies, however, as children may use them to solve one problem and then use a correct procedure to solve another (Hatano, Amaiwa, & Inagaki, 1996). These forms of error have not been as extensively studied for associativity with subtraction or to the misapplication of these principles to division, but similar confusions are likely.

Distributive property, identity properties, and inversion

There is not a sufficient amount of research on children's understanding of the distributive property of multiplication to draw conclusions at this time; e.g., $a\ (b + c) = ab + ac$. In one of the few studies of children's understanding of the distributive property (conducted in the United Kingdom), Squire, Davies, and Bryant (2004) found that less than 5% of fifth-graders could solve various forms of distributive property problems at above chance levels, as compared to 44% to 52% solving similar forms of commutative problems at above chance levels. The majority (> 74%) of sixth-graders correctly solved various forms of commutative principle problems, but only 22% to 44% of these children correctly solved various forms of distributive property problems.

Error patterns were systematic and suggested that the children confused addition and multiplication when solving distributive problems. Some of the distributive items were presented as word problems with an underlying form of if $a \text{ x } b = c$, then $(a + 1) \text{ x } b = c + b$. The first statement in a corresponding word problem item might be presented as 67 candies in each of 25 bags = 1675 candies. The next statement might then be presented as 68 candies in each of 25 bags = M candies. Children were provided with six potential answers and were given a limited amount of time to choose one of these. If the children understood the distributive property then they would choose the answer that is 1675 + 25, or 1700. If they approached the problem as an addition of 1 to both sides of the equation, then they would choose 1675 + 1 or 1676. Nearly all of the fifth-graders and the majority of sixth-graders committed this type of error.

As with the distributive property, there is not enough research to draw firm conclusions about children's understanding of identity elements. Studies of adults' mental arithmetic indicate that identity problems in addition ($a + 0 = a$) and multiplication ($b \times 1 = b$) are solved more quickly and accurately than are other problems, suggesting these may be solved by means of a rule (Miller, Perlmutter, & Keating, 1984). These findings, however, do not address the issue of whether these adults explicitly understand the mathematical concept of an identity element nor do they address the more focal issue of how children come to understand this concept. Studies of children's conceptual understanding of and ability to apply the distributive property and identity elements are clearly a priority for future research.

Inverse relations are an integral part of many aspects of mathematics. Children's first encounter with such a relation is with addition and subtraction; e.g., $a + b = c$, $c - b = a$. Studies of knowledge of the inverse relation between addition and subtraction have revealed an implicit understanding for many children by the time they enter kindergarten (Baroody, & Lai, 2007; Klein, & Bisanz, 2000; Vilette, 2002), and a growing implicit use of this relation with schooling, as reflected in problem-solving performance (Bryant, Christie, & Rendu, 1999; Gilmore & Bryant, 2006; Siegler & Stern, 1998). Developmental and experimental studies indicate that the majority of children implicitly use addition and subtraction inversion in their problem solving (as measured by a shorter time needed to solve inversion problems as compared to similar problems that cannot be solved with inversion) before they can explicitly state this relation. This pattern is common in many areas (Siegler, & Araya, 2005). An ability to explicitly state some aspect of this relation is found in many children by the end of the elementary school years (Robinson, Ninowski, & Gray, 2006).

However, many weaknesses in children's and even adults' (Robinson, & Ninowski, 2003) understanding of inverse relations are evident; many adults and most children do not have a firm grasp of the inverse relation between multiplication and division, nor do they appear to understand the concept of inversion at an abstract mathematical level. For instance, Robinson, Ninowski et al. (2006) found that knowledge of the inverse relation between addition and subtraction in sixth- and eighth-graders did not transfer to multiplication and division; they seemed to understand these relations separately for addition and subtraction and multiplication and division but did not link them together through the more general concept of inverse relations in mathematics.

d. Conclusions and Recommendations

American students do not meet the goal of fast and efficient solving of basic arithmetic combinations or execution of standard algorithms, and their competence in these areas is well below that of students in many other countries. American students have a poor grasp of most core arithmetical concepts; most American students do not understand the distributive property of multiplication, and they do not know identity elements or the inverse relation between division and multiplication, among other deficits. Mastery of these core concepts is a necessary component of learning arithmetic and is needed to understand novel problems and to use previously learned procedures to solve novel problems. Debates regarding the relative importance of conceptual knowledge, procedural skills, and the commitment of arithmetical facts to long-term memory are misguided. The development of conceptual knowledge and procedural skills is intertwined, each supporting the other. Fast access to number combinations, prime numbers, and so forth supports problem solving because it frees working memory resources that can then be focused on other aspects of problem solving.

Classroom

The development of measures that support the teacher's ability to make formative assessments of children's procedural and conceptual competencies in all key areas of whole number arithmetic should be a research priority.

Training

Teachers. For teachers to take full advantage of the above noted types of formative assessments, they must have a better understanding of children's learning and the sources of children's conceptual and procedural errors in the content areas they are teaching. As an example, many errors on conceptual tasks are systematic and can provide information on how students are misunderstanding the concept. These errors can be used in formative assessments and to focus instruction. However, as noted, for teachers to make full use of these common errors in children's arithmetic learning, they must understand how children learn arithmetic and how children conceptualize and misconceptualize core concepts.

The development of courses in mathematical cognition for inclusion in teacher training programs will be necessary to address this goal.

Researchers. Programs that support cross-disciplinary pre-doctoral and postdoctoral training in cognition, education, and mathematics are needed to ensure a sufficient number of researchers that study children's mathematical learning, and have the background needed to bridge the gap between laboratory studies and classroom practice.

Curricula

The fast and efficient solving of arithmetic combinations and execution of procedures requires considerable practice that is distributed over time. The consistent failure of American children to achieve mastery of these topics is a strong indication that most current curricula in the United States do not provide these experiences. Although definitive conclusions cannot be drawn at this time due to lack of relevant, large-scale experimental studies, the research that has been conducted suggests that effective practice should include a conceptually rich and varied mix of problems, with several features:

1) Present more difficult problems (e.g., 9 + 7) more frequently than less difficult problems (e.g., 3 + 1); this is because long-term retention of difficult problems requires more practice.
2) Highlight the relations among problems.

   For example, the inverse relation between addition and subtraction:

   4 + 7 =

   11 - 4 =
3) Order practice problems in ways that reinforce core concepts.

   For example, identity elements:

   3 x 0 =

   0 x 8 =

   6 x 0 =
4) Include key problems that support formative assessments.

   Such problems can reveal students' misconceptions and problem-solving errors:

   7 - 4 =

   4 - 7 =

   Errors on the second problem (i.e., 4 - 7 = 3) are common because children infer that the commutative relation they learned for addition also applies to subtraction.

   Errors on these types of problems may be diagnostic of this incorrect inference, which can then be addressed as part of classroom instruction.

U.S. students' poor knowledge of core arithmetical concepts—the distributive property, identity elements, the inverse relation between division and multiplication, among others—is unacceptable and indicates a substantive gap in the mathematics curricula that must be addressed.

Research

Although much is known about some areas of children's arithmetical cognition and learning, further research is needed in the areas of children's learning of complex algorithms (e.g., division algorithm); the relation between conceptual knowledge and procedural learning; and on the learning of core concepts, including the base-10 number system, the distributive property of multiplication, and identity elements, among others.

Studies are needed that focus on the translation of cognitive measures of children's learning into formative assessments that are easily understood by teachers and used in the classroom.

Funding priorities that target areas of deficit in children's arithmetical cognition and learning are recommended, along with priorities that encourage projects that bridge the gap between basic research and classroom practice.

2. Fractions, Decimals, and Proportions

Fractions, decimals, and proportions are introduced into the mathematics curriculum as early as elementary school or its equivalent in different countries, and yet solving problems with these quantities remains difficult for many adults. Nevertheless, understanding and manipulating fractions is crucial for further progress in mathematics and for tasks of everyday life, such as computing interest on a loan or deciding among risky medical treatments (e.g., Kutner, Greenburg, & Baer, 2006; Reyna & Brainerd, 2007; Wu, 2006). Central to the charge of this Panel, knowledge of fractions and related concepts has been described as "fundamental to the learning of algebra" (p. 1; Wu, 2007). In a nationally representative sample, teachers of Algebra I rated students as having the poorest preparation in "rational numbers and operations involving fractions and decimals" among 15 areas of mathematics, surpassed only by "solving word problems" (Hoffer, Venkataraman, Hedberg, & Shagle, 2007, Table 3).

a. Definitions and Interpretations

Mathematically, the definition of a fraction begins with the concept of a number line (Wu, 2007). A fraction is defined as a point on the number line, based on the concept of a part-whole relation, with the unit segment [0,1] (the segment from 0 to 1) serving as the whole. The fraction $\frac{1}{3}$, for example, is obtained by dividing [0,1] into three equal parts. Every segment on the number line, not just the unit segment, can be similarly divided into three equal parts. More generally, m/n consists of *m* adjoining short segments of $\frac{1}{n}$ each (e.g., thirds). The example of $m = 10$ and $\frac{1}{n} = \frac{1}{3}$ is shown below:

```
   0   1  2   3  4

|–|–|–|–|–|–|–|–|–|–|–|–|–
                 10
                  3
```

From this mathematical definition of a fraction, other definitions can be derived, such as the division interpretation (i.e., $\frac{m}{n} = m \div n$).

Psychologically, however, there are at least five interpretations or "subconstructs" of fractions (Kieren, 1988), defined as follows by Sophian (2007, pp. 114–115):

- measure, as representing magnitudes that can be intermediate between whole numbers of units (e.g., magnitudes between 0 and 1);
- quotient, as numerical values obtained by dividing one whole number by another;
- ratio number, as representing the relative magnitude of two non-overlapping quantities (as in a recipe that calls for 3 eggs for every 2 cups of flour);
- multiplicative operator, as representing an extending/contracting or stretching/shrinking function applied to some object, set, or number (so that, e.g., taking $\frac{2}{3}$ of a quantity stretches that quantity by a factor of 2 but then shrinks the "stretched" quantity by a factor of 3 so that the end result is smaller than the original quantity); and
- part-whole, as representing one or more parts of a whole, where the parts are formed by partitioning the whole into a number of equal units (see Behr, Harel, Post, & Lesh, 1992).

There are a number of properties of fractions that are related to one or more of these interpretations, such as inversion (that fractions become smaller as denominators become larger, assuming that the numerator is held constant; (Sophian, Garyantes, & Chang, 1997), that the effect of denominator magnitude is multiplicative (e.g., Thompson & Saldanha, 2003), that segments are infinitely divisible or dense (that there are infinite fractions between two endpoints on a number line; e.g., Smith, Solomon, & Carey, 2005), and others. Unlike mathematical definitions, which can be explained, derived, or, with the help of theorems, proved to be related to one another in precise ways, the relations among different psychological interpretations or properties are unclear. Mathematically, although a precise definition of fractions using the number line makes it possible to derive other properties of fractions, empirically, a student might successfully perform tasks that fit one psychological interpretation of fractions but fail others that are mathematically equivalent (or derivable). How interpretations relate is a question that can be answered empirically; a taxonomy of interpretations based on a process model of underlying causal mechanisms could be produced through hypothesis-driven experimentation (see Platt, 1964). However, current scientific theory is not sufficiently developed to fully answer this important question of how the understanding of different properties and interpretations of fractions are related to one another.

Furthermore, because of the lack of clarity concerning how psychological interpretations of fractions relate to one another, scholars frequently differ in the meanings they attach to such terms as *conceptual knowledge* of fractions, emphasizing varied interpretations and properties. Fortunately, researchers generally provide operational definitions of conceptual knowledge (as well as of computational facility) by precisely specifying the tasks that subjects are asked to perform. For example, subjects might be asked to judge the relative magnitude of two fractions with identical denominators but different numerators, or vice versa. Other tasks include judging equivalence of fractions, translation of pictures (e.g., of pizzas with different portions shaded) into numerically expressed fractions, ordering fractions according to magnitude, judging which of two pairs of fractions are closer in magnitude, and computation (e.g., adding, subtracting, multiplying or dividing fractions).

Certain tasks are more diagnostic than others with respect to assessing specific aspects of conceptual knowledge of fractions. For example, if subjects judge $\frac{2}{3}$ and $\frac{4}{5}$ to be equivalent they likely do not understand that the relation between numerator and denominator is multiplicative rather than additive (Sophian, 2007). Similarly, if subjects select a container with 3 winning chips out of 7 chips rather than a container with 1 winning chip out of 2 chips (the so-called numerosity effect or ratio bias), they are failing to take the magnitude of the denominator into full account (e.g., Acredolo, O'Connor, Banks, & Horobin, 1989; Hoemann & Ross, 1982; Reyna & Brainerd, 1994, 2008). Although no task is process pure in the sense that it cleanly measures one and only one psychological process, specific empirical tests have been devised to identify processes that underlie judgments involving fractions, decimals, and proportions (see Kerkman & Wright, 1988; Siegler, 1981, 1991; Surber & Haines, 1987). Therefore, in the remainder of this review, conceptual knowledge is identified with respect to performance on specific tasks that are designed to diagnose comprehension of aspects of knowledge about fractions, decimals, or proportions. Psychometric studies have distinguished computational ability from conceptual knowledge and, thus, research concerning the former is also reviewed by the Task Group.

b. Extent of the Problem

Computations involving fractions and decimals have proved challenging for every group that has been tested in the U.S. Difficulties emerge when such concepts are introduced in elementary school, and they persist through middle school, high school, and into adulthood, extending beyond those with learning disabilities in mathematics (e.g., Hecht et al., 2007; Mazzocco & Devlin, in press; U.S. Department of Education, NCES, 2003; Sophian, 2007; Stafylidou & Vosniadou, 2004). The percentage of middle school students who have difficulties with fractions and decimals, which has been estimated at 40%, far exceeds the cumulative incidence of MLD, as the Task Group reviews in the section on Learning Disabilities (Barbaresi et al., 2005; Hope & Owens, 1987; U.S. Department of Education, NCES, 1990; Smith, 1995). To illustrate, the 1990 National Assessment of Educational Progress documented that only 53% of 7th-graders and only 71% of 11th-graders could correctly subtract two mixed fractions with unlike denominators, despite the fact that such content is typically taught in elementary school (U.S. Department of Education, NCES, 1990). Recent assessments paint a similar picture. On the 1996 and 2005 NAEP tests, only 65% and 73% of eighth-graders, respectively, were able to correctly shade $\frac{1}{3}$ of a rectangle; on the 2004 NAEP test, only 55% of eighth-graders could correctly solve a word problem involving dividing one fraction by another.

Adults also perform poorly on problems involving fractions, decimals, and proportions. The most recent report of the National Assessment of Adult Literacy (NAAL) assessed literacy and numeracy in 2003 for 19,000 U.S. adults, who completed realistic tasks (Kutner et al., 2006). More adults scored in the Below Basic level on the quantitative scale (22%) of the NAAL than on any other scale, such as those measuring document or prose literacy (Kutner et al.). Most studies of adult "numeracy" assess the ability to perform simple computations or quantitative judgments concerning decimals, probabilities, percentages, and frequencies (e.g., Fagerlin, Zikmund-Fisher, & Ubel, 2005; Lipkus, Samsa, & Rimer, 2001; Schwartz, Woloshin, Black, & Welch, 1997; Woloshin, Schwartz, Byram, Fischhoff, & Welch, 2000; for a review, see Reyna & Brainerd, 2007). For example, one question from the Newest Vital Sign (NVS) test assesses the ability to calculate percentages of ingredients based on information from a nutrition label for ice cream (Weiss et al., 2005). Another well-known numeracy scale includes 11 questions, all of which pertain to fractions, decimals, and percentages (Lipkus et al.). Only 46% of adults in one sample and 24% in another were able to answer a question from this scale correctly that involved converting frequencies to percentages (i.e., In the Acme Publishing Sweepstakes, the chance of winning a car is 1 in 1,000. What percent of tickets win a car?). The percentage of adults who answered three such questions correctly ranged from 15% or 16% (Lipkus et al.; Schwartz et al.) to 38% (Black, Nease, & Tosteson, 1995; Woloshin et al.), including samples that were mostly college-educated. Scores on these tests have been found to relate to important real-world outcomes, such as patients' knowledge, health behaviors, health outcomes, and medical costs (American Medical Association Ad Hoc Committee on Health Literacy, 1999; Baker, 2006; Berkman et al., 2004; Estrada, Martin-Hryniewicz, Peek, Collins, & Byrd, 2004; Institute of Medicine, 2004).

In sum, it is clear that a broad range of students have difficulties with fractions, and these problems continue after graduation for many adults (e.g., Hecht et al., 2007; Mazzocco & Devlin, in press). The failure to attain basic facility with fractions constitutes an obstacle to progress to more advanced topics in mathematics, including algebra (although direct evidence for this link is lacking, but see e.g., Hecht, 1998; Hecht et al., 2003; Heller, Post, Behr, & Lesh, 1990; Loveless, 2003) and, presumably, to career paths that require mathematical proficiency (e.g., National Science Board Commission on 21st Century Education in Science, Technology, Engineering, and Mathematics (STEM, 2006)), as well as potentially interfering with life-and-death aspects of daily functioning, such as compliance with medication.

c. Paths of Acquisition

Informal, implicit knowledge

In order to assess competence accurately, it is important to separate children's understanding of formal fractional notation (i.e., what the line between two numbers in a fraction such as $\frac{1}{3}$ means) from their intuitive ability to understand fractional relations and perform calculations using fractional quantities (e.g., Mix et al., 1999). Illustrating the difficulty in understanding notation, children frequently add numerators and denominators together without regard for the notational convention that each numerator-denominator combination refers to a single quantity (e.g., $\frac{3}{4} + \frac{1}{2} = \frac{4}{6}$) (Carpenter et al., 1978; see also Silver, 1986; Resnick & Ford, 1981). When such notational constraints are removed, young children reveal a nascent ability to understand ratios (Geary, 2006; Mix et al.; Sophian, 2000). Preschool children's experiences with and understanding of part-whole relations among sets of physical objects, such as receiving $\frac{1}{2}$ of a cookie or having to share 1 of their 2 toys, may contribute to an early understanding of simple ratios (Correa, Nunes, & Bryant, 1998; Geary; Mix et al.).

For example, avoiding the use of conventional notation, Goswami (1989) gave 4-, 6-, and 7-year-olds a series of analogy problems using shaded portions of geometric shapes such as $\frac{1}{2}$ of a circle: $\frac{1}{2}$ of a rectangle: $\frac{1}{4}$ of a circle: ?, and the children selected an answer from among five alternatives. A simpler version of the task was also presented in which proportions did not change across shapes (e.g., $\frac{1}{2}$ of a diamond: $\frac{1}{2}$ of a circle: $\frac{1}{2}$ of a square: ?) and children selected from among four alternatives. Four-year-olds performed significantly above the chance level of 25% correct in the simpler task (56% correct), and 6- and 7-year-olds performed nearly perfectly (86% and 91% correct, respectively). However, performance for 4-year-olds was only 31% correct in the harder version of the task, though significantly above chance-level performance, and 6-year-olds remained far from perfect at 74% correct. Thus, the ability to recognize equivalent fractions undergoes significant development in early childhood, but basic competence emerges before children enter formal schooling.

Similarly, the ability to manipulate fractions—to engage in a kind of informal computation with fractions that does not involve conventional notation—is also present early. In a study of simple part-whole relations, Mix et al. (1999) administered a nonverbal task that assessed children's ability to mentally represent and manipulate $\frac{1}{4}$ segments of a whole circle. The results indicated that children as young as 4-years-old could calculate with fractional amounts of less than or equal to one, as shown by their ability to recognize fractional manipulations. For instance, if $\frac{3}{4}$ of a circle was placed under a mat and $\frac{1}{4}$ of the circle was removed, the children recognized that $\frac{1}{2}$ of a circle remained under the mat. However, it was not until 6 years of age that children began to understand manipulations that were analogous to mixed numbers; for instance, placing $1\frac{3}{4}$ circles under the mat and removing $\frac{1}{2}$ a circle. These results suggest that about the time children begin to show an understanding of part-whole relations in other contexts (Sophian et al., 1995; Sophian & McCorgray, 1994; Resnick, 1992), they demonstrate a rudimentary understanding of fractional relations. Although it is possible that children in the Mix et al. study represented the $\frac{1}{4}$ sections of the circles as single units and not as parts of a whole, this seems unlikely because solution of whole number and fraction problems differs in important ways.

Correa et al. (1998) argue that sharing forms the basis for preschool age children's ability to partition a quantity into roughly equal parts through a process of distributive counting (see also Hunting & Davis, 1991; Miller, 1984). Preschoolers also know that the term *half* refers to one of two parts (Hunting & Davis) and can use the notions of greater than half versus less than half to recognize which of two proportions are closer to a target proportion (Spinillo & Bryant, 1991; see also Singer-Freeman & Goswami, 2001; Spinillo & Bryant, 1999). However, knowing how to share equal amounts or that some fractions are greater or lesser than half is not the same as understanding the inverse relation among quantities (e.g., between numerator and denominator in a fraction when the value of the fraction is held constant or between terms in division).

For example, Sophian et al. (1997) asked 5- and 7-year-old children to determine which of two sharing scenarios would yield a larger portion for a recipient. Consistent with the idea that sharing forms an early basis for understanding fractions, children understood that if different total quantities were shared among the same number of recipients, the recipients who shared the larger total quantity would get larger portions than those who shared the smaller total quantity. However, the children also expected that sharing *equal* total quantities among different numbers of recipients would result in larger portions when shared with a larger number of recipients, compared to fewer recipients. The latter expectation violates the inverse rule that the larger the number of shares, the smaller the size of each share, or vice versa, when total amounts are held constant. About half of 6-year-olds were able to understand the inverse relation between divisor and quotient in Correa et al.'s (1998) sharing task, which is well below the age at which division is formally taught in schools.

Acredolo et al. (1989), Schlottman (2001), and others using sensitive techniques that do not require explicit numerical computation have also shown an early appreciation for the inverse relation between numerators and denominators in probability and other ratio concepts (e.g., Hoemann & Ross, 1982; Reyna & Brainerd, 1994; Reyna & Ellis, 1994). These techniques, such as functional measurement, make it possible to discern whether the perceived relation between numerator and denominator is multiplicative rather than additive; by first grade, most students correctly perceived the relation to be inverse and multiplicative (see also Jacobs & Potenza, 1991). Notably, Sophian et al. (1997) found that in a study subsequent to the one reported above, children were able to appreciate and generalize the inverse relation after just a few trials demonstrating how changes in the denominator affected the size of each share, suggesting that some level of competence was already present to build on.

In sum, studies of nonverbal or implicit knowledge of fractions show an intuitive awareness of fractions based on part-whole relations, notions of sharing, and a limited conception of inverse, multiplicative relations between numerators and denominators (or divisors and quotients) in the preschool years. Like place value in decimals (e.g., .1 vs. .0001), the symbolic notation for fractions is not yet correctly interpreted and must be explicitly taught. Despite evidence of early basic competence, these studies show considerable change in performance between ages 4 and 7 (and beyond, in some studies) and significant differences between performance with whole numbers and fractions, with competence with fractions lagging substantially behind competence with whole numbers even on relatively simple tasks (e.g., Mix et al., 1999).

Formal, mathematical knowledge

Studies of elementary and middle school-aged children have focused on the acquisition of conceptual knowledge, computational skills (e.g., multiplying fractions), and the ability to use both of these abilities in conjunction with reading comprehension to solve word problems involving fractional quantities (Byrnes & Wasik, 1991; Hecht et al., 2007; Rittle-Johnson et al., 2001). Conceptual knowledge tasks have included identifying which of several fractions is largest, judging relative magnitude (e.g., 21/18 > 1), translating pictorial representations into equivalent formal fractional representations, and vice versa. Computational tasks have involved adding, subtracting, multiplying and (rarely) dividing fractions using pictures (e.g., providing pictorial representations of answers to pictorial problems), numbers, and verbal descriptions, as in word problems (e.g., Hecht et al.). Scores on items assessing conceptual knowledge have consistently been shown to explain unique variance (beyond general intellectual and reading abilities) in performance on computational fraction problems, word problems that include fractions, and estimation tasks with fractional quantities (e.g., Byrnes & Wasik; Hecht, 1998; Hecht et al., 2003; Sophian, 2007; Hecht et al., 2007).

Consistent with these findings and illustrating the close connection between conceptual and procedural (computational) abilities, Hecht (1998) reported that fully 82% of 1,474 errors on fraction computation problems could be classified as involving a faulty procedure, as opposed to wild guesses, no attempt, or calculation errors. Children's accuracy at recognizing formal procedural rules for fractions (e.g., when multiplying, that both numerators and denominators are multiplied) and automatic retrieval of basic arithmetic facts also predicted computational skills (i.e., accuracy in adding, multiplying, and dividing proper and mixed fractions), above and beyond the influence of intelligence, reading skills, and conceptual knowledge (see Hecht et al., 2007 for a review).

In a follow up study, Hecht et al. (2003) investigated effects of conceptual knowledge of fractions, basic arithmetic skills, working memory capacity, and on-task time in mathematics class. Outcome measures included the computation of fraction sums and products; the estimation of fraction sums; and the solution of one-step word problems involving fraction addition, multiplication, or division. On-task time referred to paying attention to instruction or engaging in other forms of on-task behaviors in the mathematics classroom, which other studies have shown correlates with the acquisition of academic skills (Bennett, Gottesman, Rock, & Cerullo, 1993 in two of six samples; McKinney & Speece, 1986; Wentzel, 1991); for example, Chen, Rubin, and Li (1997) reported a correlation of .52 for sixth- and eighth-graders between on-task time and academic achievement (.52 is the concurrent simple correlation at Time 2; a cross-lagged correlation of .47 was also reported). For fraction problems, Hecht et al. found that on-task time influenced performance through its effect on conceptual knowledge. Presumably, children who engaged in more on-task behavior in class were better able to acquire and practice conceptual understanding of fractions that then contributed to their ability to solve fraction computation, estimation, and word problems. For fraction computation, on-task time influenced performance through its effect on simple arithmetic knowledge as well. That is, on-task time was associated with better knowledge of simple arithmetic, and this arithmetic knowledge contributed to better performance on fraction computation problems.

Note that on-task time refers to focused attention and practice, rather than motivation. Motivation also has positive effects on fraction learning. Schunk (1996) showed that fourth-graders who had a learning goal (trying to learn how to solve fraction problems) rather than a performance goal (trying to solve fraction problems) had higher self-efficacy, skill, and other achievement outcomes, such as number of fraction problems solved. Children in the first of Schunk's experiments were also assigned either to a self-evaluation condition (they judged their fraction capabilities at the end of each of six learning sessions) or they did not engage in self-evaluation (instead answering an attitudes question at the end of the six sessions). In the second experiment, all children engaged in self-evaluation. In both experiments, the learning goal with or without self-evaluation led to higher motivation and achievement outcomes than the performance goal. Performance goals with self-evaluation were more effective than without self-evaluation.

Taken together, these studies suggest that motivation and on-task time contribute to superior conceptual knowledge of fractions, which broadly benefits computation, estimation, and skill at solving word problems (see Hecht et al. 2003 for detailed models). Basic skills (i.e., arithmetic knowledge, reflected in rapid retrieval of basic arithmetic facts) were also important, but they more narrowly benefited the solution of fraction computation problems. Basic skills were related to fraction computation even when other factors were controlled for, such as conceptual knowledge, reading ability, on-task time, and working memory. Therefore, early levels of basic arithmetic skills may predict those children who will later have difficulty with fractions, and building such skills (e.g., Goldman, Mertz, & Pellegrino, 1989; Jordan, Hanich, & Kaplan, 2003) may enhance performance on fraction computation problems.

Several studies have examined the relation between conceptual and procedural knowledge (computational ability), and their results echo findings in other domains of mathematics learning (e.g., Hecht et al., 2003; Rittle-Johnson & Alibali, 1999; Rittle-Johnson & Siegler, 1998). Rittle-Johnson et al. (2001) demonstrated that children's skill at solving decimal fractions was related to both their conceptual and procedural knowledge of fractions and that learning conceptual and procedural knowledge occurred iteratively. That is, conceptual knowledge predicted gains in procedural skills, and vice versa (Rittle-Johnson et al.; Sophian, 1997); Byrnes and Wasik (1991), in contrast, did not find that procedural knowledge affected conceptual knowledge, but failure to detect an effect is not evidence against it. Specifically, in two experiments with fifth- and sixth-graders, Rittle-Johnson et al. found that conceptual knowledge of decimal fractions at pretest (with initial procedural knowledge controlled for) predicted changes–as a result of instruction–in procedural competence from pretest to posttest. These changes in procedural competence (again controlling for initial scores on the procedural knowledge pretest) in turn predicted changes in conceptual knowledge from pretest to posttest. The iterative model of gradual, bidirectional influence of conceptual and procedural knowledge on development has been supported in multiple domains of learning. This model explains why children might be able to pass one test of conceptual knowledge and yet fail another test; because children have intuitions about part-whole relations, for example, does not mean that they fully understand conventional fractions.

The mechanism linking conceptual and procedural knowledge appears to be children's ability to represent the decimal fraction on a mental number line, supported by correlational evidence from Experiment 1 and causal evidence from Experiment 2 of Rittle-Johnson et al.'s research (2001). This linear representation undergoes development from childhood to adulthood and has been linked to both whole number magnitude estimation (Siegler & Opfer, 2003) and fractional magnitude estimation (e.g., Opfer, Thompson, & DeVries, 2007). Adults have been shown to successfully use a mental number line to represent relative magnitude and to solve inference problems, and recent neuropsychological evidence points to an internalized representation that preserves spatial features of a physical line, such as left-to-right orientation (e.g., Bouwmeester, Vermunt, & Sijtsma, 2007; Trabasso, Riley, & Wilson, 1975; Zorzi, Priftis, & Umiltá, 2002). Prompting children to think of decimal fractions as composite representations (e.g., a certain number of tenths, a certain number of hundredths and so on) rather than common unit representations (e.g., .45 as 45 hundredths), and then mapping those representations spatially onto the number line, led to large gains in procedural knowledge (Rittle-Johnson et al., 2001). In addition, children who began with low conceptual knowledge benefited more from representational supports than children who began with higher conceptual knowledge.

Some scholars have argued that frequencies are "privileged" mental representations from an evolutionary perspective and have used this concept to explain common errors in fraction and decimal use (e.g., Brase, 2002). The claim is not dissimilar from Gelman's (1991) ideas about negative transfer from counting whole numbers, that "a frequentist representation that tends to parse the world into discrete, countable units" (p. 406, Brase) explains difficulties in dealing with part-whole relations as opposed to part-part relations (e.g., Sophian & Wood, 1997; Sophian & Kailihiwa, 1998). However, recent research disentangling effects of frequentistic representations from clarification of class-inclusion (or part-whole) relations has shown that using frequencies per se does not reduce errors (for reviews, see Barbey & Sloman, in press; Reyna & Brainerd, 2008). Making part-whole relations transparent (e.g., by using Venn diagrams or distinctively labeling classes that are nested or overlapping), however, has been found to reduce errors for children and for adults in problems involving fractions, decimals, percentages, and frequencies (e.g., Girotto & Gonzalez, 2007; Reyna & Brainerd, 1994; Reyna & Mills, in press). For more advanced reasoners who have acquired conceptual knowledge, representations that make part-whole relations salient or transparent, in contrast to making part-part relations salient, virtually eliminate errors for simple magnitude judgments (Brainerd & Reyna, 1990, 1995; Lloyd & Reyna, 2001).

Proportional reasoning involves many of the same elements as fractions, decimals, and other ratio concepts but it requires, in addition, the coordination of two ratio quantities. By this definition, judging the equivalence or relative magnitude of two fractions with unequal numerators and denominators is an example of proportional reasoning. Thus, many of the studies reviewed thus far concern *proportional reasoning* although they are not labeled as such. Early, informal competence can be detected if children are able to use perceptual cues, in particular surface area, to judge relative numerosity (Rousselle, Palmers, & Noel, 2004). Using carefully controlled stimuli, Rousselle et al. showed that 3-year-olds responded above chance, even for large numerosities, by using an analog mechanism that codes continuous perceptual dimensions. In another study involving visual displays rather than numbers or notation, 3- to 4-year-olds were able to match proportions of pizzas (divided into

8 slices) and boxes of chocolates (consisting of 4 pieces), even when the numbers did not match (e.g., matching $\frac{4}{8}$ to $\frac{2}{4}$) (Singer-Freeman & Goswami, 2001). Sophian (2000) showed that 4- to 5-year-olds were able to identify corresponding spatial ratios based on relational information rather than the exact form of the stimuli (e.g., matching large and small rectangles of similar proportions, and rejecting rectangles that matched on only one dimension). Sophian and Wood (1997) used a more difficult task involving "conflict" problems in which 5-to-7-year-olds matched sample pictures either to a test stimulus that preserved the part-whole relation or one that preserved the part-part relation. By age 7, children were able to use part-whole relations to compare proportions. Jeong, Levine, and Huttenlocher (2007) found that 6-, 8-, and 10-year-olds failed a proportional reasoning task when discrete quantities were used, but even the youngest children showed some success when proportions involved continuous quantities. Children's greater success with continuous quantities was related to the use of erroneous counting strategies in two discrete conditions: they counted the parts (and compared them) rather than comparing parts to wholes. In the continuous conditions, children were presumably able to rely on their earlier developing ability to perceptually compare relative surface areas.

Hence, although there is some debate about exact ages and some variation across tasks, the studies indicate that children prior to formal schooling can recognize proportional analogs when they can perceptually compare relative amounts of surface area. In contrast, when problems involve numbers and simple ratios, children generally perform poorly until 7 or 8 years of age, although gaps in understanding remain after these ages (Dixon & Moore, 1996; Fischbein, 1990; Kieren, 1988; Moore, Dixon, & Haines, 1991; Nunes, Schliemann, & Carraher, 1993; Singer, Kohn, & Resnick, 1997). Ahl, Moore, and Dixon (1992) compared the relation between informal, intuitive and formal numerical proportional reasoning in fifth-graders, eighth-graders, and college-aged subjects. In a temperature mixing task, varying amounts of water (1, 2, or 3 cups) at varying temperatures (20˚, 40˚, 60˚, and 80˚) were added to a container that was either cool (40˚) or warm (60˚), and children were asked what the resulting temperature would be. In the intuitive version of the task, quantities and temperatures were described verbally (e.g., cold, cool, warm, and hot), but in the numerical version, numbers were used (and students were told to use mathematics). Half of the students received the intuitive task first, and the other half received the numerical task first. Performing the intuitive task first improved performance in the numerical task; performance in the intuitive task did not change when it followed the numerical task. Therefore, students were able to use their intuitive understanding—elicited without numbers—to inform their numerical performance.

In sum, studies of elementary and middle school-aged children's abilities to solve fraction problems indicate that conceptual knowledge broadly determines performance in such tasks as estimation, word problems, and even computation. Procedural knowledge, too, influenced performance on such tasks, and worked hand-in-hand with conceptual knowledge to determine the benefit derived from instruction. A key mechanism linking conceptual and procedural knowledge is the ability to represent fractions on a mental number line, which also supports reasoning performance in adults. On-task time, motivation, working memory, and well-learned basic arithmetic skills (in addition to general intelligence and reading ability) were also determinants of performance. Studies of preschool and older children's ability to solve proportional reasoning problems mirror findings for fraction problems inasmuch as intuitive or pictorial versions of tasks are mastered early, in the preschool period. The tendency to rely on perceptual cues (comparing relative surface areas) continues in the elementary years, and children perform better using such intuitive strategies compared to numerical strategies. Among older elementary and middle school students, receiving an intuitive version of a proportional reasoning problem aids performance on a numerical version, but not vice versa.

d. Obstacles to Mastery

Many observers have remarked on the contrast between the relative ease of learning to count, which is engaged in spontaneously and seems to build easily on prior intuitions, and the relative difficulty of learning fractions (e.g., Moss, 2005; Sophian, 2007). Indeed, some have attributed the difficulties children have with fractions to the lack of fit with properties of counting (Gelman, 1991); for example, 3 > 2 and therefore children infer that 1/3 > 1/2. Because of the property of infinite divisibility, fractions, unlike counting numbers, do not form a sequence in which each number has a fixed successor. Therefore, it has been argued that the one-to-one and stable-order principles that are important to counting are misleading when children attempt to generalize from whole numbers to fractions.

Gelman (1991) examined kindergarten and first-grade children's interpretations of pictorial and numeral representations of fractions to determine whether children try to generalize from counting to fractions. Consistent with points made earlier about lack of familiarity with notation, young children read fraction symbols such as "$\frac{1}{2}$" as combinations of whole numbers (e.g., "one and two" rather than "one-half"). As others have reported, children also incorrectly judged fractions with larger denominators to be larger than those with smaller denominators (e.g., that "$\frac{1}{4}$" was more than "$\frac{1}{2}$"). Finally, most children were unable to correctly place pictorial representations of proper and mixed fractions (e.g., $\frac{1}{3}$ of a circle, $1\frac{1}{2}$ circles) on a number line on which the values 0, 1, 2, and 3 were marked. Consistent with Gelman's analysis that each of these effects had to do with negative transfer from knowledge of whole numbers, Vamvakoussi and Vosniadou (2004) found that just over half ($\frac{9}{16}$) of a sample of ninth-graders expressed the view that fractions form a series (such as $\frac{1}{8}, \frac{2}{8}, \frac{3}{8}$ and so on) rather than being infinitely divisible. However, strong conclusions cannot be drawn from Vamvakoussi and Vosniadou's relatively small sample, and a subsequent study by Smith et al. (2005) found that elementary school children were able to express the idea of infinite divisibility when prompted (e.g., endorsing the idea that one could divide numbers in half forever without ever getting to zero).

Although the concept of infinite divisibility is of interest because it distinguishes fractions from whole numbers, this does not mean that children do not inappropriately apply knowledge about whole numbers to fractions regardless of what they believe about divisibility. It appears that when children do not understand the conventions of reading fractions, they overgeneralize from their knowledge of whole numbers. For instance, they often judge the relative magnitude of two fractions as corresponding to the relative magnitudes of the numbers within them, which is sometimes correct (e.g., $\frac{2}{5} \langle \frac{3}{5}$ and $\frac{2}{3} \langle \frac{4}{5}$), but sometimes is not correct (e.g., $\frac{2}{3} \rangle \frac{2}{5}$, $\frac{3}{4} \rangle \frac{5}{8}$, and $\frac{3}{4} = \frac{6}{8}$ although 3 < 5, 4 < 8, and 3 < 6) (Sophian, 2007). Similarly, they add compound fractions by reading left to right, such as adding 2 + $\frac{3}{8}$ to get $\frac{5}{8}$ (Mack, 1995; cf. Sophian, 2007). The reliance on knowledge of counting and whole numbers leads to predictable errors in judging relative magnitudes or equivalence of fractions. However, it is not clear that this negative transfer occurs because of conflicts with innate counting mechanisms. Rather, it may stem from lack of knowledge of conventional notation, an argument that is strengthened by the demonstration of accurate intuitions when such notation is not used.

Another potential obstacle to mastery of fractions is the use of pictorial representations in early demonstrations, without sufficient emphasis on the nature of wholes in part-whole relations and the importance of equal-sized parts (Sophian, 2007). For example, if fractions are represented as slices of pizza as they often are, it becomes difficult to conceptualize improper ($\frac{6}{5}$) fractions. The number line representation presented earlier would seem to be more robust, easily representing quantities less than and greater than one. However, little research has been conducted comparing the relative effectiveness of different representational formats and whether, for example, pizza slices or other part-whole pictorial representations introduce difficulties when children move to fractions beyond the unit segment. Despite assertions made about the relative merits of different formats (e.g. using discrete vs. continuous quantities to represent fractions), few experiments with sufficient sample sizes and appropriate dependent measures (i.e., learning outcomes) have been conducted. A straightforward, randomized assignment experiment, for instance, pitting initial instruction using the number line against pizza slices or other pictorial formats could be used to answer this question. Both near (using problems resembling examples from training) and far (using superficially different problems) transfer could then be assessed.

In addition to pictures, words seem to influence the mental representations that children form concerning fractions. Several studies have confirmed that being a speaker of English, Croatian, or other languages that do not demarcate parts and wholes in fraction names is an obstacle to mastery of fractions. In East Asian languages, the part-whole relation is reflected in the corresponding names for fractions; for example, “one-fourth” is “of four parts, one” in Korean (Geary, 2006). Children whose languages demarcate parts and wholes in fractions names are able to demonstrate conceptual knowledge (e.g. they are able to correctly associate numerical fractions with pictorial representations) prior to formal instruction in fractions. For example, Miura, Okamoto, Vlahovic-Stetic, Kim, and Han (1999) found that 6- and 7-year-old Korean children grasped the part-whole relations represented by simple fractions (e.g., $\frac{1}{2}, \frac{1}{4}$) before formal instruction in first and second grade and before Croatian- and English-speaking children, whose languages do not have transparent word names for fractions. Such evidence is correlational, however, and subject to alternative interpretations based on differences in culture and experience. However, Paik and Mix (2003) demonstrated that when nontransparent, whole-number representations were used, U.S. and Korean children made similar errors in a fraction-identification task (although Korean children still scored better overall). When presented with fraction names that explicitly referred to parts and wholes on analogy with Korean names, U.S. children's performance improved and their scores exceeded those of the same-grade Korean children. In order for such labeling to be effective without additional training, it must build on fundamentally sound intuitions. These studies introduce a manipulation—fraction names with explicitly marked parts and wholes—that resembles the class-inclusion effects noted earlier (e.g., using Venn diagrams and tagging sets) because they, too, highlight part-whole relations. Both kinds of interventions are effective without additional training, suggesting that confusion about parts and wholes in working memory, rather than a total lack of conceptual knowledge, is responsible for unaided errors (e.g., Brainerd & Reyna, 1990; 1995).

As noted earlier, working-memory limitations are an obstacle to mastery of fractions (although we use the language of limited capacity, interference rather than capacity may offer a more satisfactory explanation of developmental and individual differences) (Dempster, 1992). English and Halford (1995) analyzed the working-memory demands of different tasks, and argued that different fraction interpretations entail different information-processing demands. A ratio interpretation, for example, a 2:3 ratio between red and blue chips involves just binary relations, because only two subsets need to be related to each other. In contrast, conceiving of the same array as corresponding to the fraction $\frac{2}{5}$ entails "ternary" relations, because three sets are related, the total set of all chips and each of its subsets, red chips and blue ones. Assessing equivalence relations between two fractions, as in the expression $\frac{1}{2} = \frac{3}{6}$, entails "quaternary" relations, because relations among all four quantities must be considered; judging relative magnitudes of fractions with unequal denominators and numerators, such as $\frac{5}{14} \rangle \frac{11}{34}$, makes similar demands. Quotient interpretations of fractions (e.g., sharing 3 pizzas among 4 people) are more demanding of memory resources than part-whole interpretations because they involve a more complex series of mappings (see Sophian, 2007). Formally similar tasks can have different information-processing demands. For example, area models of fractions (such as a partitioned rectangle) are assumed to be lower in demands for memory resources than set models (such as an array of red and blue chips) because the whole is more salient in the area model and the nonselected parts (e.g., the nonshaded segments) are less salient. English and Halford's claims once again reinforce the importance of making part-whole (or class-inclusion) relations salient or transparent.

Individual differences in working memory have been associated with performance on fraction tasks (e.g., Hecht et al., 2003; see Hecht et al., 2007, for a review). Effects of working memory were independent of effects of conceptual knowledge, which means that both factors are important and neither can be reduced to the other. Specifically, in the Hecht et al. (2003) study, individual differences in *working memory* were assessed with a counting-span task and individual differences in *conceptual knowledge* were assessed with tasks involving providing numerical representations of pictorially presented fractions and vice versa, providing a pictorial representation of the sum of two pictorially-represented fractions, or identifying the larger of two numerically-represented fractions. The effects of working memory on fraction computation were mediated by differences in fifth-graders' mastery of basic arithmetic facts (assessed by measures of accuracy and speed of retrieving basic addition and multiplication facts). That is, working memory uniquely contributed to variability in basic arithmetic knowledge (e.g., direct retrieval requires less working memory resources than counting to solve basic problems), and basic arithmetic, in turn, influenced fraction computation. Working memory directly influenced the solution of word problems, without any mediation through effects of basic arithmetic knowledge or conceptual knowledge (and when factors such as reading ability were also controlled for). Working memory predicted accuracy at, for example, setting up or translating word problems, as the Task Group elaborates in the section on Algebra. Although working memory is described as an individual difference, that does not mean that it cannot be changed or that strategies cannot be learned that make the most of whatever capacity an individual has, so that performance surpasses that of individuals with greater basic capacity. For example, strategies such as chunking (recoding a multidimensional concept into fewer dimensions) or segmentation (breaking a task into a series of steps, each of which is not too resource demanding) can reduce the working-memory demands of a task (Sophian, 2007). Moreover, conceptual knowledge carried the greatest weight in predicting performance on all three outcome measures (computation, estimation, and word problems), whereas working memory only affected word problems and only indirectly affected computation through knowledge of basic arithmetic facts (see Table 4, p. 290; Hecht et al., 2003).

In sum, despite evidence of early appreciation of part-whole relations prior to formal schooling, children lack sufficient conceptual knowledge of conventional fractions, which is a stumbling block to performance on such fraction tasks as estimation, computation, and word problems. Conceptual knowledge is assessed using a variety of tasks, such as judging equivalence or rank ordering quantities according to magnitude, but it should be pointed out that these tasks do not tap identical competence; tasks such as rank ordering decimals and fractions may be harder than judging equivalence (Mazzocco & Devlin, in press). When students do not understand conventional fraction notation, they will often generalize inappropriately from whole number counting to fractions. However, they seem to have a rudimentary understanding of infinite divisibility, so the generalization from counting has exceptions, and they can build on intuitions about part-whole relations. Different representational formats, such as pictures and fraction names that separate parts and wholes, allow those intuitions to be tapped to support better performance, prior to explicit instruction. Intuitive versions of proportional reasoning problems are solved earlier, are easier, and improve performance on subsequent numerical problems. Even complicated operations, such as division, seem to be supported by earlier kinds of knowledge, for example, about sharing. Representations that make part-whole relations salient or transparent, in contrast to making part-part relations salient, improve performance across tasks and age groups. Effects of different representational formats (e.g., discrete objects vs. portions of shapes) on more advanced problem solving, such as adding improper and mixed fractions, has yet to be definitively determined. Among other differences, students with low working memory capacity are less able to bring arithmetic facts to mind quickly and automatically, without drawing on mental resources (e.g., counting to solve basic problems) that could be used for other aspects of complex problem solving, compared to typically achieving students (Hecht et al., 2007). Despite the ubiquity of differences in working-memory capacity for low- and high-achieving groups in studies of mathematics learning, however, recent reviews of the literature on fractions assign greater weight to lack of conceptual knowledge in accounting for performance (e.g., Hecht et al.; Sophian, 2007). Conceptual knowledge has been shown to promote procedural knowledge (or computational ability), and vice versa, and development progresses iteratively, with gains in conceptual and procedural knowledge reinforcing, and bootstrapping, one another.

e. Conclusions and Recommendations

A basic interpretation of a fraction is a part-whole relation of two or more values, although there are other interpretations of fractions. Fractions can be represented as proper fractions (e.g., $\frac{1}{8}$), mixed numbers (e.g., $2\frac{1}{8}$), or in decimal form (e.g., 0.8), but are often represented using pictures during early instruction. Difficulty with fractions is pervasive and is an obstacle to further progress in mathematics, and, thus, is likely to constrain achievement in science and pursuit of scientific careers (e.g., Sadler & Tai, 2007). The inability to understand and compute fractions, decimals, and proportions has important real-life implications, and has been linked to poor health outcomes, among other harmful effects.

Classroom

The learning of arithmetic facts provides a foundation for learning fractions. Committing such facts to memory reduces working memory demands of problem solving and thus allows attention to be focused on other problem features. Therefore, children should begin fraction instruction with the ability to quickly and easily retrieve basic arithmetic facts. Instruction focusing on conceptual knowledge of fractions is likely to have the broadest and largest impact on problem-solving performance (provided that it is aimed at accurate solution of specific problem types that tap conceptual knowledge). Procedural knowledge is also essential, however, and although it must be learned separately, is likely to enhance conceptual knowledge and vice versa. Successful interventions reported in the scientific literature could be transferred easily to classrooms. These interventions include using fraction names that demarcate parts and wholes, using pictorial representations that are mapped onto the number line, and linking composite representations of fractions to representations of the number line. Conceptual and procedural knowledge about fractions less than one do not necessarily transfer to fractions greater than one (i.e., improper and mixed fractions), and must be separately instructed. Appropriate intuitions about sharing, part-whole relations, and proportional relations can be built on in classrooms to support acquisition of conceptual and procedural knowledge of fractions.

Training

Teachers. Training of teachers should include sufficient coverage of the scientific method so that teachers are able to critically evaluate the evidence for proposed pedagogical approaches and to be informed consumers of the scientific literature (who can keep up with advances in scientific knowledge after graduation from training programs). Teachers should be aware of common conceptions and misconceptions involving fractions, based on the scientific literature, and of effective interventions involving fractions. Thus, training should include comprehensive courses on cognitive development focusing on mathematics learning that draw on the primary literature in this area (i.e., refereed journal articles).

Future researchers. Many of the best researchers in the basic science of mathematics learning are currently not engaged in directly relevant educational research. New funding should be provided to train future researchers, to begin new interdisciplinary degree programs with rigorous quantitative training, and to establish support mechanisms for career shifts for rigorous researchers that are similar to K awards from the National Institutes of Health.

Curriculum

The curriculum should allow for sufficient time on task to ensure acquisition of conceptual and procedural knowledge of fractions and of proportional reasoning, with the goal for students being one of learning rather than performance. However, there should be ample opportunity in the curriculum for accurate self-evaluation. The curriculum should include representational supports that have been shown to be effective, such as number line representations, and encompass instruction in tasks that tap the full gamut of conceptual and procedural knowledge, such as ordering fractions on a number line, judging equivalence and relative magnitudes of fractions with unequal numerators and denominators, estimation, computation, and word problems. The curriculum should make explicit connections between intuitive understanding and formal problem solving.

Research

Basic. Studies suggest that preschool and early elementary-school children have a rudimentary understanding of very simple fractional relations, but the mechanism underlying this knowledge is not yet known. The relation between this informal, often implicit knowledge, and the learning of formal mathematical fractional concepts and procedures is not well understood, and is an area in need of further study. Similarly, the mechanisms that contribute to the emergence of formal competencies in school are not fully understood, but involve a combination of instruction, working memory, and the bidirectional influences of procedural knowledge on the acquisition of conceptual knowledge and conceptual knowledge on the skilled use of procedures. Therefore, research is needed that tests specific hypotheses designed to uncover these mechanisms, including linking earlier intuitive understanding with later formal problem solving. In addition, research on understanding and learning of fractions should be integrated with what is known and with emerging knowledge in other areas of basic research, such as neuroscience, cognition, motivation, and social psychology. Research on mental representations and retrieval in memory, as well as on intuitive versus analytical reasoning, are especially relevant and currently not integrated with research on fractions. Ironically, the absence of a coherent and empirically supported theory of fraction tasks (i.e., how tasks are related to one another in terms of underlying processes) is a major stumbling block to developing practical interventions to improve performance in this crucial domain of mathematics. Such a theory would, for example, provide scientific guidance concerning how instruction in different fraction tasks should be ordered.

Classroom. Classroom-relevant research need not be conducted physically in classrooms, and constraints on funding that require that relevant research be performed in classrooms should be removed. Conversely, many interventions demonstrated to be effective in experiments should be scaled up and evaluated in classrooms. In order to produce a steady supply of high-quality research that is relevant to classroom instruction, a pipeline of research must be funded that extends from the basic science of learning to field studies in classrooms. Incentives should also be provided to encourage partnerships between basic and applied researchers, and to support research that includes both laboratory and field-based research, in a way that will provide converging operations.

3. Estimation

Estimation is an important part of mathematical cognition, one that is pervasively present in the lives of both children and adults. Consider just a few everyday examples. How many people were at the game? How fast can that Lamborghini go? About how much is 65 × 29? Estimation may be used more often in everyday life than any other quantification process.

In addition to its pervasive use, estimation is quite strongly related to other aspects of mathematical ability, such as arithmetic skill and conceptual understanding of computational procedures, and to overall math achievement test scores (Booth & Siegler, 2006; Dowker, 2003; Hiebert & Wearne, 1986; LeFevre, Greenham, & Waheed, 1993; Geary, Hoard, Byrd-Craven, Nugent, & Numtee, 2007). It usually requires going beyond rote application of procedures and applying mathematical knowledge in flexible ways. This type of adaptive problem solving is a fundamental goal of contemporary mathematics education.

Yet another basis of the importance of estimation is practical—most school age children are surprisingly bad at it and even many adults are far from good. Standardized scores on the part of the NAEP that tests estimation proficiency are below those for the mathematics test as a whole (Mitchell, Hawkins, Stancavage, & Dossey, 1999). This limited proficiency, together with the pervasiveness of estimation in everyday life, its relation to general mathematical ability, and its embodying the type of flexible problem solving that is viewed as crucial within modern mathematics education, have led the National Council of Teachers of Mathematics (NCTM) to assign a high priority to the goal of improving estimation skills within each revision of its Math Standards since 1980 (e.g., NCTM, 1980, 2000), as well as in its recent *Focal Points* (NCTM, 2006).

Despite the importance of estimation both in and out of school, far less is known about it than about other basic quantitative abilities, such as counting and arithmetic. One reason for the discrepancy is that estimation includes a varied set of processes rather than a single one. Some estimation tasks, for example estimating the distance between two cities or the cost of a bag of groceries, require knowledge of measurement units such as miles or dollars. Other estimation tasks, for example estimating the number of coins in a jar or the answers to arithmetic problems, do not. Similarly, some uses of estimation, for example estimating the cost of a pizza or the speed of a Lamborghini, require prior knowledge of the entities whose properties are being estimated (i.e., pizzas, Lamborghinis). Other uses, such as estimating the length of a line on a page or the number of fans at a game, do not.

In this discussion, the Task Group focuses on *numerical estimation*, the process of translating between alternative quantitative representations, at least one of which is inexact and at least one of which is numerical. This category includes many prototypic forms of estimation. For example, computational estimation involves translating from one numerical representation (e.g., 75 × 29) to another (about 2,200). Number line estimation either requires translating a number into a spatial position on a number line (e.g., given: 0____________100, place a mark on the line where 71 would fall) or translating a spatial position on a number line into a number. Numerosity estimation requires translating a nonnumerical quantitative representation (e.g., a visual representation of the approximate volume and density of candies in a jar) into a number (e.g., about 300 marbles.) Because this task group's focus is on the learning of mathematics, excluded from consideration are tasks that require knowledge external to mathematics, in particular knowledge of measurement units (e.g., pounds, hours, miles) or real-world entities (e.g., population of Russia, number of people with AIDS). We also exclude from consideration trivial applications of estimation, such as rounding to the nearest 10, which unfortunately are the predominant focus of instruction in estimation in many U. S. classrooms.

a. Understanding the Goals of Estimation

Many children have highly distorted impressions of the goals of estimation, especially the goals of computational estimation. As noted by LeFevre et al. (1993), accurate computational estimation requires understanding of the simplification principle (the understanding that mental arithmetic is easier with simple operands) and the proximity principle (the understanding that the main aim of estimation is to obtain estimates close in magnitude to the correct answer). LeFevre et al. found that fourth- and sixth-graders understood the principle of simplification, but they showed little if any understanding of the importance of generating an estimate close in magnitude to the correct answer. When asked to define estimation, most said that it was "guessing" or indicated that they did not know. When asked to estimate the products of multidigit multiplication problems, only 20% of fourth-graders produced reasonable estimates (estimates that varied systematically with the product).

Sowder and Wheeler (1989) found that even middle and high school students typically do not understand that the goal of estimation is to generate estimates that are close to the correct value, rather than following some prescribed procedure. They based this conclusion on the reluctance of even ninth-graders to accept that both of two alternative estimates could be acceptable and on their infrequent use of compensation to correct for distortions introduced by rounding. When asked to generate estimates, some students went as far as to calculate the correct answer and then to round to a nearby number. The problem seemed to be that the children viewed estimation as a rigid algorithmic procedure that required following preset rounding rules rather than as a flexible attempt to approximate the magnitude of an answer using whatever means made sense in the particular situation. This blind execution of an algorithm is reflected in Sowder and Wheeler's observation that, "Some students in both Grades 5 and 7 objected to rounding 267 to 250 rather than 300, arguing, 'You're always taught to go up if it's past five,' or 'Seven is above five, so you have to go up, not down (p. 144).'"

On the other hand, Sowder and Wheeler (1989) also noted that by fifth grade, the large majority of children, when presented hypothetical estimation procedures in which rounding was or was not followed by compensation for the distortions introduced by rounding, recognized that rounding with compensation was superior. This finding suggests that some conceptual understanding of the importance of the proximity principle is present by fifth grade. Instruction in estimation clearly needs to convey to students earlier and more consistently that the purpose of estimation is to generate values close in magnitude to the correct value.

b. Development of Estimation Skills

Computational estimation

Development of computational estimation skills (the ability to answer an arithmetic problem with the goal of approximating the correct magnitude rather than calculating the exact answer) begins surprisingly late and proceeds surprisingly slowly. In one study, more than 75% of third- and fifth-graders did not agree that two alternative estimates of the sum of two addends could both be acceptable (Sowder & Wheeler, 1989). Similarly, Dowker (1997) found that early elementary school children often could perform exact computations in a numerical range but could not estimate answers in the same range. Thus, 7- and 8-year-olds who were able to compute the correct answer on problems with sums less than 100 failed to generate reasonable estimates (estimates within 30% of the correct answer) on 33% of these problems.

Computational estimation does improve considerably, albeit gradually, with age and experience. Adults and sixth-graders are more accurate than fourth-graders in estimating the sum of 2 three-digit addends (Lemaire & Lecacheur, 2002), sixth- and eighth-graders are more accurate than fourth-graders in estimating the sums of long strings of addends (Smith, 1999), and fourth-graders are more accurate than second-graders in estimating the sums of two-digit addends (Booth & Siegler, 2006). Similarly, adults are more accurate than eighth-graders, who in turn are more accurate than sixth-graders, in estimating the products of multidigit multiplication problems (LeFevre et al., 1993). Improvements in the speed of estimation of the answers to both addition and multiplication problems follow a similar course to improvements in accuracy over the same age range (Lemaire & Lecacheur).

From early in the development of computational estimation, individual children use a variety of strategies (Reys, 1984). Evidence for such strategic variability comes both from observations of ongoing behavior and from immediately retrospective self-reports (LeFevre et al., 1993; Sowder & Wheeler, 1989). The following is a list of some of the most common estimation strategies for addition and multiplication (Dowker, Flood, Griffiths, Harriss, & Hook, 1996; LeFevre et al.; Reys et al., 1982; Reys et al., 1991; Sowder & Wheeler):

1) *Rounding*: Converting one or both operands to the closest number ending in one or more zeroes (e.g., on 297 × 296, both multiplicands might be converted to 300).
2) *Truncating*: Changing to zero one or more digits at the right end of one or more operands (e.g., on 297 × 296, both multiplicands might be converted to 290).
3) *Prior compensation*: Rounding the second operand in the opposite direction of the first before performing any computation (e.g., on 297 × 296, 296 might be rounded to 290 rather than 300 to compensate for the effect of rounding 297 to 300).

4) *Postcompensation*: Correcting later for distortion introduced by earlier rounding or truncation (e.g., on 297 × 296, multiplying 300 × 300 and then subtracting 2% of 90,000).
5) *Decomposition*: Dividing numbers into simpler forms (e.g., on 282 × 153, multiplying 280 by 100 and then by 1.5).
6) *Translation*: Simplifying an equation (e.g., by changing the operation, on 44 + 53 + 51 + 47, multiplying 50 × 4).
7) *Guessing.*

As might be expected, some of these strategies are used more often than others. Rounding is the most common approach (Lemaire et al., 2000; LeFevre et al., 1993; Reys et al., 1982; Reys et al., 1991). Compensation tends to be among the least frequent approaches, although it is among the most useful. For example, in Lemaire et al.'s study of estimation of multidigit sums, fifth-graders used rounding on 64% of trials and compensation on 2%.

This use of multiple strategies is not a result of individuals using only one approach but differing in what that approach is. Instead, both children and adults often know and use a variety of computational estimation strategies. This is especially true among mathematically sophisticated individuals. For example, Dowker et al. (1996) examined the multiplication and division estimates of four groups of adults: mathematicians, accountants, and students majoring in psychology or English at Oxford University. The strategies that they used were remarkably diverse: For example, the 176 participants used 27 different strategies for solving the single problem 4645 ÷ 18. Individuals in each of the four groups averaged more than five strategies apiece. Strategic variability was evident even within a single person solving the same problem on two occasions. When problems were presented to participants a second time, mathematicians used a different strategy on 46% of items and psychology students on 37%.

Both children and adults adapt their strategy choices to problem characteristics. One form that this adaptation takes is to use rounding more often on problems where it introduces less distortion. For example, on multidigit addition problems, the closer an addend is to the nearest 10, and therefore the less distortion introduced by rounding, the more often fourth- and sixth-graders round (Lemaire & Lecacheur, 2002). Similarly, on multidigit multiplication problems, sixth-graders, eighth-graders, and adults more often round both of the multiplicands when each includes two or three digits, but often only round the larger multiplicand when the smaller one is a single digit (LeFevre et al., 1993). This choice pattern minimizes distortion, because rounding two or three digit multiplicands to the nearest 10 changes the product by a smaller percentage than rounding single-digit multiplicands to the nearest 10.

The range and appropriateness of computational estimation strategies increase with age and mathematical experience. Adults use a considerably greater variety of multiplication strategies than do sixth- or eighth-graders (LeFevre et al., 1993). Similarly, mathematicians and accountants, who have unusually extensive numerical experience, use a greater variety of appropriate estimation strategies than do even the highly selected psychology and English students at Oxford University (Dowker et al., 1996). The latter two groups used a greater variety of inappropriate estimation strategies than did the former two, which indicates that ability to generate *appropriate* variants is what distinguishes the mathematicians and accountants, rather than greater variation per se.

A second type of change in strategy use involves the sophistication of the strategies that are used. Use of compensation, a strategy that requires a good conceptual understanding of estimation, shows especially substantial growth. In estimating the answers to multidigit addition problems, far more ninth-graders use post-compensation than do third- or fifth-graders (Lemaire et al., 2000; Sowder & Wheeler, 1989). The quality of strategy choices in multidigit multiplication also increases with age and mathematical experience. LeFevre et al. (1993) provided the example of strategy choices on 11 × 112. Among adults, 75% rounded the problem to 10 × 112, a computationally tractable approach that yields an answer within 9% of the correct answer. Although this approach would seem well within the capabilities of sixth-graders (LeFevre et al.), no sixth-grader used it. Instead, they rounded either to 10 × 100 or to 10 ×110.

All of the studies reviewed in this section involve computational estimation with whole numbers. Far less is known about computational estimation with fractions. Two findings that have emerged are that even high school students are very poor at computational estimation with fractions and that the main problem seems to be inadequate conceptual understanding of the magnitudes of fractions (Carpenter, Corbitt, Kepner, Lindquist, & Reys, 1981; Hecht et al., 2007). When more 13- and 17-year-olds estimate that $\frac{12}{13} + \frac{7}{8}$ is roughly equal to 19 than estimate that it is roughly equal to 2, there clearly is a serious problem in their understanding of the relation between fractional notation and the magnitudes that are being estimated (Carpenter et al.).

Number line estimation

The number line task has proved highly informative, not only for improving understanding of estimation but also for providing useful information about children's understanding of the decimal number system more generally. On this task, children are presented a line with 0 at one end, another number such as 100 or 1,000 at the other end, and no other numbers or hatch marks in between. The child is presented a new line and number to be estimated on each trial, until the child has estimated the magnitudes of numbers throughout the range (e.g., 0–1,000). Then each estimate is translated into a numerical value, and the relation between the number presented and the estimate is examined for the full set of numbers. Ideally, the estimated value should increase linearly with the actual value in a 1:1 fashion, in accord with the equation $y = x$. Thus, on a number line with 0 at one end and 1,000 at the other, the estimate for 20 should be 2% of the way between 0 and 1,000, the estimate for 230 should be 23% of the way, the estimate for 760 should be 76% of the way, etc.

Although this task seems easy, elementary school children's estimates consistently depart from correct values in predictable ways. Moreover, similar departures from correct estimates are seen at the same ages on other types of estimation tasks; the deviations are indicative of broader difficulties with mathematics. This number-line task and findings have inspired an educational intervention that succeeded in improving a broad range of numerical skills in low-income preschoolers (Ramani & Siegler, 2008).

Number line estimation improves steadily during the elementary school years, with accuracy at any given age being greater on smaller numerical scales than on larger scales. On 0–10 number lines, Petitto (1990) found that the percent of absolute error decreased from 14% late in first grade to 4% late in third grade. On 0–100 number lines, the same children's percent of absolute error decreased from 19% late in first grade to 8% late in third grade. On 0–1,000 number lines, Siegler and Opfer (2003) found that percent absolute error improved from 21% in second grade to 14% in fourth grade, 7% in sixth grade, and 1% in adulthood. The number line estimates of children from low-income backgrounds and children with learning disabilities in mathematics are far less accurate than those of typically achieving children from middle-income families, though they also improve with age (Geary, Hoard, Nugent, & Byrd-Craven, in press; Siegler & Ramani, in press). The superiority of the estimates on the smaller scale (0–100) indicates that at least through fourth grade, children use their knowledge of particular numbers, rather than general understanding of the decimal system, to estimate.

Children use two primary mental representations of numerical magnitude on number line estimation tasks. One common approach (the correct one with whole numbers) is to use a linear representation, that is, a representation in which numerical magnitude increases linearly with the size of the number. Another common approach is to employ a logarithmic representation, in which representations of numerical magnitudes increase logarithmically with numerical size. When children use such a logarithmic representation on a number line estimation task, the spatial positions they choose increase very quickly in the low range of numbers and then increase only slowly in the upper part of the range. For example, Siegler and Opfer (2003) found that on 0–1,000 number lines, differences between second-graders' estimates for 5 and 86 were much larger than the differences between their estimates for 86 and 810. Such logarithmic representations of quantities and other magnitudes are common across many species and tasks (Dehaene, 1997), and for good reason: To a hungry animal, the difference in importance between 5 and 86 pieces of food often is far larger than the difference between 86 and 810 pieces. This is not the case within the formal number system, however.

With age and experience, children progress from using the less accurate logarithmic representation to the more accurate linear one on the number line task. For example, kindergartners' median estimates on 0–100 number lines are better fit by the logarithmic function than by the linear function, first-graders' estimates are fit equally well by the two functions, and second-graders' estimates are better fit by the linear function (Geary et al., in press; Laski & Siegler, in press; Siegler & Booth, 2004). The same progression is seen in children with learning disabilities in mathematics, but it occurs more slowly and at older ages (Geary et al.). On 0–1,000 number lines, second-graders' estimates are fit better by the logarithmic function, whereas fourth-graders' are fit better by the linear function (Booth & Siegler, 2006; Opfer & Siegler, 2007). The same child often uses different representations depending on the scale of numbers they are asked to estimate. For roughly half of the second-graders in Siegler and Opfer (2003), the best fitting function for number line estimates was linear on the 0-100 line but logarithmic on the 0–1,000 line.

Number lines can be used to examine estimates of the magnitudes of fractions as well as whole numbers. Results from such studies, like studies of computational estimation with fractions, show poor understanding of fractional magnitudes at all ages. Fifth- and sixth-graders' estimates of the magnitudes of decimal fractions do not even maintain the correct rank order (Rittle-Johnson et al., 2001). Even many adults in the United States have poor understanding of the magnitudes of common fractions. Opfer et al. (2007) found that most adults estimate the magnitudes of common fractions with numerators of "1" as a linear function of the distance between their denominators (even though the magnitudes of such fractions actually follow a logarithmic function). For example, adults estimate the magnitudes of $\frac{1}{1}$ and $\frac{1}{60}$ to be much closer than those of $\frac{1}{60}$ and $\frac{1}{1440}$), even though the magnitudes of the fractions in the initial pair are more than 60 times as discrepant.

c. Individual Differences in Estimation

Both children and adults show substantial individual differences in skill at computational estimation (Dowker, 2003) that are associated with broader individual differences in mathematical understanding. Proficiency at computational estimation correlates positively, and often substantially, with mathematics SAT scores (Paull, 1972), mathematics achievement test scores (Booth & Siegler, 2006; Siegler & Booth, 2004), performance on other estimation tasks (Booth & Siegler), and arithmetic fluency scores (Dowker, 1997, 2003; LeFevre et al., 1993).

Accuracy and linearity of number line estimation also is highly associated with general mathematical ability. Significant and substantial correlations—typically between r = .50 and r = .60—have been found between mathematics achievement test scores and linearity of number line estimates among kindergartners, first-graders, and second-graders on 0–100 number lines (Geary et al., in press; Siegler & Booth, 2004) and among second-, third-, and fourth-graders on 0–1,000 number lines (Booth & Siegler, 2006). Individual differences in linearity of number line estimates also are closely associated with individual differences in linearity on other estimation tasks (Booth & Siegler). These results suggest that performance on a variety of estimation tasks reflects a common underlying representation of numerical magnitude and that the closer this representation is to the formal linear mathematical system the better the overall mathematics achievement.

d. Improving Children's Estimation

Findings with number line estimation have raised the question: What leads children to shift from logarithmic to linear representations of numerical magnitude? One common activity that seems likely to contribute is playing board games with linearly arranged, consecutively numbered, equal-size spaces (e.g., Chutes and Ladders©). Such board games provide multiple cues to both the order of numbers and the numbers' magnitudes. In the games, the greater the number in a square, the greater a) the distance that the child has moved the token, b) the number of discrete moves the child has made, c) the number of number names the child has spoken, d) the number of number names the child has heard, and e) the amount of time since the game began. The linear relations between numerical magnitudes and these visuospatial, kinesthetic, auditory, and temporal cues provide a broadly based, multi-modal foundation for a linear representation of numerical magnitudes.

To determine whether playing number board games produces improvements in numerical understanding, Siegler and Ramani (in press) and Ramani and Siegler (2008) randomly assigned preschoolers at Head Start centers, all of whom came from low-income families, to play one of two board games. The games differed only in the board that children encountered. One board included linearly arranged, equal-spaced squares that progressed from 1–10 from left to right. The other board was identical except for the squares varying in color rather than number. Each child played the number board game or the color board game with an experimenter for four 15-minute sessions within 2 weeks.

Playing the numerical board game for this 1 hour period increased the Head Start children's proficiency not only at number line estimation but also at three other key numerical skills: counting, identifying printed numerals, and comparing the relative sizes of numbers. The gains in all four skills remained when children were tested nine weeks after the game playing experience. Gains were comparable for African-American and white children; they also were comparable for children who, relative to their low-income peers, entered the game with more or less numerical knowledge. Classmates who played the color board version of the game did not improve on any of the skills. The effect sizes of differences between the groups were substantial: *d*'s between .69 and 1.08 on the four measures on the immediate posttest and between .55 and .80 on the 9-week follow-up.

Ramani and Siegler (2008) also tested whether board game experience in the everyday environment is related to numerical knowledge and whether it might contribute to the knowledge differences between children from low- and middle-income backgrounds. They asked children from the initial experiment, as well as age peers from middle-income backgrounds, about their experience playing board games, card games, and video games at their own and other people's homes. The children from middle-income homes reported having more experience playing board games and card games in both contexts (though less experience playing video games). Of particular interest, the number of board games that the Head Start children reported playing at their own and other people's homes correlated positively with their skill at all four numerical tasks examined in the study. In contrast, the preschoolers' experience playing card games and video games was only minimally related to their numerical knowledge. Thus, playing numerical board games appears to be a promising (and inexpensive) way to improve low-income preschoolers' numerical knowledge and to reduce discrepancies in the numerical knowledge that children from low- and middle-income homes bring to school.

A different procedure has been found effective for improving elementary school children's number line estimation. By second grade, a large majority of children generate linear representations of magnitudes in the 0–100 range but logarithmic ones in the 0–1,000 range. Opfer and Siegler (2007) reasoned that a dramatic error of a number line estimate in the 0–1,000 range might lead children to search for an alternative approach, that their representations of numbers in the 0–100 range provided such an alternative, and that the children would draw the analogy to the 0–100 range and quickly improve their estimates in the 0–1,000 range. This proved to be the case. Providing the second-graders with feedback on their estimate of the single number 150—the number where the logarithmic and linear functions are most discrepant—led 80% of the children to shift from a logarithmic to a linear approach after that single feedback problem. Almost all of these children continued to use the linear approach on all subsequent trials. Thus, feedback on well-chosen problems is another means of improving children's estimation.

Improving elementary school children's numerical representations also can improve their skill at learning arithmetic. Presenting first-graders with accurate number line representations of the magnitudes of addends and sums enabled the children to recall the correct answer more often than children who were told the correct answer but were not presented the number line representations (Booth & Siegler, in press). Providing the number line representations also led to errors being closer in magnitude to the correct answer. Thus, numerical magnitude representations influence learning of arithmetic as well as a variety of other numerical skills and knowledge.

e. Conclusions and Recommendations

Numerical estimation is an important part of mathematical cognition. It is used frequently by both children and adults, in both academic and nonacademic contexts; is closely related to arithmetic skill, conceptual understanding of computational operations, and mathematics achievement test performance; and receives a considerable amount of attention in elementary school mathematics textbooks and classroom instruction. Moreover, estimation performance often reveals both subtle and gross deficiencies in numerical understanding.

Despite its importance and the substantial attention that it receives, most children's proficiency at estimation is poor. This in part reflects the emphasis in many classrooms on rounding procedures, to the exclusion of conceptually richer aspects of estimation, such as compensating for the distortions introduced by rounding. Many students do not even know that the goal of estimation is to generate values that are close to the correct value or that there is often more than one reasonable estimation procedure.

From kindergarten or first grade onward, most children's estimates of the magnitudes of whole numbers accurately reflect the rank order of the numbers. However, children from low-income backgrounds often do not even know the rank order of the numbers 0–10 when they enter school. Proficiency develops first in the 0–10 and 0–100 ranges, and then in the 0–1,000 and larger ranges. However, many elementary school students fail to discriminate adequately among the magnitudes of numbers in the hundreds or thousands.

Studies of estimation of the magnitudes of fractions show little if any understanding, even among middle school and high school students. Estimates often do not even maintain the rank order of the fractions' magnitudes. There is a strong need to develop effective procedures for remedying most students' lack of understanding of fractional magnitudes.

Classroom

Teachers should broaden instruction in computational estimation beyond rounding. They should insure that students understand that the purpose of estimation is to approximate the correct value and that rounding is only one of several means for accomplishing this goal.

Teachers should provide examples of alternative procedures for compensating for the distortions introduced by rounding, should emphasize that there are many reasonable procedures for estimating rather than just a single correct one, and should discuss reasons why some procedures are reasonable and others are not.

Teachers in Head Start and other facilities serving preschoolers from low-income backgrounds should be made aware of the usefulness of numerical board games for improving the children's knowledge of numbers and of the importance of such early knowledge for long-term educational success.

Teachers should not assume that children understand the magnitudes represented by fractions even if the children can perform arithmetic operations with them, because the arithmetic competence may only represent execution of memorized procedures. Examining children's ability to perform novel estimation tasks, such as estimating the positions of fractions on number lines, can provide a useful tool for assessing children's knowledge of fractions. Providing feedback on such number line estimates can improve children's knowledge of the fractions' magnitudes.

Training

Teachers in preservice and in-service programs should be informed of the tendency of elementary school students not to fully understand the magnitude of large whole numbers, should be taught how to assess individual students' understanding, and should be taught research-based techniques for improving the children's understanding.

Teachers should be made aware of the inadequate understanding of the magnitudes of fractions of elementary school, middle school, and high school students. The teachers also should be familiarized with the usefulness of feedback on number line estimates of the magnitudes of fractions for overcoming these difficulties.

Curriculum

Textbooks need to explicitly explain that the purpose of estimation is to produce accurate approximations. Illustrating multiple useful estimation procedures for a single problem, and explaining how each procedure achieves the goal of accurate estimation, is a useful means for achieving this goal. Contrasting these procedures with others that produce less accurate estimates, and explaining why the one set of procedures produces more accurate estimates than the other, is also likely to be helpful.

Research

Research is needed regarding simple instruments that teachers can use in the classroom for assessing children's estimation skills, and regarding instruction that can efficiently improve children's estimation.

Research is needed on how the inadequate representations of whole number numerical magnitudes that have been identified by studies of estimation influence learning of other mathematical skills, such as arithmetic.

Research is needed on how children can be taught to accurately estimate the magnitudes of fractions and on how learning to estimate those magnitudes influences acquisition of other numerical skills involving fractions, such as arithmetic and algebra.

Research is needed on how estimation is used by students (e.g., to solve complex problems, to improve test performance) and by adults in everyday life and professional tasks (e.g., to rule out implausible answers and thus reduce human error).

4. Geometry

Geometry is the branch of mathematics concerned with properties of space, and of figures and shapes in space. Euclidean geometry is the domain typically covered in mathematics curricula in the United States, although a separate year-long course is not usually taught until high school. Units on geometry as well as measurement are frequently included in middle school mathematics classes, whereas only the latter tends to be emphasized in the elementary grades.

a. Geometry Performance of U.S. Students on International Mathematics Assessments

Although geometric concepts and skills are typically taught in both elementary and middle school classrooms in the United States, international assessments indicate that the achievement levels of U.S. students are comparatively poor in this mathematical domain. To begin with, the 2003 Trends in International Mathematics and Science Study (TIMSS) showed no significant improvement in geometry for U.S. eighth-graders between 1999 and 2003, despite significant gains in algebra during this same time period (Gonzales et al., 2004). Moreover, of the five mathematical content areas assessed by TIMSS (number, algebra, geometry, measurement, and data), U.S. eighth-graders' performance in geometry items was weakest (Mullis et al., 2004).

Similarly, a report from the American Institutes for Research (Ginsburg et al., 2005) reexamined the 2003 mathematics performance of U.S. students on the TIMSS fourth- and eighth-grade assessments, as well as the Program for International Student Assessment (PISA)—relative to a common set of 11 other countries which had also participated in these studies (including Australia, Hong Kong,[1] Japan, and New Zealand, among others). The content areas that were evaluated included 1) number/quantity, 2) algebra/change and relationships, 3) measurement, 4) geometry/space and shape, and 5) data/uncertainty. The United States ranked 8th, 9th, and 9th out of the 12 countries on the TIMSS-4, TIMSS-8, and PISA, respectively. And again the performance levels of U.S. students were found to be significantly weakest in the area of measurement in Grade 4 and in geometry in Grade 8, as compared against the average U.S. score across all content areas. Furthermore, the United States was found to devote only about half as much time to the study of geometry as the other countries.

[1] Hong Kong is a Special Administrative Region (SAR) of the People's Republic of China.

b. Importance of Geometry for Learning Algebra

Given that the primary charge to the National Mathematics Advisory Panel concerns preparation for and learning of algebra, one may ask what geometry has to do with the acquisition of algebraic concepts and skills? Moreover, as the teaching of high school level geometry usually follows the first course in algebra, what if any geometric concepts should be learned in the middle school years, if not earlier, to ensure that students are best prepared to acquire a thorough understanding of key algebraic concepts and expressions? As noted in the Conceptual Knowledge and Skills Task Group report, the single aspect of geometry that is most directly relevant for early learning of algebra is that of similar triangles. In particular, the proof that the slope of a straight line is independent of the two points selected depends logically on considerations of the properties of similar triangles. This is because the corresponding angles of similar triangles are congruent and their corresponding sides are proportional. Therefore, the Conceptual Knowledge and Skills Task Group contends that it is crucially important for students to be given the opportunity to acquire these and other essential facts about similar triangles prior to the formal study of algebra. Furthermore, they point out that whereas students do not need to learn to construct the proofs of these theorems until they take a course in Euclidean geometry, they should nonetheless be able to make use of them.

Consistent with this perspective, the NCTM's (2006) *Focal Points* underscores (as do some state frameworks) the importance of these ideas in its section on algebra and connections to geometry for eighth-graders: "Given a line in a coordinate plane, students should understand that all 'slope triangles' triangles created by a vertical 'rise' line segment (showing the change in y), a horizontal 'run' line segment (showing the change in x), and a segment of the line itself—are similar. They also [should] understand the relationship of these similar triangles to the constant slope of a line" (p. 20).

What are the essential aspects of similar triangles? Acknowledging the need for learning how the relations between various properties of triangles underlie the fact that the slope of a straight line is independent of the two points selected, the question arises as to what kinds of concepts students need to acquire to understand the "basic aspects" of similar triangles? Certainly, to comprehend that the corresponding sides of similar triangles are proportional requires at minimum an understanding of length, equal angles, right triangles, and correspondence, as well as the crucial concepts of ratio and proportion. At this point, the Task Group notes that the difficulties associated with acquiring a sound conceptual understanding of ratio and proportion in and of themselves (as outlined in the section on Fractions in the this report) clearly constitute a significant obstacle to mastering how the slope of a straight line is derived from the properties of similar triangles.

Moreover, some additional difficulties may arise from the way in which the concept of similarity is often defined for students in school mathematics. For example, Wu (2005) has argued that rather than defining the similarity of figures as "same shape but not necessarily the same size," the most mathematically accurate and potentially effective way to define it is two figures are similar if one figure is congruent to a dilated version of the other. Naturally, understanding this definition would necessitate learning the meanings of congruence and dilation. Although a common way of defining congruence in school mathematics is "same size and same shape," Wu contends that a more mathematically correct and grade-level appropriate (i.e., for middle school students) definition is a composition of translations, reflections and rotations. It follows that to make sense of this definition students would first have to learn the meanings of these various transformations of the plane (more commonly referred to as slides, flips and turns, respectively). Furthermore, students would have to learn the meaning of dilation—a transformation of the plane that expands (or contracts) all points away from (or toward) a central point by a common scale factor. This mathematically accurate definition is clearly rather complicated in comparison with the more commonly used definition: a transformation that changes a figure's size, while its shape, orientation, and location remain the same. To sum up, in order to understand the mathematics underlying the proof that the slope of a straight line is independent of the choice of the points selected, students must successfully develop a conceptual understanding of the following: points, lines, length, angle, right triangle, correspondence, ratio, proportion, translation, reflection, rotation, dilation, congruence, and similarity.

c. Limitations of the Relevant Research-Based Literature

For the purposes of the present section, it is important to understand the developmental course that children take in learning the concepts that are required for understanding the properties of similar triangles. Whereas empirical studies of the key components of congruence, similarity, transformations, and so forth have indeed been conducted, it is difficult to draw firm, scientific conclusions from the relevant research literature. The reasons for this include, among others, 1) numerous studies of convenience samples with small numbers of participants, 2) the frequent use of a single age group or grade level, 3) the almost complete lack of longitudinal data, 4) an overemphasis on interview data and anecdotal reports, 5) a lack of rigor in study designs with comparatively limited use of experimental manipulations, and perhaps of greatest concern 6) a paucity of programmatic and cumulative efforts that could yield a clearer picture of the development of geometric thinking and reasoning. Thus for the most part, the Task Group is in agreement with Clements and Sarama's (2007a) conclusion following a recent extensive and intensive review of the relevant literature, with a focus on early childhood mathematics:

> Although far less developed than our knowledge of number, research provides guidelines for developing young children's learning of geometric and spatial abilities. However, researchers do not know the potential of children's learning if a conscientious, sequenced development of spatial thinking and geometry were provided throughout their earliest years. Insufficient evidence exists on the effects (efficacy and efficiency) of including topics such as congruence, similarity, transformations, and angles in curricula and teaching at specific age levels. Such research, and longitudinal research on many such topics, is needed (p. 517).

Nevertheless, the Task Group reviewed influential theories and literatures on children's geometric learning and provided directions for future research in this area.

d. Paths of Acquisition

Piaget's theory of spatial development. One of the earliest and most influential theories of the development of spatial and geometric concepts was put forth by Piaget and Inhelder (1967), who proposed that young children initially conceptualize space and spatial relations topologically as characterized by the following properties: proximity, order, separation, and enclosure. With development, children subsequently begin to represent space in a projective fashion, that is, in relation to different points of view, and then sometime between middle and late childhood the Euclidean conceptual system emerges permitting preservation of metric relationships such as proportion and distance.

Although numerous studies have been carried out to test the validity of this theory of "topological primacy," the consensus of investigators who have reviewed the empirical literature is that evidence supporting this developmental model is comparatively weak. One central criticism of this theory has been that Piaget and Inhelder's uses of terms such as *topological*, *separation*, and *proximity* are mathematically erroneous (Clements & Sarama, 2007a; Kapadia, 1974). For example, as Clements and Sarama point out, Piaget and Inhelder (1967) maintain that children do not synthesize the concepts of proximity, separation, order, and enclosure to construct the notion of continuity until the emergence of the formal operations stage (at approximately 11 or 12 years of age). However, in direct contrast to comprising a synthesis of these four properties, continuity is itself a central concept in topology (McCleary, 2006). And thus as Clements and Sarama note, the claim that this concept does not develop until early adolescence undermines the argument for the primacy of topological concepts (Darke, 1982; Kapadia). Concomitantly, these authors indicate that classifying figures as topological or Euclidean is problematic given that all figures possess attributes of both.

Clements and Sarama (2007a) conclude that although the empirical evidence does in fact suggest that the spatial abilities of children develop considerably throughout the school years, young children are more competent than hypothesized by Piaget and Inhelder as they can reason about spatial perspectives as well as distances. Indeed, Liben (2002) cites research suggesting that implicit Euclidean concepts are present perhaps as early as birth or soon thereafter, and that visual experience may not even be necessary for this system to develop. Furthermore, she points out Piaget himself subsequently replaced his topological primacy model with a different theory (i.e., intra, inter, and transfigural relations; Piaget & Garcia, 1989).

More recently, Dehaene, Izard, Pica, and Spelke (2006) tested adult and child participants from an isolated Amazonian community to determine whether they possess intuitive geometric conceptions, notwithstanding their lack of formal schooling, experience with maps, and a language containing an abundance of geometric terms. These investigators demonstrated that both children and adults spontaneously made use of foundational geometric concepts, including points, lines, parallelism, and right angles when trying to identify intruders in simple pictures, and used distance and angular relationships in geometrical maps to locate hidden objects. Finally, although a comparison group of American adults performed at a higher level overall than the Amazonian adults, the two groups showed similar profiles of difficulty. Dehaene et al. concluded that the existence of core Euclidean geometrical knowledge in all humans is inconsistent with Piaget's hypothesis of a developmental progression from topology to projective to Euclidian geometry.

A final body of evidence emanating from research with adults has yielded findings that are also inconsistent with the developmental trajectory proposed by Piaget and Inhelder. According to Newcombe and Huttenlocher (2006), although the Piagetian approach has stimulated much research on spatial thought for many decades, investigators who have focused on spatial cognition in adults have questioned the accuracy of characterizing mature spatial thought as explicitly Euclidean. They go on to describe how a good deal of evidence supports the assertion of cognitive psychologists that the spatial representations of adults are "inevitably erroneous, biased, and fragmentary" (p. 738).

Taken together, the mathematical inaccuracies of Piaget's topological primacy thesis along with the mounting, negative empirical evidence to date leads the Task Group to conclude that this theory lacks the kind of compelling support needed to make it useful for continuing to inform the design and testing of instructional approaches in geometry.

The van Hiele model of the development of geometric reasoning. The van Hiele model (1986) has been the dominant theory of geometric reasoning in mathematics education for the past several decades. According to this model the learner moves sequentially through five levels of understanding:

> ***Level 0: Visualization/Recognition*****—**Students can name common geometric figures but usually recognize them only by their shapes as a whole, not by their parts or properties.
>
> ***Level 1: Description/Analysis*****—**Students can judge a shape to be a certain type of figure based on its properties and can analyze component parts of the figures but cannot explain the interrelationships between figures and properties; they still do not understand definitions.
>
> ***Level 2: Informal Deduction or Ordering*****—**Students can form definitions, establish interrelationships of properties within and among figures, and follow informal proofs but cannot construct one.
>
> ***Level 3: Formal Deduction*****—**Students understand the significance of deduction as a way of establishing geometric theory within an axiomatic system, and comprehend the interrelationships and roles of undefined terms, axioms, definitions, theorems, and formal proof.
>
> ***Level 4: Rigor*****—**Students can reason formally about different axiom systems.

The majority of high school geometry courses are taught at Level 3.

Battista (2007) has recently carried out a review and analysis of the strengths and weaknesses of this theory, and new developments pertaining to it, including 1) extending the level descriptors from two-dimensional to three-dimensional shapes, 2) reexamining the nature of levels, 3) elaborating the levels and proposing alternatives though related conceptions, 4) considering the idea that different types of reasoning develop simultaneously but at different rates, 5) judging whether the developmental periods should be viewed as stages or levels, and 6) evaluating extant methods of assessment. Battista concludes that the van Hiele theory provides a generally valid description of the development of students' geometric reasoning, especially pertaining to the learning of shapes (see Clements & Battista, 1992, for a detailed review of the supporting evidence).

Cognitive processes underlying the van Hiele levels. From the standpoint of the Task Group's analysis, empirical examinations of the cognitive processes underlying van Hiele levels of geometric reasoning is a much more challenging endeavor. Nonetheless, such an approach is crucial for making further progress in this area, as well as designing appropriate assessment tasks for use in both research and practice. The Task Group thus concurs with Battista's (2007) comment that "…it is one thing to devise broad categories of behavioral descriptors; it is another to determine the cognitive processes underlying these categories of behaviors. This has been and will continue to be, a major challenge facing researchers" (p. 854).

Although theories such as the hierarchical interactionism model developed by Clements and colleagues (Clements & Battista, 1992; Clements & Sarama, 2007a) represent a major step in this direction, the Task Group agrees with Battista's (2007) perspective on the state of the science:

> Although a number of theories and studies have been reviewed in an attempt to describe the cognitive processes by which students progress through the early van Hiele levels, this area of research is still in its infancy. This is due in great part because researchers are investigating cognitive processes that cannot be observed. To achieve progress in this domain, it is important for mathematics education researchers to heed the work of researchers in other fields such as cognitive science and neuroscience. Such research can provide valuable insights into these difficult-to-observe processes (pp. 858–859).

Numerous advances have been made in recent years regarding the development of spatial cognition, including topics such as spatial visualization, spatial relations, spatial orientation, spatial perception, spatial memory, spatial reasoning, and spatial and visual imagery. Excellent reviews of this rich research literature can be found in Liben (2002), Newcombe and Huttenlocher (2000; 2006), and Tversky (2004). Nevertheless, comparatively few cognitive or developmental psychologists have explicitly studied the development and learning of Euclidean geometric concepts and skills. Having said this, it should be noted that Koedinger, Anderson, and colleagues have been applying theory-based cognitive processing approaches to instructional interventions in geometry for many years (see Ritter, Anderson, Koedinger, & Corbett, 2007 for an overview of this and related work, as well as recent work by Kao & Anderson, 2006, 2007; and Kao, Roll, & Koedinger, 2007). Finally, research on the cognitive neuroscience of geometric reasoning is just beginning to get off the ground (Kao & Anderson, 2006).

e. Obstacles to Mastery

Earlier, the Task Group mentioned how difficulties in learning about ratio and proportion can present obstacles to understanding the meaning of similarity of triangles. Additionally, research on some intriguing geometric misconceptions are described below which are also relevant to understanding characteristics of shapes, albeit with respect to the concept of area. As such, future efforts to design instructional approaches for overcoming these kinds of errors may assist students in acquiring concepts crucial to understanding the definition of similarity.

Same-perimeter/same-area misconception. Dembo et al. (1997) examined a common misconception involving the relationship between area and perimeter—namely, that shapes with the same perimeter must have the same area. In fact, shapes with the same perimeter frequently have different areas, and the area increases as the shape becomes more regular. Thus, as these authors note, for rectangles having a constant perimeter, area increases as the figure approaches a square and decreases as it approaches a line. Dembo et al. tested the effects of schooling on this misconception by comparing the performance of two groups of Israeli students: one that attended ultra-orthodox schools and had received virtually no instruction in math and science, and the other that attended mainstream schools and received extensive instruction in these areas.

Surprisingly, the ultra-orthodox 12- to 14-year-old group correctly solved the geometric misconception problems more frequently than did their mainstream peers. The authors suggested two possible explanations of this unexpected finding. According to one perspective, the relatively strong performance of the ultra-orthodox students may have resulted from a curriculum which cultivated proficiency in applying general cognitive strategies and in carefully implementing rules of analytical reasoning to solve problems. Alternatively, early conventional instruction in geometry and related topics may actually have had an adverse effect on mainstream students' geometric reasoning. That is, initial formal instruction may have inadvertently promoted this misconception as a consequence of students being presented with the concepts of perimeter and area pertaining to the same shapes and kinds of problems—and during one and the same course. As these investigators point out, since the same factors are used to compute perimeter and area for many types of shapes (e.g., length of the sides for squares, rectangles, and right triangles), students may deduce that area and perimeter are determined by the same variables, leading them to erroneously infer that when the perimeter remains the same under some transformation, the area must as well.

Illusion of linearity. De Bock and colleagues have recently reviewed numerous studies demonstrating what has come to be known as the "illusion of linearity." Essentially, this phenomenon consists of a misconception that the linear (or proportional) model can pertain to situations where it is in fact not applicable. More specifically, many students incorrectly believe that that if the perimeter of a geometric figure is enlarged *k* times, its area (and/or volume) is enlarged *k* times as well (De Bock, Verschaffel, & Janssens, 1998, 2002; De Bock, Van Dooren, Janssens, & Verschaffel, 2002; Freudenthal, 1983; Modestou, Gagatsis, & Pitta-Pantazi, 2004). Apparently, this misconception emerges not only in geometry, but also in elementary arithmetic (Van Dooren, De Bock, Hessels, Janssens, & Verschaffel, 2005), probability (Van Dooren, De Bock, Depaepe, Janssens, & Verschaffel, 2003), algebra and calculus (Esteley, Villarreal, & Alagia, 2004).

Remarkably, a series of studies has shown that even with considerable scaffolding (e.g., supplying drawings, presenting the problem in another format, or providing meta-cognitive hints), the vast majority of 12- to 16-year-old students fail to solve these problems due to a strong tendency to inappropriately apply linearity (De Bock et al., 1998; De Bock, Van Dooren, et al., 2002; Modestou et al., 2004). Van Dooren, De Bock, Janssens, & Verschaffel (2005) note that additional research suggests that this tendency is attributable to a set of closely related factors, including "the intuitiveness of the linear model, shortcomings in students' geometrical knowledge, inadaptive attitudes and beliefs towards mathematical (word) problem solving, and a poor use of heuristics" (p. 266).

De Bock, Verschaffel, Janssens, Van Dooren, and Claes (2003) explored the potential utility of two additional factors with 13- to 16-year-olds: 1) the authenticity of the testing context (i.e., prefacing the test with well-chosen, meaningful video fragments and linking all test items directly to these), and 2) the integrative use of drawings (i.e., having students draw a reduced copy of the figure described in the problem before trying to solve it). Neither of these manipulations improved performance. Additionally, both factors actually produced a negative effect. After exploring several possible explanations for this unanticipated finding, the authors conclude that, "Most likely, only a long term classroom intervention, not only acting upon students' deep conceptual understanding of proportional reasoning in a modeling context, but also taking into account the social, cultural and emotional context for learning, can produce a positive effect in defeating the illusion of linearity" (p. 460).

f. Conclusions and Recommendations

Classroom

Teachers should recognize that from early childhood through the elementary school years, the spatial visualization skills needed for learning geometry have already begun to develop. In contrast to Piagetian theory, young children appear to possess at least an implicit understanding of basic facets of some Euclidean concepts, although proper instruction is needed to ensure that children adequately build upon and make explicit this core knowledge for subsequent learning of formal geometry. Additionally, whereas children can reason to some extent about the properties of and relationships among different shapes, their developing abilities to acquire more detailed information about the metrics of these properties and the changes that occur under various transformations in the plane is by no means simple and straightforward.

Training

Teachers. Teachers need to learn more about the latest research concerning the development of children's spatial abilities in general and their geometric conceptions and misconceptions in particular. Acquiring knowledge of the spatial skills children bring to school with them, the limitations of these early developing competencies, and their use and misuse of shape words and names can help teachers capitalize on children's strengths and aid them in overcoming their weaknesses.

Future researchers. The next cohort of researchers who will be investigating geometry learning need to have a firm grounding in cognitive development and spatial information processing, in addition to mathematics education. Although some math education researchers have explicitly linked their work to advances in these areas, future progress in studying students' geometry learning will require a blend of content knowledge, proficiency with multiple research methods, and theoretical sophistication across several different disciplines. In addition, research teams composed of people who each possess a relevant area of expertise may be even more likely to help advance the study of geometric reasoning and evidence-based approaches to instruction in this domain.

Curriculum

Early exposure to common shapes, their names, and so forth appears to be beneficial for developing young children's basic geometric knowledge and skills. However, comparatively little is known about what the long-term effects would be of including a foundational treatment of more complex geometric concepts in preschool through second-grade curricula. Moreover, despite the widespread use of mathematical manipulatives such as geoboards, dynamic software, and so forth during the elementary school years, rigorous evidence is lacking as to precisely when and how these should be implemented to help children acquire a foundational understanding of concepts such as congruence, similarity, transformations, and angles. Finally, while a judicious reliance on manipulatives may enhance the initial acquisition of some concepts under specified conditions, students must eventually transition from concrete (hands-on) or visual representations to internalized abstract representations. The crucial steps in making such transitions are not clearly understood at present.

Research

***Basic*.** Longitudinal studies are needed to assess more directly how developmental changes in spatial cognition can inform the design of instructional units in geometry. Studies are needed to demonstrate whether and to what extent knowledge about similar triangles enhances the understanding that the slope of a straight line is the same regardless of the two points chosen, thus leading to a more thorough understanding of linearity.

***Classroom*.** More research is needed that specifically links cognitive, theory-driven research to classroom contexts. At the same time, cognitive theorizing pertaining to geometry learning needs to take into account more facets of classroom settings if it is to eventually have a large impact on the design of instructional approaches.

5. Algebra

This Task Group acknowledges the existence of arguments for early algebra learning, that is, implicit knowledge in elementary-school children's solving of arithmetic and other problems (Carraher & Schliemann, 2007). At this point, it is not known if the early algebra achievement of elementary school children reflects an actual implicit understanding of aspects of algebra, or if their performance is the result of the mathematical relation between algebra and arithmetic and not an indication of accumulating implicit knowledge. In either case, the Task Group focuses on explicit algebra content typically encountered in middle school to high school algebra courses. The bulk of the cognitive literature related to learning of this content focuses on simple linear equations and word problems; the Task Group summarizes the major findings from these studies below. The research literature for many of the remaining conceptual and procedural competencies identified within the Major Topics of School Algebra listed in the *Report of the Task Group on Conceptual Knowledge and Skills* is not sufficient for this Task Group to draw conclusions about the cognitive processes that contribute to these aspects of algebra learning; in many cases, sound studies are simply nonexistent.

a. Algebraic Equations

The majority of cognitive and learning research on algebra examines the processes underlying the solution of linear equations and the sources of problem-solving error. Some of the studies also include more complex equations (e.g., quadratic), but not enough research is available to discuss findings on these types of equations separately. The Task Group does, however, note a few common sources of problem-solving errors when students attempt to solve more complex algebraic equations.

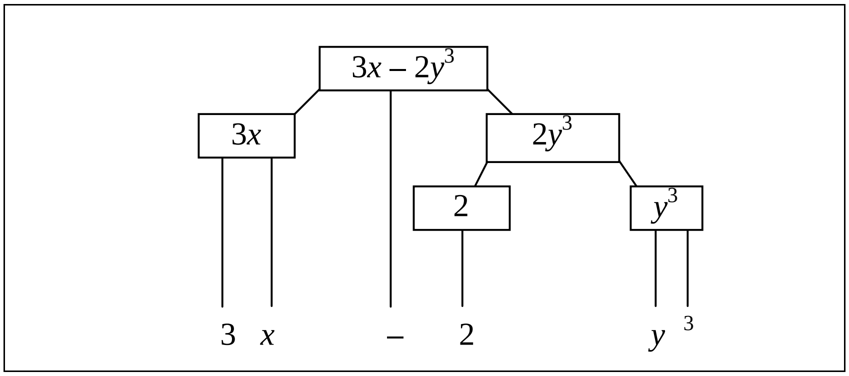
Paths of Acquisition

Conceptual and declarative knowledge. Studies of skilled adults and high school students who have taken several mathematics courses reveal that the processing of algebraic expressions is guided by an underlying syntax or system of implicit rules that guides the parsing and processing of the expressions (Jansen et al., 2003, 2007; Kirshner, 1989; Ranney, 1987). The learning of this syntax is not completely analogous to learning the syntax and grammar of natural language, because learning the syntax of algebra is strongly influenced by schooling. Learning of algebraic syntax is determined, in part, by earlier learned arithmetic rules, such as the order of operations; use of the commutative, associative and distributive properties; and by knowing the mathematical meaning of symbols, such as parentheses or summation signs, that note subexpressions within the equation.

Following Jansen et al. (2007), the parsing tree in Figure 2 illustrates the basic process followed by mathematicians when they solve algebraic equations, as revealed by an experimental method (Restricted Focus Viewer) that restricts the amount of information that can be viewed at one time and tracks the pattern with which components of the equation are processed. They typically scan the equation from left to right, but the variables, numbers, and exponents are not processed as individual symbols but rather as meaningful chunks, each of which is decomposed in turn. The processing is also influenced by core symbols that define chunks, including brackets, parentheses, horizontal bars in division, summation notations (∑), and so forth. For instance, mathematicians initially scan the following rational subexpressions from top to bottom, not strictly from left to right. Without an understanding of the mathematical meaning of these subexpressions, a person who is unfamiliar with algebra may view the subtraction sign and division lines as a continuous horizontal line that would be scanned from left to right and then top to bottom.

$$\frac{5x - 2}{2y + 7} - \frac{3y^2 - 1}{4}$$

Figure 2: Processing Linear Equations



Note: Experts scan algebraic equations in terms of meaningful chunks of information. For this expression, "3*x*" and "$2y^3$" are processed as chunks, and "$2y^3$" is then decomposed into the coefficient "2" or the variable, "y^3."

These methods indicate that the solving of algebraic equations rests, in part, on learning the basic rules of arithmetic, mathematical meaning of core symbols, and eventually the automatic parsing of equations on the basis of this knowledge. Evidence for automaticity comes from the finding that skilled problem solvers scan and process basic sub-expressions in these equations in a fraction of a second (e.g., Jansen et al., 2007). Comparisons of novices and skilled problem solvers reveal that this fast and efficient processing is possible because the skilled problem solvers have formed long-term memory representations of the basic structure of algebraic equations and the sequences of procedural steps that can be used to solve them (Sweller & Cooper, 1985).

A small-scale ($n = 33$) experimental study of college students' algebraic rule learning (e.g., when multiplying variables with exponents, add the exponents, $y^3\ x\ y^7 = y^{10}$) revealed substantial benefits to cumulative practice. This group practiced already learned rules with newly introduced rules and was contrasted with groups that only practiced rules individually or received follow-up reviews and practice of individual rules (Mayfield & Chase, 2002). In comparison to the two other conditions, cumulative practice resulted in better short-term and long-term retention of individual rules and a better ability to apply rules to solve problems that involved the integration of multiple rules. One potential reason for the advantage of cumulative practice is that it provides a context for comparing, contrasting, and eventually discriminating between rules that might otherwise be used inappropriately (e.g., confusing the rule for $(y^3)^7$ with the rule for $y^3 \times y^7$).

One method used to experimentally demonstrate the existence of such long-term memory representations is to compare the memory spans of experienced students and novices for meaningful and meaningless equations. In these studies, increasing skill is associated with longer memory spans for mathematically meaningful expressions but not for meaningless expressions with the same number of characters. When given 90 seconds to remember expressions such as $6y + 5\ (2x - 7)$, 11th-graders with several years of high school mathematics could remember strings of 10 to 12 symbols, whereas they remembered $4\frac{1}{2}$ to $5\frac{1}{2}$ symbols when the same strings were presented in a meaningless way; e.g., 5 (2*x* + – 7 6*y*. Memory span for meaningful but not meaningless expressions increased with number of mathematics classes and with practice (Sweller & Cooper, 1985).

Students who are first learning algebra and adults who are not skilled in mathematics do not have long-term memory representations of basic forms of linear equations or the sequences of procedural steps that can be used to solve these equations. The absence of or failure to access these long-term memory representations does not necessarily preclude the solving of linear equations, as long as the individual understands the general arithmetical and algebraic concepts and rules needed to solve the problem. Unfortunately, there are often substantive gaps in this knowledge. The result is that many students make mistakes. Problem solving is sometimes complicated by the execution of mathematically correct, but unnecessary, procedures. As an example, when presented with $a(y + 3z) - x = 4a$ and asked to solve for x, many students will unnecessarily expand $a(y + 3z)$ [i.e., $ay + 3az$]; the problem can still be solved but now requires several added steps.

Birenbaum et al. (1993) used a promising strategy for identifying sources of common errors such as these. A diagnostic test in which individual problems varied systematically in terms of the knowledge needed for correct solution was administered to eighth- and ninth-grade students in Israel. The problems ranged from relatively simple (e.g., $3 + x = 6 + 3 \times 2$) to those with more complex subexpressions [e.g., $6 + 4(x - 2) = 18$]. The pattern of correct and incorrect solutions across problems allowed for the identification of the most likely sources of error. The most common errors occurred because many students failed to correctly divide when terms included a coefficient and a variable (e.g., $9x$), and had difficulty applying the commutative and distributive properties [e.g., $4(x - 3)$]. Other common errors resulted from a failure to correctly order the operations, and to correctly add and subtract numbers on both sides of the equation, especially signed numbers (e.g., $5x - 4$). Using the same methods, Birenbaum and Tatsuoka (1993) found that many Israeli 10th-graders did not recall the laws of exponents. The two most common errors resulted from failure to recall that $X^0 = 1$ and that $(X^m)^n = X^{mn}$. Incorrect factoring was also a common source of error.

Similar types of errors have been found in the United States and other countries. In these studies, moving terms from the left to the right side of an equation was a common point at which errors occurred (Anderson, Reder, & Lebiere, 1996; Cooper & Sweller, 1987; Lewis, 1981). In keeping with the division errors found by Birenbaum et al. (1993), for the problem, $\frac{[2(x+6)]}{y} = z$ (solve for x), one type of error involves moving y from the left to the right, rather than multiplying both sides of the equation by y. With this error, the right side of the equation reads $\frac{z}{y}$, rather than zy. These types of errors often reflect a poor conceptual understanding of the syntax of algebraic expressions.

A poor understanding of the concept of mathematical equality and the meaning of the "=" is common for elementary school children in the United States, and continues for many children into the learning of algebra. Many elementary school children believe that the equal sign is simply a signal to execute an arithmetic operation. On typical problems such as 3 + 4 + 5 = __, this misinterpretation does not cause any difficulty. However, on less typical problems (at least in U.S. mathematics textbooks), such as 3 + 4 + 5 = __ + 5, it causes most third- and fourth-graders either to just add the numbers to the left of the equal sign, and answer "12," or to add all numbers on both sides of it, and answer "17" (Alibali & Goldin-Meadow, 1993).

Knuth et al. (2006) extended this research to 177 U.S. middle school children. They assessed children's understanding of the equal sign as expressed in arithmetic (e.g., the meaning of "=" in 4 + 8 = __) and how this knowledge of equality was related to their ability to solve simple linear equations; e.g., $5x - 5 = 30$. As found with third- and fourth-graders, many of the eighth-graders in this study interpreted the "=" as indicating the outcome of an arithmetic operation. Only 31% of the eighth-graders understood it as representing the equality of the terms on the left and right side of the equation. When solving the linear expressions, 33% of the eighth-graders used an algebraic strategy and these students always (100%) got the correct answer. The other two thirds of students used a "guess and check" strategy—as is often found for U.S. students (Cai, 2004; Johanning, 2004)—or some type of arithmetic strategy and frequently erred in solving the equation.

About 75% of the eighth-grade students who understood mathematical equality used algebra to solve linear equations, compared to less than 20% who understood "=" as a signal to perform an operation. The relation between understanding the concept of mathematical equality and skill at solving linear equations held, when standardized mathematics achievement scores and algebra course work were statistically controlled.

One potential source of U.S. students' poor understanding of the equal sign is the way in which problems are presented in textbooks. McNeil et al. (2006) provided a systematic examination of four commonly used textbooks series in middle school and found that the most frequent presentation of "=" was in the context of 'operate-equals-answer' format; e.g., $4 + 7 = 2x + 3 = 11$ (see also Seo & Ginsburg, 2003). Other studies have indicated that use of this format contributes to students' interpretation of "=" as operational rather than relational (Baroody & Ginsburg, 1983; McNeil & Alibali, 2005). A relational interpretation of "=" is most common for problems for which operations are needed on both sides of the equation (e.g., $4 + 5 = 11 - 2$; $3x + 5 = x + 15$). Yet, less than 5% of problems in middle school textbooks in the United States use this format, reaching a maximum of 9% of problems in eighth-grade textbooks.

Although it has not been empirically assessed, it is possible that the tendency of simple arithmetic problems to be presented vertically in U.S. textbooks may make the transition to left to right horizontal processing of algebraic expressions more difficult than it needs to be. This is a readily testable hypothesis and, if correct, can be easily remedied with the presentation of simple arithmetic problems in a horizontal format beginning with first-grade textbooks.

Procedural bugs. Errors in the solving of algebraic equations are sometimes classified as procedural bugs in a way analogous to the "buggy rules" noted earlier for subtraction (Birenbaum, Kelly, Tatsuoka, & Gutvirtz, 1994; Schoenfeld, 1985; Sleeman, 1984; Sleeman, Kelly, Martinak, Ward, & Moore, 1989; Wenger, 1987). These errors can occur due to overgeneralized use of procedures that are correct for some problems or from a misunderstanding of the procedure itself. Schoenfeld described a number of these types of procedural errors, a few of which are illustrated in Figure 3.

Figure 3: Algebraic Bugs

Expression	Buggy/Incorrect Translation	Potential Source of Confusion
$(X + Y)^2$	$X^2 + Y^2$	$2(X + Y) = 2X + 2Y$
$\sqrt{(X + Y)}$	$\sqrt{X} + \sqrt{Y}$	$\sqrt{(XY)} = \sqrt{X}\sqrt{Y}$
$\frac{X}{Y+Z}$	$\frac{X}{Y} + \frac{X}{Z}$	$\frac{Y+Z}{X} = \frac{Y}{X} + \frac{Z}{X}$

Unfortunately the nature of these bugs often differs from one student to the next and often for the same student from one equation to the next (Birenbaum et al., 1994; Sleeman et al., 1989). The problem of stability arises because the same equation can elicit several different types of bugs, and many students make errors on the same problem from one time to the next for different reasons. Although many bugs do not occur with enough consistency to inform specific classroom practices, a few bugs may be consistent across and within students. Preliminary studies by Sleeman et al. suggest that remediation that focuses on these specific bugs can reduce their frequency. Follow-up studies—perhaps using the classification methods described by Birenbaum and colleagues (1993, 1994; Birenbaum & Tatsuoka, 1993), focusing on identifying the sources of error underlying stable bugs or classes of bugs and assessing the effectiveness of instructional strategies in correcting them—are needed.

b. Word Problems

The Task Group's review of algebraic word problems includes studies of college students, due to a shortage of studies of middle and high school students' performance on such problems. Results for these college samples are likely to underestimate the difficulty of solving word problems for high school students. In a few places, the Task Group includes research on multistep arithmetical word problems, because the core processes and sources of error appear to be similar for arithmetical and algebraic word problems. The Task Group also notes that the focus of some of this research is on problem-solving processes and not the learning of specific algebraic content. A review of the these studies is, nonetheless, needed because of the wide use of word problems in the mathematics curriculum, because the application of algebraic skills (e.g., in physics classrooms) is often in the context of word problems, and because student difficulty with solving word problems was identified as an area of concern in the National Survey of Algebra Teachers (Hoffer et al., 2007, Table 3).

Paths of Acquisition

Problem translation and solution. Mayer (1982) proposed that the solution of algebraic word problems requires two general sets of processes: problem translation and problem solution. Problem translation involves transforming the verbal statement of the problem into a set of algebraic equations. It determines how the student forms a mental representation of the problem. The generation of the representation starts with an understanding of the text within which the problem is embedded (Kintsch & Greeno, 1985). Text comprehension involves understanding not only the meaning and mathematical implication of specific words (e.g., "speed" implies a rate problem), but also the structure of the entire problem.

In an analysis of word problems presented in algebra textbooks, Mayer (1981) found that most problems included four types of statements: assignment statements, relational statements, questions, and relevant facts. Assignment statements, not surprisingly, involve assigning a particular numerical value to some variable. Relational statements specify a single relationship between two variables. Questions involve the requested solution (e.g., "What is *X*?"). Relevant facts involve any other type of information that might be useful for solving the problem. Problem translation involves taking each of these forms of information and using them to develop corresponding algebraic equations. The translation of assignment statements, questions, and relevant facts does not pose much of a problem for most high school and college students (Lewis & Mayer, 1987; Mayer, 1982; Wenger, 1987). However, discriminating relevant from irrelevant information (Low, Over, Doolan, & Michell, 1994) and determining if the problem is solvable (Rehder, 1999) are potential sources of difficulty for many students. Translation errors most frequently occur during the processing of relational statements.

An example is provided by a simple problem: "There are six times as many students as professors at this university" (Clement, 1982, p. 17). Clement presented this problem to freshman engineering students at a major state university and asked them to write an equation that represented the relation between the number of students (s) and the number of professors (p). Thirty-seven percent of the engineering students committed an error on this problem, typically $6s = p$. This type of error is fairly common (Hinsley et al., 1977) for at least two reasons. The first is that the syntax, or structure, of the relational statement suggests a direct (though incorrect) translation into an algebraic expression. So "six times ... students" is literally translated into $6s$. Second, many students appear to interpret relational statements as requesting static comparisons. In this example, $6s$ is used to represent the group of students and p to represent the group of professors. In other words, for many students the "=" does not represent the actual equality of $6p$ and s, but rather simply separates the two groups. Students who correctly translate this relational statement understand that s and p represent variables, not static groups. These students understand that to make the number of professors equal to the number of students, some type of operation has to be performed; the smaller quantity, p, has to be increased so as to make it equal to the larger quantity, s. This translation leads to the correct algebraic expression, $6p = s$. The finding that types of errors are common in college students who intend to major in engineering implies that mistranslations of relational statements are likely to be widespread.

In a large-scale study that included 8th- to 10th-grade students, MacGregor and Stacey (1993) demonstrated that it is not simply the syntax of the relational wording that makes the translation process a common point of error. Errors occurred even for problems in which the statement could be directly translated into an equation; e.g., "z is equal to the sum of 3 and y." For one problem—"I have \$$x$ and you have \$$y$. I have \$6 more than you. Which of following must be true?"—students were asked to choose from among five alternatives. Across grades, only 34% to 38% of the 8th- to 10th-graders chose the correct, $x = y + 6$, equation. There was no predominantly incorrect response across the four potential mistranslations, although mistranslations that involved addition or subtraction (13% to 22% of choices; e.g., $6 + x = y$) were more common than mistranslations that involved multiplication (3% to 15%, e.g., $6x = y$) (p. 222). Capraro and Joffrion (2006) also found a variety of translation errors for a sample of 668 middle school students, as did Sebrechts, Enright, Bennett, and Martin (1996) when undergraduates solved algebraic word problems from the quantitative section of the Graduate Record Exam (GRE). The pattern indicates there are many ways to mistranslate the same word problem, just as there are many potential procedural bugs when solving algebraic equations.

At the same time, relational information conveyed in a word problem can sometimes aid problem solving if this relational information is consistent with students' previous out-of-classroom experiences and if these experiences and the corresponding situational representation can be used to create non-algebraic solution strategies (Bassok, Chase, & Martin, 1998; Koedinger & Nathan, 2004; Martin & Bassok, 2005). Bassok et al. found that the majority of word problems presented in one U.S. textbook series across first to eighth grade used story situations that were consistent with everyday activities or with everyday uses of objects described in the problems. Koedinger and Nathan discovered that these types of word problems are sometimes easier to solve (i.e., lower error rate) than corresponding linear equations. An analysis of solution strategies and error rates revealed that this was due to the frequent use of non-algebraic, arithmetic-based strategies for solving the word problems; these might involve a guess and test approach whereby the presented quantities are added, multiplied, etc. until an answer is obtained. Although these high school students were more successful with use of these non-algebraic strategies, the question of whether this contributes to their learning of formal algebraic representations of problem situations remains to be determined (Koedinger & Nathan).

Hembree's (1992) large-scale meta-analysis of students' ability to solve mathematical word problems from first grade to college level also reveals contextual effects. Abstract problems were more difficult to solve than concrete problems (mean $r = -.14$; mean across studies) but the largest effect was for familiarity ($r = .40$). Familiarity was defined in such a way that it included familiarity with classes of word-problem (e.g., interest, compare, distance problems) or familiarity with the cover context (e.g., baseball, travel). Evidence for the importance of familiarity of problem class comes from the finding that contexts that were based on the students' personal interests ($r = .04$) or preferences ($r = -.04$) were not related to problem-solving skill; that is, it was familiarity with solving the class of problem (e.g., rate) and not students' personal interests.

The second general set of processes suggested by Mayer (1982) as necessary for the solution of algebraic word problems or problem solution refers to the actual use of algebraic or arithmetical procedures to solve the resulting equations. The same potential sources of error described for solving of linear equations can occur during this stage of solving word problems.

Schema development. Hinsley et al. (1977) showed that successful translation of algebraic word problems, as well as the solution of algebraic equations and many other problem types, is guided by *schemas*—these include the syntax of equations. Sweller and Cooper (1985) provided a useful definition: "Schemas are defined as mental constructs that allow patterns or configurations to be recognized as belonging to a previously learned category and which specify what moves are appropriate for that category" (p. 60).

In short, a schema is a long-term memory representation that makes possible fast and automatic recognition of key elements of an equation or word problem, enables the classification of the problem into a conceptual group (e.g., velocity problems, interest problems), and has a linked system of procedures that can be used to solve the problem (e.g., Larkin, McDermott, Simon, & Simon, 1980). It should be noted that this cognitive science approach to schema development differs from Piaget's less precisely defined concept of schema.

Morales et al.'s (1985) study of third-, and a combined group of fifth- and sixth-grade children's conceptual understanding and ability to solve arithmetic word problems illustrates the usefulness of the concept of schema. The children's conceptual knowledge was inferred based on how they sorted word problems into categories. The question was whether the sorts were based on conceptual similarity (e.g., combine versus change problems), (Carpenter & Moser, 1984; Riley, Greeno, & Heller, 1983) or on unimportant (surface) similarities in how the problems were worded (e.g., both about baseball). For both grade levels, more than two-thirds of the errors were conceptually based—e.g., using a procedure appropriate for some class of problem but not the current problem—rather than due to computational error. More important, the categories formed by third-graders were more strongly influenced by the surface structure of the problems than by any underlying conceptual similarities, whereas the categorizations of the fifth- and sixth-graders were more strongly influenced by conceptual similarities. Third-graders who tended to organize the problems on the basis of conceptual categories, rather than surface structure, were much more accurate at solving the problems than were their peers who focused on surface features; the fifth- and sixth-graders did not make enough errors to conduct this type of analysis. The emerging ability to categorize word problems based on is underlying concepts (e.g., whether the problem asks for quantities to be combined or compared) and the corresponding reduction in problem-solving errors is consistent with development of category-specific schemas.

Sweller and colleagues have demonstrated that one way in which schema development can occur with both algebraic equations and word problems is through the use of worked examples (e.g., Cooper & Sweller, 1987; Sweller & Cooper, 1985). Worked examples provide students with a sequence of steps that can be used to solve these problems. The students then solve a series of related problems that are in the same category (e.g., interest problems) and involve the same or a very similar series of problem-solving steps. Studies by Reed and colleagues (Reed & Bolstad, 1991; Reed, Willis, & Guarino, 1994) reveal that, at least for word problems, worked examples that include an explanation of procedures (e.g., rate as related to work per unit of time or distance per unit of time) and several examples are more effective than simply providing students with the procedural steps.

The associated experimental and quasi-experimental studies have shown that the use of well-developed worked examples has several advantages over more conventional teacher instruction followed by worksheet practice (Carroll, 1994; Cooper & Sweller, 1987; Sweller & Cooper, 1985; Zhu & Simon, 1987). It is likely to have similar advantages over unguided discovery learning. Because of the many potential bugs that students can commit when solving algebraic equations and the common translation errors for word problems, teacher presentation of a problem or two followed by worksheet practice or homework can result in students' repeatedly committing (and therefore practicing) these bugs or translation errors.

In comparison to conventional practice, students provided with worked examples solve problem in the same class (e.g., distance) faster and with fewer errors (Carroll, 1994). The benefits of worked examples for solving different classes of problems (i.e., far transfer) are mixed. Some studies have revealed no skill transfer from one class of worked examples to another (Sweller & Cooper, 1985), but other studies have found positive transfer (Cooper & Sweller, 1987). Cooper and Sweller provided preliminary evidence consistent with the hypothesis that worked examples promote the learning of schemas and the memorization of embedded procedures. A tentative conclusion is that use of worked examples promotes the automatization and transfer of procedures used across classes of problems but does not appear to promote the transfer of the schema, that is, the specific sequence with which embedded procedures are used to solve the different classes of problem.

These same studies also have implications for unguided discovery. By definition, students using discovery approaches to learn algebra are novices and similar in some ways to students given conventional worksheets for problem-solving practice; students in both situations typically devise their own problem-solving strategies. Unlike students provided with worked examples, students engaging in conventional practice attempt to solve problems using the general problem-solving means-ends heuristic (i.e., working backward from the goal) (Newell & Simon, 1972), and often fall back on arithmetical rather than algebraic representations of the problem (Koedinger & Nathan, 2004). Even students provided with worked examples for one class of algebra problem revert to the means-ends heuristic when asked to solve problems for which a problem-solving schema has not yet been learned.

Although use of the means-ends heuristic can be an effective approach in many problem-solving situations, it requires considerable working memory resources and results in an attentional focus on the problem-solving goal and not on learning the sequence of problem-solving steps (Sweller, 1989). An attentional focus on the problem-solving goal appears to interfere with learning the sequence of these steps; that is, learning the underlying schema and connecting the goal to the sequence of steps needed to attain it (Cooper & Sweller, 1987; Sweller et al., 1983).

These and other results suggest that the effectiveness of worked examples appears to be due, at least in part, to a reduction in working memory demands that accompany use of means-ends problem solving. The elimination of these working memory demands allows attention to be focused on learning the sequence of steps that can be used to solve the class of problem (e.g., velocity) illustrated in the worked examples (Sweller, 1989). Worked examples can also provide a means of practicing embedded procedures.

Although it has not been as extensively studied in the context of mathematics learning, research in other areas reveal limits on the effectiveness of worked problems. Worked problems are most effective during the initial stages of learning and lose their advantage over other methods, such as exploration, as the learners' level of competence in the domain increases (Kalyuga et al., 2001; Tuovinen & Sweller, 1999).

The above noted limitations in terms of transfer for worked examples were demonstrated by Blessing and Ross (1996). These researchers found that undergraduates who had attended a high school for mathematically and scientifically gifted students, Illinois Mathematics and Science Academy, were skilled at translating algebraic word problems into appropriate equations. Their skill was, in part, related to fast and automatic access to problem solving schemas (e.g., distance = rate × time) associated with common word problems (e.g., those involving motion, interest, etc.) and the situations presented in these problems (e.g., an investor receiving dividends). However, when the "cover story," or the way the problem was presented, was modified, but the underlying algebra needed to solve the problem was left unchanged, these students committed more errors. In a series of studies that included undergraduates, and 9th- and 10th-grade mathematics honors students, Bassok (1990) found that transfer from one problem type (e.g., banking) to another (e.g., manufacturing) occurred when students recognized that the problems were asking the same basic question, such as questions about the rate of change. This spontaneous transfer did not occur for all students and largely disappeared if the problem contexts were too different, even if the underlying similarity of the problems was not changed.

Reed and colleagues (Reed, 1987; Reed, Dempster, & Ettinger, 1985) found the same pattern for undergraduates. They also found that if students' recognized conceptual similarities between problems they were much more likely to draw the analogy and use the same problem-solving procedures; for example, mixture problems that involve determining percentage of acid in a solution are conceptually the same as alloy problems (e.g., percentage of tin in bronze)

In other words, retrieval of the problem-solving schema is tightly tied to the ways in which the corresponding class of problem (e.g., distance, interest) is typically presented in word problems. Modification of this cover story can result in failure to retrieve the appropriate procedural sequence or retrieval of the wrong sequence. Hembree's (1992) meta-analysis sheds some light on student attributes that might promote transfer across classes of word problem. Hembree found that students who were skilled problem solvers also were skilled at analogical reasoning ($r = .56$) and at drawing inferences ($r = .49$). Other student-level traits, such as creativity ($r = .22$) and critical thinking ($r = .37$) were less closely related to successful problem solving. The importance of analogical and inferential reasoning is consistent with transfer effects in other areas (e.g., Holyoak & Thagard, 1997), and Reed's (1987) studies of algebraic word problems.

Obstacles to mastery

Stumbling blocks to the mastery of algebraic equations and the ability to translate word problems into algebraic expressions and equations are multifold and reflected in the many sources of problem-solving errors described in the preceding sections. *In addition, a common obstacle to ability to solve algebraic equations is inadequate preparation in arithmetic.*

In keeping with this conclusion, Hembree's (1992) meta-analysis revealed that for ninth-graders, the best predictors of the ability to solve word problems were computational skills (r = .51) and knowledge of mathematical concepts (r = .56). Other predictors were intelligence (r = .44), reading ability (r = .44), and vocabulary (r = .26).

The deficiency in basic arithmetic skills includes poor knowledge of the properties of arithmetic (e.g., order of operations; commutative, associative, and distributive property; the laws of exponents) and committing arithmetic errors, especially the manipulation of signed numbers (e.g., -5) and rational expressions. Students who struggle with algebraic equations also make factoring errors and use algebraic procedures incorrectly (i.e., commit bugs). In comparison to skilled problem solvers, poor problem solvers do not process algebraic equations by breaking them into mathematically meaningful subexpressions, and often do not even understand the significance of symbols that signal the existence of a subexpression, such as parentheses. Many do not even have a good understanding of mathematical equality or the "=" sign. Translation of word problems, especially relational information, into appropriate algebraic expressions and the discrimination of relevant and irrelevant information are consistent sources of student difficulty.

At a cognitive level, problem-solving errors and learning the syntax of algebraic expressions and algebraic schemas are influenced by working memory (Ayres, 2001, 2006; Cooney & Swanson, 1990; Lee, Ng, Ng, & Lim, 2004; Pawley, Ayres, Cooper, & Sweller, 2005). Working memory limitations also make the processing of relational sentences in word problems and the discrimination of relevant from relevant information especially difficult (Cooney & Swanson). As noted for whole number arithmetic, the commitment of procedures, rules, and often-used facts to long-term memory will reduce the working memory demands associated with solving the problem, thus freeing resources for processing less familiar problem features. As described above, well-designed worked examples may be effective in allowing students to focus working memory resources on learning classes of algebraic problem and sequences of problem-solving steps. At the same time, worked examples that include redundant or extraneous information may increase the working memory demands of processing these examples and thereby make them less effective (Pawley et al.). Pawley et al. found that redundant information for one student may, however, be helpful for another; thus, the effects of including redundant or irrelevant information appears to vary with the working memory capacity and mathematical competency of the student.

In a small-scale experimental study, Kalyuga and Sweller (2004) demonstrated that the potential cost of including irrelevant information can be addressed with use of faded worked examples. Here, the amount of information provided is reduced as students' skill level increases. Use of these learner-adapted worked examples resulted in moderate gains (d = .46) in the ability to solve linear equations. Use of other learner-adapted systems, especially cognitive tutors, is also associated with improved skills in geometry and algebra, if the tutor is well integrated with classroom instruction and the overall curriculum (Koedinger, Anderson, Hadley, & Mark, 1997; Ritter et al., 2007). In a large-scale quasi-experimental study, Koedinger et al. demonstrated substantial gains (d = 1.2) in the ability to translate algebraic word problems into equations. More modest gains were found for scores on two standardized algebra tests (d's = .30).

In addition to working memory, accuracy at solving various forms of mathematics word problems, such as those found on the SAT, is also related to spatial abilities across samples ranging from gifted middle school students to college students (Casey, Nuttall, Pezaris, & Benbow, 1995; Casey, Nuttall, & Pezaris, 1997; Geary, Saults, Liu, & Hoard, 2000; Johnson, 1984). The relation between spatial abilities and problem-solving accuracy may be due to the skilled use of visuospatial diagrams [see Larkin and Simon (1987) for general discussion of the utility of diagrams] or representations of the core relationships described in the problem, particularly the translation of relational information. Providing diagrams or instruction on the use of diagrams reduces errors rates when college students solve multi-step arithmetical word problem (Johnson; Lewis, 1989; Lewis & Mayer, 1987).

In Hembree's (1992) meta-analysis, the use of diagrams was more strongly related to the ability of fourth- and seventh-graders to solve word problems (r = .54) than was the use of other heuristics. From second grade to college, direct instruction on use of diagrams was much more effective for promoting their correct use than was practice alone (d = 1.16). Findings from a recent study of Japanese (n = 291) and New Zealand (n = 323) algebra students are also consistent with the usefulness of diagrams, at least for some types of problems (Uesaka, Manalo, & Ichikawa, 2007). These results are promising. However, the benefits and limitations of diagrams for facilitating the solving of different types of word problems remain to be determined and are more strongly related to instructional issues than to learning of specific algebraic content.

The ability to solve word problems is also related to reading ability and nonverbal reasoning ability, above and beyond the influence of working memory (Lee et al., 2004). It is also very likely that other factors reviewed earlier, including motivation, self-efficacy, anxiety, and so forth contribute to skill development in algebra (e.g., Casey et al., 1997). Although cause-effect relations cannot be determined, Hembree (1992) found that skill at solving word problems was related to positive attitudes towards mathematics (r = .23) and problem solving (r = .20), self-confidence in mathematics (r = .35), and self-esteem (r = .27). These correlations are consistently lower than those found between measures of mathematical preparation (e.g., computational skills) and cognitive factors (e.g., use of diagrams) and skill at solving word problems. Studies that simultaneously assess all of these constructs are needed to fully understand their relative contributions. As an illustration, one small-scale (n = 42) correlational study simultaneously assessed several of these constructs as related to achievement gains in Algebra II (Jones & Byrnes, 2006). The analyses allowed for estimates of the unique contribution of multiple constructs and classroom-related behaviors. Completion of homework ($ß$ = .36) was associated with higher end of class achievement test scores. Higher preclass knowledge of algebra ($ß$ = .32) and self-regulation ($ß$ = .33; e.g., ability to organize and self-check work) were also predictive of higher postclass scores. Frustration was associated with lower scores ($ß$ = -.26).

Research on learning in general (described in the General Principles: From Cognitive Processes to Learning Outcomes section in this report) indicates a benefit for practice that is distributed across time, as contrasted with the same amount of practice massed in a single session. Pashler, Rohrer, Cepeda, and Carpenter (2007) provide a recent review, including discussion of one area of mathematics. Initial experimental studies with mathematics, specifically teaching probability, are consistent with the more general literature. As with other content areas, distributed and massed practice reveal no difference after 1 week (70 to 75% correct), but after 4 weeks, the distributed practice group correctly solved 64% of the permutation problems as compared to 32% for the group trained with massed practice (Rohrer & Taylor, 2006).

A unique study of the longer-term benefits of distributed practice was provided by Bahrick and Hall (1991). In this study, algebra and geometry tests were carefully constructed based on textbooks and New York State Regents Examinations from 1945 to 1985. These exams were administered to about one thousand 19- to 84-year-olds. Information was obtained on high school and college course work, grades, and standardized test scores (for a sub-sample), as well as mathematics-related occupations (e.g., math teacher) and other activities that would involve rehearsal of algebra or geometry after completion of formal schooling. The retention interval for algebra began with the last algebra course taken in high school or college; for geometry, it began with the completion of plane geometry in high school. These data allowed for estimates of the degree of retention over a 50-year period as a function of these variables.

Overall, there was a steady decline in algebraic skills once the last course was taken. Over a 50-year interval between their last mathematics course to the time of the study assessment, about two-thirds of concepts and procedures typically taught in Algebra I was lost. Students who received an A in Algebra I, but took no other mathematics courses, retained more than students who received a B and these students in turn retained more than C students. The rate of decline in algebra skills was similar across groups; across a 50-year interval, the performance of all of these groups remained above that of a control group who had not taken high school algebra. The best predictor of long-term retention of competencies in algebra was the number of mathematics courses taken beyond Algebra I. Students taking college calculus showed a 20% decline in algebra performance over the 50 years, controlling for occupational and other potential confounds. Students taking a course beyond calculus showed no decline in algebra skills during this interval. A similar pattern emerged for content typically covered in Algebra II, but the effects of additional course work were less pronounced for geometry. The course work results suggest that the distributed review and integration—which was likely to have occurred more consistently for algebra than geometry—of the material across years contributes to the retention of the material throughout adulthood.

c. Conclusions and Recommendations

Too many students in high school algebra classes are woefully unprepared for learning even the basics of algebra. The types of errors these students make when attempting to solve algebraic equations reveal they do not have a firm understanding of many basic principles of arithmetic (e.g., commutativity, distributivity), and many do not even understand the concept of equality. Many students have difficulty grasping the syntax or structure of algebraic expressions and do not understand procedures for transforming equations (e.g., adding or subtracting the same value from both sides of the equation) or why transformations are done the way they are. These and other difficulties are compounded as equations become more complex and when students attempt to solve word problems.

With respect to policy, the situation is not likely to improve substantively without concerted and sustained federal efforts to make focused changes in teaching and curricula from elementary school forward, and efforts to change the ways in which teachers and future researchers are trained. There are many gaps in our current understanding of how students learn algebra and the preparation that is needed by the time they enter the algebra classroom. Funding to encourage scientists to enter research in this area is needed and to encourage the formation of research teams that will translate basic science findings into the design of instructional interventions to be assessed for effectiveness in the classroom.

Classroom

Teachers should not assume that all students understand even basic concepts, such as equality. Many students will not have a sufficient understanding of the commutative and distributive properties, exponents, and so forth to take full advantage of instruction in algebra.

Many students will likely need extensive practice at basic transformations of algebraic equations and explanation as to why the transformations are done the way they are; for instance, to maintain mathematical equality across the two sides of the equation. Common errors, as illustrated in Figure 3, may provide an opportunity to discuss and remediate overgeneralizations or misconceptions.

The combination of explanation of problem-solving steps combined with associated concepts is critically important for students to effectively solve word problems. For both equations and word problems, it is important that students correctly solve problems before given seatwork or homework. If students are making mistakes, then there may be a risk they will continue to make these errors and thus practice them during seatwork or homework.

Training

Teachers. Teachers should understand how students learn to solve equations and word problems and causes of common errors and conceptual misunderstandings. This training will better prepare them for dealing with the deficiencies students bring to the classroom and for anticipating and responding to procedural and conceptual errors during instruction.

Future researchers. To implement the recommendations that follow, the next generation of researchers to study algebra learning will need multidisciplinary training in mathematics, experimental cognitive psychology, and education. This can be achieved through interdisciplinary doctoral programs or at a federal level postdoctoral fellowships that involve work across these disciplines.

Curriculum

There are aspects of many, if not all, current textbook series in the United States that contribute to the poor preparation and background of algebra students. Modifying textbooks so that operations (arithmetical and algebraic) are presented on both sides of the equation, not just the typical operate-equals-answer format, is just one example of how textbooks can be improved.

The use of worked-out examples that include conceptual explanation, procedural steps, and multiple examples holds promise for teaching students to solve common classes of problems.

Retention of algebraic skills into adulthood requires repeated exposure that is distributed over time. This occurs as core procedures and concepts are encountered across grades. In much of mathematics, distributed practice should naturally occur as students progress to more complex topics. However, if basic skills are not well learned and understood, the natural progression to complex topics is impeded. This is because students will continue to make (and potentially practice) mistakes. As an example, procedures for transforming simple linear equations are embedded in more complex equations and thereby practiced as students solve them. The practice will not be effective, however, if students incorrectly transform basic equations, as they often do.

Research

Basic**.** The development of assessment measures that teachers can use to identify core deficiencies in arithmetic (whole number, fractions, and decimals), and likely sources of procedural and conceptual errors in algebra are needed. The early work of Birenbaum and colleagues appears promising in this regard.

Research that explicitly explores the relation between conceptual understanding and procedural skills in solving algebraic equations is needed. Research on how student's solve linear equations and where and why they make mistakes needs to be extended to more complex equations and other Major Topics of School Algebra identified by the Conceptual Skills and Knowledge Task Group.

The issue of transfer, that is, the ability to use skills learned to solve one type or class of problem to solve another type or class of problem, needs considerable attention. Of particular importance is determining the parameters that impede or facilitate transfer, as illustrated by the work of Reed and Sweller.

Research on instructional methods that will reduce the working memory demands associated with learning algebra is needed. Although there are individual differences in working memory capacity, aspects of instruction (e.g., using faded worked examples) may be modifiable in ways that reduce working memory demands. Instructional or curricular changes that reduce working memory demands appear to provide students with an enhanced potential to learn the procedure or concept that is the focus of instruction.

Longitudinal research is needed to identify the early (e.g., kindergarten, first-grade) predictors of later success in algebra.

Classroom. A mechanism for fostering translation of basic research findings into potential classroom practices and for scientifically assessing their effectiveness in the classroom is needed. Cognitive tutors for algebra illustrate how this can be achieved. Equally important, mechanisms for reducing the lag time between basic findings and assessment in classroom settings need to be developed.

E. Differences Among Individuals and Groups

1. Sex Differences

For large, nationally representative samples, the average mathematics scores of boys and girls are very similar; when differences are found, they are small and typically favor boys (Appendix C). However, there are consistently more boys than girls at the low and high ends of mathematical performance (Hedges et al., 1995). The overrepresentation of boys at the high end of mathematical performance has garnered considerable media attention and debate, but it has obscured the fact that average differences are small, if they are found at all, and has been a distraction from the goal of improving the mathematical competencies of both boys and girls.

An overview of sex differences in overall performance across a variety of national and international data sets is presented in Appendix C. Mean differences often favor boys but are small, with effect sizes ranging from -01 to .16. In adulthood, men have a small advantage on measures of quantitative literacy, but this gap has narrowed since 1992 (d = .21 in 1993, d = .11 in 2003). These results are consistent with similar analyses (Hedges & Nowell, 1995) and with meta-analyses that include smaller-scale studies (Hyde, Fennema, & Lamon, 1990). The magnitude of the gap may have diminished, but any such changes have not been consistent across grades or tests (Nowell & Hedges, 1998). More consistent sex differences have been found for some measures and for more select samples. As an example and as recently reviewed by Halpern et al. (2007), the male advantage (d ~ .40) on the SAT mathematics test has been remarkably stable during the past 40 years.

Differences are also consistently found at the low and high ends of performance, with more boys than girls at these extremes (Hedges & Nowell, 1995; Strand & Deary, 2006). In a large-scale prospective study (see section on Learning Disabilities later in this report), Barbaresi et al. (2005) found that about two boys for every girl met one or several diagnostic criteria for a learning disability in mathematics sometime before the end of high school. The ratio of boys to girls at the high end tends to increase as the cutoff becomes more selective. Across multiple national studies, Nowell and Hedges (1998) found the ratio of boys to girls in the top 10% of mathematics scores ranged from 1.3:1 to 2.5:1. In these same studies, the ratio of boys to girls in the top 1% ranged from 2.6:1 to 5.7:1. Differences at the extremes begin to emerge in elementary school (Mills, Ablard, & Stumpf, 1993) and possibly before kindergarten (Robinson, Abbott, Berninger, & Busse, 1996), and in past decades has been quite large in mathematically talented adolescents (Benbow & Stanley, 1983).

Because the mean differences on mathematics measures are not large, and because several recent books and reviews have discussed potential mechanisms underlying differences at the high end of the distribution (Ceci & Williams, 2007; Gallagher & Kaufman, 2005; Halpern et al., 2007), the Task Group does not provide an extensive review of these mechanisms. The Task Group notes that the differences in the ratio of boys to girls and men to women at the high end of mathematical performance is likely to be related to a combination of factors, including stereotypes and beliefs regarding the mathematical abilities of boys and girls; the advantage of boys and men in some forms of spatial cognition (these differences can be reduced with practice on spatial tasks; Terlecki, Newcombe, & Little, in press); greater interest of boys and men in abstract, theoretical occupations and activities; and, the typically greater variability of boys and men in many cognitive domains (for a review of the evidence see Halpern et al.). Additional studies that simultaneously assess all of these potential mechanisms are needed to determine the relative importance of each of them.

2. Race and Ethnicity

One explicit charge to the National Mathematics Advisory Panel is to determine the processes by which students from diverse backgrounds learn mathematics. It is widely documented that black and Hispanic students perform substantially less well in our nation's schools than their white and Asian counterparts. These achievement and attainment gaps are found across a host of schooling indexes, including grade point average; performance on district, state, and national achievement tests; rigorous course-taking; as well as, across behavioral indicators such as school drop-out, suspension and referral rates, and differential placements in special education, and programs for the talented and gifted.

a. The Achievement Gap

As documented in Appendix D and elsewhere, the mathematics performance gap is found from preschool to college (Ryan & Ryan, 2005), and across the full range of mathematical content areas. Even early on it tends to be manifested more on measures of mathematical concepts than on measures of mathematical computation (U.S. Department of Education, 2006; Hall, Davis, Bolen, & Chea, 1999).

It is instructive to examine mathematics performance differences for high schools that serve white, black and Hispanic students together. Byrnes (2003) has done so by analyzing NAEP outcomes. Results from this national data base show that, overall, these mixed race schools (and that had at least one student scoring above the 80th percentile in mathematics), enrolled 79%, 13%, and 8% white, black and Hispanic students respectively. Yet, among the students who scored at or above the 80th percentile in mathematics, 94% were white, whereas only 3% were black and 3% were Latino. Representing these numbers somewhat differently, 26% of the white students enrolled in these schools performed at or above the 80th percentile, as compared to only 7% of their black and Hispanic peers. White students were almost four times more likely than black and Hispanic students to reach this performance level.

Hughes (2003) found mathematics performance differences when comparing third-grade black and white students attending schools in a generally affluent school district. Specifically, differences were found even in the midst of a wealth of material and human resources available to black and white students. Elsewhere, Schmidt (2003) showed a black-white difference in performance on the TIMSS even when controlling for socioeconomic status (SES), and Nettles (2000) has reported a 100-150 point difference on the SAT that holds up across all income levels.

Defying easy explanation, in particular, are data from the most recent NAEP tests. Using average main NAEP mathematics scores for eighth-graders broken down by race and parents' highest level of education, it was found that 23 scale points separated black and white test scores for students whose parents did not finish high school. Yet, white scores were 37 scale points higher than black scores for students whose parents graduated from college. The pattern is similar for white-Hispanic test score differences. The gap favoring white students was 9 scale points for students whose parents did not finish high school but 22 scale points for children of college graduates. For 12th-graders, the white-black difference was 16 scale points for students of parents who did not finish high school; this difference jumped to 37 scale points for students whose parents were college graduates. The respective Hispanic-white differences were 8 and 22 scale points.

It seems that whatever explanations are offered for these patterns, they cannot simply be reduced to a focus on social standing or SES (to the extent that parents' education level is a marker for SES). The findings defy this straightforward explanation.

Attempts to close these achievement gaps should be done in ways that raise achievement for all students, while simultaneously raising levels at a steeper rate for black and Hispanic students.

b. Potential Sources of the Achievement Gap

In this section, the Task Group reviews research literature on potential explanations of why mathematics performance is comparatively low for black and Hispanic students, and potential approaches for raising their mathematics achievement levels.

Socioeconomic status (SES)

The conventional explanations for poor math performance for black and Hispanic students center on inadequate social experiences and learning opportunities linked to low socioeconomic status. Because black and Hispanic children are disproportionately poor, and because poor children perform less well, this then identifies the root cause of such performance deficiencies.

SES is a generic construct that has had many definitions over the years, including family income, parental education, and occupational prestige, among others. As documented in Appendix E, whether SES is defined in terms of parental education, poverty level, parental income, or a composite index, there is a consistent association between SES and mathematics achievement. The mechanisms linking these broad constructs to mathematical learning and achievement are not well understood, nor are the relationships among SES, ethnicity, and mathematics learning.

With respect to the latter issue, the findings are inconsistent. Stevenson, Chen, and Utal (1990) found that white-Hispanic-black differences on mathematics curriculum tests essentially disappeared when controlling for parental education and income level among fifth-grade but not third-grade students; yet, differences remained for reading scores across both grade levels. They also found that black and Hispanic mothers rated their children's performance in mathematics as more important than did white mothers. Schultz (1993) found that SES was a major predictor of mathematics performance for fourth- to sixth-grade black and Hispanic students from urban school districts.

Stewart (2006) focused on results obtained across multiple administrations of the National Education Longitudinal Study of 1988 (NELS:88) data set and found that among black secondary level students, the presence of household educational resources common in higher SES households, such as books, encyclopedias, and computers, predicted combined mathematics and science performance, in the 12th-grade. For these students, neither family income nor parental educational level was directly related to mathematics and science achievement. Hall et al. (1999) found that fifth- to eighth-graders' performance on the mathematics concepts and computations sections of the California Achievement Test was correlated with parental background (a measure which included but was not limited to highest level of formal education and highest math course taken) for white students but not for black students.

In another analysis of the NELS:88 data, Thomas (1999) found that both home-based and school-based factors predicted performance outcomes across ethnic groups. When controlling for school-based and home-based factors, the mathematics performance gaps across white, black, and Hispanic students diminished substantially. A similar result was obtained in a study by Byrnes (2003). Drawing on the NAEP for 12th-graders, classroom experiences and learning opportunity factors accounted for more of the variance in mathematics scores across white, black, and Hispanic students than did SES. After statistically controlling for differences in parental background and school-based factors, the performance gap among these groups was substantially reduced.

Are learning processes among ethnic groups similar or different?

The weight of evidence supports the conclusion that learning processes are more similar than different across ethnic groups. This is not to say that there are no differences in how children from different ethnic groups approach the learning of mathematics, but rather that there are many similarities.

Thomas (1999), for example, found that the configuration of variables that predict mathematics achievement for white, black, Hispanic, and Asian 10th-graders are generally the same. Stevens, Olivarez, Lan, and Tallent-Runnels (2004) also found that the same constellation of predictors of mathematics achievement generally held for white and Hispanic high school students. This result was essentially duplicated by Stevens and his colleagues (Stevens et al., 2006) in a study of Hispanic and white students from 4th through 10th grade.

In a more process-oriented study, Malloy and Jones (1998) found that the spontaneous approaches to mathematics problem solving that emerged in a sample of middle-class black eighth-graders were highly similar to those found in studies of white students. Similar results have been reported by Kuhn and Pease (2006) and Rhymer, Henington, Skinner, and Looby (1999).

c. Potential Cognitive and Social Influences

The literature focusing on cognitive and social influences on the mathematics learning and performance of black and Hispanic students does not have a sufficient number of experimental studies to provide definitive results. Much of the research in this area is correlational, but many studies have nonetheless incorporated sophisticated multivariate analyses that can be used to control for potential confounding variables and to provide at least weak tests of potential "causal pathways."

Many of these studies have drawn on national secondary data sets rather than on primary data. The advantages are large samples with results that can be generalized across the nation. The disadvantages include greater reliance on self-report data and constructs based on these data that are often formulated in a post hoc fashion, and thus may not measure the potential mechanism as precisely as is possible in an experimental study. Further, many of the studies have not formulated hypotheses about specific social or cognitive mechanisms, nor about whether there are racial or ethnic differences on mechanisms that can be changed in ways that help to close the performance gap.

Nevertheless, in recent years several hypothesized conceptions and processes have shown promise with respect to explaining and potentially narrowing ethnic differences in mathematics performance. Prominent among these are 1) stereotype threat; 2) cognitive load; 3) engagement, effort, and efficacy; 4) strategy use; 5) constructive and supportive academic interactions; 6) collaborative learning; and 7) culturally and socially meaningful learning contexts.

Stereotype threat

In the last decade, there has been increasing research attention given to the concept of stereotype threat as a contributing factor to group differences under certain specified conditions. This conception, first offered by Steele (1992), and then elaborated by Steele (1997), and Steele and Aronson (1995, 1998), hypothesizes that groups can be subjected to societal stereotypes that stigmatize their ability to perform in certain domains. For historical and sociological reasons, blacks have been viewed in the United States as having low intellectual ability and women as having low mathematical ability. These perceptions are particularly vexing because often attached to them is the presumption that the diminished ability is inherent and thus an unalterable characteristic of the group. When placed in a relevant performance setting, members of the stigmatized group are vulnerable to performing below their potential because of anxiety about upholding the negative stereotype.

Relation to ethnic differences in mathematics performance

Stereotype threat is a promising conception that has offered plausible explanations for certain group differences in academic performance. Unlike many other hypotheses in this area, stereotype threat has been investigated using experimental methods, giving greater confidence in results that confirm this hypothesis. Yet, there are limitations that temper claims that this concept can account for ethnic differences in mathematics performance among school-aged children and youths.

First, much of the research on stereotype threat and mathematics performance has focused on gender differences rather than on race or ethnicity differences. The study of stereotype threat in black and Hispanic samples has focused on general academic ability or intelligence; hence, the outcome measures for ethnic minority samples have typically been more general academic or test-performance related and not mathematics learning per se.

Second, the preponderance of the most rigorously executed research on stereotype threat has been done with college students. This is an important factor because the effects of stereotype threat are predicted to be more evident among group members who have a great investment in doing well and are typically high performers to begin with. In other words, a preoccupation with performance under stereotype-threat conditions is only predicted to affect performance when students are concerned about doing well on the task; students who are not invested in learning mathematics may not be influenced by any stereotype that involves mathematics.

As a result, there is not a sufficient research base testing the potential influence of stereotype threat in school-aged populations or focusing on mathematics performance of black and Hispanic students. Theoretically (as explained in the next section), it is unclear whether stereotype threat for mathematics can speak to the performance outcomes of black and Hispanic students who are not substantially invested in doing well in academic contexts. Nevertheless, studies addressing this issue are urgently needed.

Potential mechanisms

A recent study illustrates mechanisms that may link stereotype threat to performance outcomes. Keller (2007) investigated mathematics performance in a sample of 108 secondary level students in Germany (race was not specified, but they are presumably largely white). The students were randomly assigned to a stereotype threat or a no threat condition. Mathematics tasks were either difficult or easy. Those assigned to the threat condition were told in advance that for the mathematics tasks they were about to perform, gender differences in achievement had been found. The students in the no-threat condition were told gender differences had not been obtained. Further, the extent of identification with doing well in mathematics was assessed for all participants. Girls who value doing well in math and who were placed in the threat condition had larger decreases in mathematics task performance, from a pre-established baseline, when they worked on more difficult items. For difficult items, girls who did not value doing well in math performed better under the threat condition than under the nonthreat condition. There were no effects for the easy items. A similar result with a college student sample was obtained in an experiment by Beilock et al. (2007).

Ryan and Ryan (2005) offered a conceptual model for the processes underlying how stereotype threat influences quality of academic performance. When the conditions of stereotype threat are present—for individuals in which the domain is of importance to them and a negative stereotype exists—reminding the individual of the stereotype results in performance avoidance goals. These in turn result in heightened anxiety and lowered self-efficacy. As with mathematics anxiety, heightened anxiety under these conditions can result in thoughts about competence intruding into working memory, which functionally lowers this core capacity. Poor self-efficacy can result in diminished effort when problems become difficult. Although no study to date has tested the full model proposed by Ryan and Ryan, recent research has confirmed that each of these processes individually is linked to lowered performance outcomes in the face of stereotype threat.

Consider the work of Smith, Sansone, and White (2007) involving a sample of white college females. They found that in the presence of a salient stereotype threat, participants who were high on achievement motivation were more likely to spontaneously adopt performance avoidance goals when working on a mathematics task than were students who were not high in achievement motivation. Schmader and Johns (2003) provided evidence consistent with the hypothesis that stereotype threat interferes with mathematics performance by reducing individuals' working memory capacity. In this investigation, white men did better than white women on a mathematics task in a stereotype threat condition, and this difference was associated with reduced working memory resources for the women. No gender differences on the mathematics task or a working memory measure were found in a nonthreat control condition. Additional research revealed essentially the same pattern for Hispanic students.

For a sample of undergraduate women, Beilock et al. (2007) extended the work of Schmader and Johns (2003) by using a mathematics task where the level of working memory demands could be manipulated. Women were assigned to either a threat or nonthreat condition and asked to solve high- and low-demand problems. For women in the threat condition, performance was particularly poor for high-demand problems. These women reported worries about the task and had thoughts about confirming the stereotype during problem solving; women in the nonthreat condition did not report these concerns. The authors reasoned that these thoughts and worries functionally reduced working memory capacity which resulted in worse performance on high-demand problems. These results confirm the hypothesis that threat can result in intrusive thoughts about confirming the stereotype—thoughts that in turn lower working memory capacity and thereby lower performance.

Ryan and Ryan (2005) also hypothesized that anxiety could influence performance under conditions of stereotype threat, and there is some supporting evidence. The work of Osborne (2007) is notable in this regard. His research was also done with college students, but race of participants was not specified. Men and women were randomly assigned to either a threat or a nonthreat condition. For women, when using indexes of heightened anxiety, there were lowered levels of skin temperature, elevated levels of skin conductance, and heightened levels of diastolic blood pressure under the threat condition. No gender differences in physiological reactance occurred under the nonthreat condition. Moreover, women performed worse than men on the mathematics measure in the threat condition but not in the nonthreat condition. Ben-Zeev, Fein, and Inzlicht (2005) also found evidence for heightened arousal levels in women under conditions of stereotype threat.

Curiously few investigations have tested ways to alleviate the adverse influences of stereotype threat on performance. In one of the few studies that have done so, Beilock et al. (2007) found that extended practice on the difficult mathematics problems, which should make solving these problems more automatic and less dependent on working memory, eliminated the decrease in performance associated with stereotype threat.

A study by Good, Aronson, and Inzlicht (2003) is unique in that they attempted to enhance performance for stereotyped groups through a systematic intervention for school-aged children. This was a field experiment employing a sample of predominantly low-income, predominantly ethnic-minority seventh-graders; 67% Hispanic, 13% black, and 20% white. For the treatment condition, these students were mentored across an academic year by college students who encouraged them to regard intelligence as pliable rather than fixed and/or to attribute academic difficulties in the seventh grade to the uniqueness of the academic setting; but mentors also explained that academic performances can be improved over time. A control group of students was provided information linked to an antidrug campaign. The outcome measure was performance on a statewide standardized test of mathematics and reading achievement. The results revealed that girls' performance was substantially better in mathematics under the treatment condition than under the control condition. Boys performed essentially the same across conditions, with the exception of marginally significant ($p < .06$) better performance in the treatment condition (i.e., mentored) than the control condition. For reading, there was an overall main effect (across gender) for condition such that treatment students did better than control students.

These are striking results, but in this investigation, stereotype threat was not directly manipulated. The findings are encouraging in that academic performance was significantly improved in groups that often are stereotyped as doing poorly on academic measures. Because of the design of the study, however, it is not known if the improved performance was due to alleviation of vulnerability to stereotype threat or to other factors such as increased effort.

Cognitive load

As the Task Group described in previous sections, there is considerable evidence that when the working memory system is overloaded, performance in many domains including mathematics suffers. Putting in place procedures to reduce this load can enhance performance. The Task Group has documented how task practice leads to more automatic processing and thus reduces the working memory demands of the task. In the previous section, it was reported that practice at a task reduced vulnerability to stereotype threat in a sample of college women. Interventions that reduce cognitive load should improve the performance of all students. It would seem to follow that interventions which improve working memory functioning for low-achieving black and Hispanic students have high potential value.

Engagement, effort, and self-efficacy

In the earlier section in this report on Social, Affective and Motivational Influences on Learning, the Task Group reviewed work that indicated in general the positive influences that engagement, effort, and self-efficacy can have on mathematics performance. In reviewing research more specifically targeted to mathematics learning and performance of black and Hispanic students, the evidence strongly suggests that to the extent that such processes are positively manifested, mathematics performance can be improved. Findings in support of this conclusion have been documented across the full kindergarten to 12th-grade spectrum. These factors are more likely to be linked directly, rather than indirectly (e.g., as indexed by SES), to mathematics performance, and account for much more variance in mathematics outcomes than do global family background factors. Moreover, these processes are substantially malleable and can be changed in learning and classroom settings.

On the other hand, evidence suggests that important processes, such as effort devoted to school performance, are comparatively low for black and Hispanic students in traditional learning and classroom settings. The typical research investigating processes such as engagement, effort, and efficacy in black and Hispanic populations has not made use of experimental paradigms. These types of studies need to be conducted to determine how and why these processes influence mathematics learning and performance in ethnic minority populations, and how they can be improved in these populations.

In recent years, research has documented that general motivation level is functionally linked to mathematics outcomes for black and Hispanic students. A recent study by Borman and Overman (2004) is a case in point. They set out to determine the factors that differentiate between academically successful and unsuccessful black, Hispanic, and white students from low-income backgrounds. They examined such students' trajectory from third-grade to sixth-grade performance using the Comprehensive Test of Basic Skills, Fourth Edition (CTBS/4) math scores from the *Prospects* national data set. This was a congressionally mandated study conducted between 1991 and 1994 as part of the federal evaluation of Title I at the elementary school level. The focus was students who performed comparably in the third grade but whose performance diverged substantially in the sixth grade. Students whose scores increased substantially were labeled *resilient* and those whose scores declined were termed *nonresilient*. The percentile ranks for the two groups were 39th and 38th respectively in third grade. In the sixth grade, the percentile ranks were 59th and 11th, respectively, for the resilient versus the nonresilient group. Students were polled each of the four years of the investigation on certain beliefs, attitudes, and practices pertaining to their schooling experiences, and for each factor, average ratings were calculated. One factor that distinguished the resilient from the nonresilient children was having a positive attitude toward school.

In the previously cited Stewart (2006) study, the one factor that stood out as a predictor of combined mathematics and science achievement for the black students was general motivation level. This measure included items such as the importance of getting good grades and satisfaction from doing well in school. A similar result was obtained by Byrnes (2003) with 12th-grade black and Hispanic students, as well as white students. In a recent study, Balfanz and Byrnes (2006) found that self-reported effort emerged as a significant predictor of yearly gains in mathematics performance for black and Hispanic middle school students from an "urban background;" the gains were in terms of whether the students' performance exceeded what would have been expected by average yearly grade-equivalent increments. This outcome, by implication, suggests that interventions such as the one described earlier (Blackwell et al., 2007) which focus on the importance and malleability of effort, have the potential to help reduce achievement differences in mathematics across racial and ethnic groups.

Sirin and Rogers-Sirin (2004) found that student engagement in school was among the two strongest correlates (among a host of variables) of combined math and English grades in a sample of middle class adolescent black students. Borman and Overman (2004) also found that student engagement differentiated between the academically successful and nonsuccessful students. In this investigation, student engagement was not a self-reported measured but instead was indexed by the extent to which teachers agreed that a student conveyed attitudes and manifested behaviors indicative of an interest in school work and a desire to learn.

It is of interest that Borman and Overman (2004) also reported significant race differences for predictor variables. It was found that black students overall had substantially lower student engagement scores than did their white and Hispanic counterparts. However, the previously described experimental study by Blackwell et al. (2007) indicates that engagement scores can be raised for low-income minority students through certain targeted interventions. They deployed an intervention strategy similar to that used by Good et al. (2003). For a description of Blackwell et al., see the *Goals and Beliefs About Learning* section in this report.

Self-efficacy has also been found to be an important correlate of mathematics achievement. In the Borman and Overman (2004) study, self-efficacy differentiated between resilient and nonresilient students. Elsewhere, Stevens et al. (2006) reported that across 4th to 10th grade self-efficacy was a significant correlate of math achievement for Hispanic and white students (SES level was not reported). Similar findings have been obtained in many other recent studies; Navarro, Flores, and Worthington (2007) for Mexican-American 8th-graders; Long, Monoi, Harper, Knoblauch, and Murphy (2007) for black low-income 8th- and 9th-graders; Stevens et al. (2004) for Hispanic and white 9th- and 10th-graders (41% of the sample were from low-income backgrounds); Byrnes (2003) for white, black, and Hispanic 12th-graders.

Two studies have found that Hispanic students have lower-levels of self-efficacy, on average, than their counterparts from other ethnic groups. In the Borman and Overman (2004) study, Hispanic students had lower self-efficacy scores than did black or white students (d = .27), and in the Stevens et al. (2004) study, Mexican American students had lower mathematics self-efficacy than their white school counterparts (d = .25).

These results, however, do not directly address the questions of the antecedents of self-efficacy and the factor(s) that can increase self-efficacy. At least two studies speak to these issues for ethnic minority populations. For a sample of black high school students, Gutman (2006) found that exposure to mastery goals in the classroom were associated with increased mathematics self-efficacy, as well as to higher mathematics grades. Similarly, students who espoused mastery goals had higher mathematics self-efficacy and higher mathematics course grades.

In a related study, Fuchs et al. (1998) produced noteworthy results through an intervention experiment designed to heighten students’ mastery goal orientations. For the relevant part of this investigation, participants were second- to fourth-graders who began the school year at or near the bottom of their classes in mathematics performance; 78% of these participants were black. The dependent variable was performance on a curriculum-based mathematics test at the end of the school year. Instruction focused on fostering mastery-oriented beliefs through targeted activities across a full 17–18 weeks of the academic year. Over the course of the study, students were also provided opportunities to receive assessment feedback. Students were randomly assigned to one of three conditions: 1) mastery-focused plus assessment feedback, 2) assessment feedback only, and 3) a standard classroom instruction control condition. It was found that the mastery plus assessment treatment led to the highest end-of-year mathematics test scores, followed by the assessment only condition, whose participants in turn had higher scores than those receiving only the standard classroom instruction (for the mastery versus control difference, d = .94; for the mastery versus assessment feedback only difference, d = .42; and for the assessment feedback versus control difference, d = .43).

Strategy use

Studies of explicit instruction of problem-solving strategies indicate it is a potentially useful intervention for improving the mathematics achievement in racial and ethnic minority populations. Although strategy use has generally been understood to foster greater academic performance (e.g., Pressley and Woloshyn, 1995), much of this work has centered on reading performance. Of the many studies that have focused on mathematics, only a few have focused squarely on racial and ethnic minority populations.

In a study of the correlates of mathematics performance, Schultz (1993) found that for black and Hispanic fourth- through sixth-graders, higher self-reported academic motivation (for which self-regulatory strategies figured prominently) was associated with higher mathematics achievement test scores. Malloy and Jones (1998) found that in comparing successful and unsuccessful mathematics problem solvers among their sample of black eighth-graders, the more successful students were more likely to use a mix of strategies and more often verified their procedures than their less successful peers. The less successful students often guessed. Examples of the strategies employed by the successful students were drawing diagrams, looking for patterns, or systematic guessing and checking. Among the verification procedures employed were rereading problems, checking calculations, or re-doing the problems.

Fuson, Smith, and Lo Cicero (1997) conducted a classroom based year-long intervention with first-grade Hispanic students from low-income backgrounds to determine if explicitly teaching certain strategies would improve their mathematics outcomes. Specifically, the children were taught to think of two-digit numbers as quantities of 10s and 1s. By year's end, these children could add and subtract two digit numbers with regrouping on par with similarly aged children in eastern Asian nations.

Learning opportunities and constructive, supportive academic interactions

At the heart of Walberg's (1984) productivity model is the assumption that students will learn more if they are given more opportunities, more contact, and more exposure to settings where they can actually learn what is demanded, expected, or required of them. Correlational and quasi-experimental evidence supports this claim. There is also evidence that the broader settings in which the learning occurs can be important. Specifically, socially supportive learning contexts are tied to enhanced academic performance (Patrick, Kaplan, & Ryan, 2007), and there is accumulating evidence that these contexts are particularly effective for black and Hispanic.

With respect to learning opportunities, Byrnes' (2003) earlier described study is especially telling. Byrnes used several opportunities to learn variables, along with key attitudinal variables. Among these variables were number of algebra and calculus courses taken; use of worksheets; and student attitudinal factors such as self-efficacy in relation to and liking of (taken together as a composite variable) mathematics, perceived utility of mathematics, and the perception that mathematics is more than just memorization. When comparing white versus black and Hispanic students who scored at or above the 80th percentile in mathematics performance on the NAEP tests, there were no differences across these variables.

However when Byrnes (2003) compared black and Hispanic students who scored above the 80th percentile to black and Hispanic students who scored below this level, notable differences were found. Eighty-five percent of the minority students who scored above the 80th percentile had taken courses beyond Algebra I, whereas only 47% of minority students who scored below the 80th percentile took these courses. Twenty-nine percent of those scoring above the 80th percentile had worksheets at least once a week versus 59% of those below the 80th percentile. Sixty-nine percent of those above the 80th percentile expressed self-efficacy for/ liking of mathematics, as compared to only 35% of those scoring below the 80th percentile. Moreover, 75% of those scoring above the 80th percentile agreed that mathematics is more than just memorization, but this was found for only 25% of lower-scoring students. In contrast, perceived utility of mathematics was not a differentiating factor in these comparisons. For that matter, it was not a predictor of mathematics outcomes in this study. Yet, courses beyond algebra, worksheet use, and math memorization were all significant predictors (self-efficacy and liking as significant predictors were discussed in a previous section). While these are correlational data where cause and effect cannot be determined, the study nevertheless reveals significant difference within black and Hispanic students in attitudes towards and views of mathematics.

With respect to constructive and supportive social interactions, a qualitative study by Brand, Glasson, and Green (2006) deserves mention. They conducted in-depth interviews with five black students (four high school seniors and one college freshman) who were participating in a program designed to encourage them to become teachers. This is a highly selective program, in which students who finish high school are guaranteed four-year scholarships to college. Among other things, students in the study were asked to describe their experiences in mathematics class. One central theme across students, in terms of school success, was having meaningful interactions with their teachers. This was taken to mean experiences that included having teachers who validated their capabilities, were accessible and approachable, were supportive, and held high expectations for them.

These qualitative insights are consistent with empirical data from other investigations. Mooney and Thornton (1999) polled black and white seventh-graders from a range of SES backgrounds regarding their attributions for success in school. Although the relative endorsement of the various attribution types was the same within race (for example, effort was most favored by both black and white students), cross-race comparisons revealed some important differences. White students, more so than their black counterparts, attributed success to a student's own abilities. Black students, in contrast and to a much greater extent than white students, attributed success to rapport with their teachers. Also worth noting is a study by Casteel (1997) that asked the question, "Whom do you most want to please with your class work?" Of the more than 1,600 black and white middle school respondents from diverse SES backgrounds, 71% of all black respondents answered "my teacher," whereas only 30% of the white respondents answered in this way. The more common response from the white students was "my parents." This pattern of results suggests that for many black students, they do not just learn *from* their teachers, but also they learn *for* their teachers.

In a study focusing on low-income black students in 1st to 12th grades, Tucker and colleagues (2002) found that higher levels of classroom engagement were found when students reported that teachers were caring and interested in their doing well in school, and showed a personal interest in them. This teacher variable was the strongest predictor of student engagement in the study. In fact, the path analysis indicated that student engagement in class was directly related to this teacher factor; this pattern of findings had not been found in previous studies of white students. Other aspects of teacher behavior such as teacher structure—the extent teachers have fair and consistent consequences in response to student behavior, or provide clear feedback—influenced engagement only indirectly.

Other studies demonstrate the connection between interpersonal academic context and mathematics performance. In the Borman and Overman (2004) study, another variable that differentiated resilient from nonresilient elementary students in their mathematics test performance was positive teacher-student interactions in the classroom.

In a study of 12th-grade black students, Stewart (2006) found that a positive perception of the school environment (i.e., the perception that students get along well with teachers, have caring teachers, and teachers provide praise for good efforts) was a significant predictor of mathematics and science achievement. Elsewhere, Balfanz & Byrne (2006) report that the greater the number of "supportive classrooms" middle school black and Latino students participated in over time, the more likely the math performance gap would be closed between them and other racial/ethnic groups.

These results are consistent with claims concerning importance of supportive social contexts, especially support provided by teachers, for the mathematics achievement of black and Hispanic students. However, definitive results await use of experimental tests of potential causal mechanism. One possibility is that these social contexts result in greater engagement and increased effort in the classroom and through this better mathematics achievement. Another possibility is that these contexts reduce stereotype threat effects, namely cognitive overload, increased anxiety, or the promotion of performance avoidance goals. Perhaps students from certain social or cultural backgrounds have been socialized such that they are more responsive to the combined power of the school and classroom context.

Research, however, has been gathered in recent years to suggest that the interpersonal relationships that do occur in classrooms serving low-income African American and Hispanic students are often not supportive and may result in disengagement. Ferguson's (2003) research review and analyses are relevant to this point. He has gathered evidence that teacher expectations of future performance for black students are regularly more negative than they are for their white counterparts. Further, teacher perceptions and future expectations may affect future mathematics performance of black students, both positively and negatively, to a greater degree than that of white students. On the basis of data he had reported in 1998, Ferguson found that the relative influence of the teacher was nearly three times larger for black than white students in elementary school, whether the outcome was mathematics grades or mathematics achievement scores. In the Ferguson (1998) article, the data of interest examined the extent to which teacher perceptions of students' "performance, talent, and effort" measured in the fall semester predicted students' math achievement scores and math grades the following spring term (p. 286). The corresponding effect sizes for the prediction were .14 and .37 for white and black students, respectively, on the math achievement test, and .20 and .56, respectively, for mathematics grades.

A recent meta-analysis provides further evidence concerning teacher expectations (Tenenbaum & Ruck, 2007). Their review covered research done between 1968 and 2003. The majority of these reviewed studies focused on elementary students only (approximately 60%). The remainder included students at the secondary or university level, students across school levels, or in a few cases, unspecified sample characteristics. They found that teachers had more positive expectations for white than for black ($d = .25$) or Hispanic students ($d = .46$). Moreover, teachers directed more positive speech in the form of praise, affirmations, and positive feedback toward white than minority children. White students also received more product- and process-based questions, and therefore black and Hispanic students had fewer overall opportunities to respond academically in their classrooms. At the same time, the review did not reveal differences in the amount of negative speech directed at white, black, or Hispanic children. A study by Hauser-Cram, Sirin, and Stipek (2003) adds another dimension to this line of inquiry. They found that elementary school teachers held lower expectations for the future mathematics success of their current students to the extent that they perceived social and educational value differences between themselves and a student's parents. Although the finding was marginally significant ($p < .06$), elementary school teachers also perceived the difference between themselves and parents to be larger for black parents than for white parents.

Collaborative learning

The available evidence suggests that when properly structured, generally speaking, collaboration for learning can have a positive influence on mathematics performance and may be relatively important for minority students, particular those from low-income backgrounds. This finding appears to be especially robust at the elementary school level. Research for or against the effectiveness of collaborative learning at the middle and high school level is generally absent from the literature.

Perhaps the best source to assess the effects of collaborative learning on mathematics outcomes in elementary school comes from a meta-analytic review conducted by Rohrbeck, Fantuzzo, Ginsburg-Block, and Miller (2003). They set certain conditions for inclusion of studies in their review. Among these criteria, there had to be ethnic group comparisons, explicit peer assistance with interdependent reward contingencies, and independent accountability or evaluation procedures. The latter two conditions were necessary because the extant literature on peer-assisted learning indicates these conditions are crucial to positive outcomes. Other qualifiers were that all studies had to appear in peer-reviewed journals, and had to have used experimental or quasi-experimental designs. Moreover, the interventions had to be classroom-based and occur for more than 1 week. Ninety studies published between the years 1966 and 2000 met these criteria.

Overall, peer-assisted learning led to greater mathematics performance outcomes than did individual or competitively structured learning. But, the magnitude of these effects varied. Larger effects were found for: 1) urban versus rural and suburban settings, 2) low-SES versus middle and higher SES, and 3) minority status (black and Hispanic) versus majority (white) status. The effect sizes were .44 and .23 for urban and suburban/rural locations, respectively. In the case of SES, the mean effect size was .45 when more than 50% of the sample was low SES, and .32 when less than 50 % of the sample size was low SES. For minority status, the mean effect size was .51 when more than 50 % of the sample was minority status (black and/or Hispanic) and .23 when less than 50 % of the sample size was of minority status. The largest effect size was obtained for samples consisting of primarily black and/or Hispanic students, and the magnitude of the effect of collaborative learning on mathematics performance was largest when contrasting these ethnic minority children with their white counterparts.

To illustrate the type of research assessed in this meta-analysis, consider a study conducted by Ginsburg-Block and Fantuzzo (1997). These researchers contrasted a reciprocal peer-tutoring dyad condition with a condition where students worked individually. The dyads met 2 times a week across a 10 week intervention period. The sample consisted of fourth- to sixth-grade black students from low-income backgrounds. For the reciprocal peer-tutoring condition, the two students alternated between tutor and tutee. As the tutee answers test questions or performs a given task, the tutor prompts, provides feedback, and offers evaluative comments. The dyad work toward a common goal, that is, their reinforcement was contingent on the performance of both students. In this study, participants in the reciprocal peer-tutoring condition had higher mathematics classroom performance outcomes than those in the practice control condition; and they received higher ratings of teacher-observed task-relevant behaviors during mathematics lessons. It was also found that these reported engagement levels were positively related to scores on a mathematics curriculum-based computation test.

Socially and culturally meaningful learning contexts

One final area with promising research is with respect to socially and culturally meaningful learning contexts. The goal is to better link what happens in school to experiences, values, and practices that are salient in the lives of black and Hispanic students (Perry, Steele, & Hillard, 2003; Ladson-Billings, 1997; Moll, Amanti, & Neff, 2005; Sternberg, 2006). Much of the actual scholarship done to establish such links has typically not brought systematic empirical research to bear, and what empirical research that has been done, has most typically been linked to reading rather than math performance. There are a few exceptions, but even these have not used experimental methods or explored underlying processes that directly link social and cultural processes to academic outcomes. Yet, results from these few recent investigations indicate that this is an area of investigation that merits further study.

A recent randomized field experiment by Cohen, Garcia, Apfel, and Master (2006) addressed the usefulness of linking school performance to matters of personal relevance for black students. The participants were black and white seventh-graders from middle- to lower-middle-class backgrounds, and they all attended the same school. The authors describe their experimental treatment as a *self-affirmation* intervention. Early in the school year, students randomly assigned to this condition selected one or more of their most important values and then wrote brief paragraphs in which they justified why these values were chosen. The exercise was presented to the students as a normal lesson and took approximately 15 minutes to complete. After this exercise was completed, the teacher resumed the focal subject lesson. Students who had been randomly assigned to the "control" condition were asked to select one or more of the least personally important value(s) and write about why these values might be important to someone else. The same procedural protocol was followed for the control condition. Teachers themselves were blind to which students participated in what condition. Two parallel studies were conducted, separated by one year. In the first, students completed the exercise once; in the second, students completed the exercise twice in the fall semester. In the first, participants wrote about only one value; in the second they could choose up to three.

For both studies, the first semester course grades of the black students in the treatment condition were significantly higher than those obtained for the black students who participated in the control condition. No treatment effect was obtained for the white students in either study. The black students in the treatment condition did even better than their black control group counterparts in other courses for which the treatment did not occur. For this investigation, the actual course subject in which the treatment was provided was not specified. But, given that the authors stated that the subject was not one linked to gender stereotype, it is very likely these were not mathematics classrooms.

Although the Task Group noted earlier in the report concerns about sociocultural claims regarding learning, and specifically that many claims have not been scientifically evaluated, there are several studies from the sociocultural perspective that might provide insights for more fully interpreting some of the results described earlier.

Another relevant approach has been to focus on cultural values or themes that may be more prominent in certain populations than in others and that may enhance learning and performance outcomes for these populations. One such theme is *communalism* (Boykin, 1986; Boykin & Ellison, 1995), which has been hypothesized as being particularly prominent for many people of African descent, including African Americans. To be sure, there is no claim that all black people are communal or that communalism is a fixed trait of a given person or group of people. Rather, if this theme is more salient in the communities of blacks, then the corresponding social expectations may influence how children interpret and perform in school settings (Boykin & Allen, 2004).

Communalism as a cultural theme implies that a premium is placed on collaborative interdependence. If this is salient for many black Americans, then this could be one factor potentially contributing to the receptiveness to collaborative learning for black students described in the previous section. Moreover, if communalism is a culturally meaningful theme, then performance enhancements in group settings would occur even in the absence of individual incentives to perform well in collaborative settings, as found with reciprocal peer tutoring. This hypothesis has been tested in several experimental investigations, but only in a handful that have used mathematics achievement as an outcome variable.

One such experiment was conducted by Hurley, Boykin and Allen (2005). In this study, fifth-grade black children from low-income backgrounds were given opportunities to learn effective strategies for solving mathematics estimation problems, and then to examine their subsequent performance on a mathematics estimation test. Students were randomly assigned to one of two conditions. In one condition, and after completing a 15-item mathematics estimation pretest that used grade-appropriate multiplication problems, children were given a 20-minute practice exercise in which they had to complete a workbook to help them become more facile with mathematics estimation. During the learning/practice phase, these students were encouraged to work alone and prompted to exercise their individual effort and autonomy. They were also offered a reward if their posttest performance reached a certain criterion level. This was the individual learning condition. The other children were assigned to the communal condition. After the pretest, these children were formed into groups of three and given a prompt during the learning phase. The prompt emphasized the importance of working together for the good of the group so that everyone in the group could benefit and learn that it is important to help each other. These children were not offered a reward for good performance. They were told to work together but not told how to work together. It was reasoned that if interdependence were a salient theme for them, they would not need external incentives to do well, nor would they require explicit instructions on how to work together. Participants in both conditions worked on the follow-up 15-item mathematics estimation posttest on an individual basis. Results revealed that performance on the posttest was superior for those who had worked in the communal learning condition ($d = .56$).

This study had certain limitations, not the least of which was that the intervention only lasted for 20 minutes. However, results reported in a recent doctoral dissertation (Coleman, 2003) tentatively suggested that these effects can extend across a 4-week intervention done in conjunction with the actual classroom teaching of a fractions unit to third- and fourth-grade low-income black students.

An intriguing study with an international comparison is also worth mentioning. Huntsinger, Jose, Fong-Ruey, and Wei-Di (1997) examined cultural differences in early mathematics learning among European Americans, second-generation Chinese Americans, and Chinese students residing in Taiwan. They sought to determine if there were differences in family practices related to mathematics and if any family differences were related to children's mathematics outcomes in school. The focus was on children at the preschool and kindergarten levels. Families from all three comparison groups were from middle-class/professional backgrounds. It was found that Chinese Americans and Chinese families in Taiwan structured more daily time for homework or music practice and encouraged their children to participate in mathematics-related activities more so than did the white parents. Chinese American parents, and to a lesser extent the Chinese parents in Taiwan, engaged in more direct, formal teaching of mathematics to their children than did white parents. Moreover, Chinese American and Chinese children in Taiwan performed better than white children on the Test of Early Mathematics Ability (TEMA-2).

It was found that children who received more formal teaching at home and spent more time doing homework had higher mathematics test scores. Certainly this was a fairly small-scale study, with narrow SES backgrounds of the participants. Furthermore, it is not clear if the differences were actually due to cultural value factors, per se. But the implications for the importance of family organization of children's activities, as it relates to mathematics outcomes, is relevant to the Task Group's review of group differences in the mathematics competencies children bring to school.

3. Learning Disabilities

At least 5% of students will experience a significant mathematics learning disability (MLD) before completing high school, and many more children will show learning difficulties in specific mathematical content areas. Intervention studies are in the early stages and should be a focus of future research efforts. Further research also is needed to identify the sources of MLD and learning difficulties in the areas of fractions, geometry, and algebra.

The issues of diagnostic criteria and the percentage of children with an MLD remain to be fully resolved. Change in the stringency of the diagnostic criteria (e.g., cutoff on a mathematics achievement test) used to diagnose MLD can significantly influence the pattern of identified deficits and explains some differences in results across studies (Murphy, Mazzocco, Hanich, & Early, 2007). Nevertheless, progress has been made in the past decade. Using a population-based birth cohort sample that provided medical, academic, and other information on 5,718 individuals from birth to age 19 years, Barbaresi et al. (2005) assessed the incidence of MLD using different diagnostic criteria. On the basis of the two criteria that involved at least a one standard-deviation difference between an intelligence quotient (IQ) score and a math achievement score, 6% to 10% of children showed evidence of MLD before they completed high school (the potential relations among IQ, mathematics learning, and MLD are not yet known and thus control of IQ is important). An additional 6% of children were diagnosed as MLD using a more lenient criterion. The two discrepancy-based criteria yielded estimates similar to the 5% to 8% of children estimated as having MLD in previous studies (Badian, 1983; Kosc, 1974; Gross-Tsur, Manor, & Shalev, 1996; Ostad, 1998; Shalev, Manor, & Gross-Tsur, 2005). In one of these studies, Shalev and colleagues identified 5% of 3,029 5th-graders as having MLD and found that 40% of these children remained at or below the 5th percentile in math achievement in 11th grade. Almost all of the remaining children were in the lowest quartile in math achievement, despite average IQ scores, and most would have been diagnosed as MLD using at least one of the Barbaresi et al. criteria. The pattern across studies suggests that 5% to 10% of children will meet at least one relatively strict criterion for MLD before reaching adulthood and at least another 5% might be diagnosed as MLD using more lenient criteria.

These large-scale studies are important for identifying the percentage of children who likely have some form of MLD. Although they do not provide detailed information on the nature of the underlying deficits in mathematics learning or in the cognitive mechanisms (e.g., working memory) that contribute to these deficits they are nonetheless informative regarding the early deficits of children with MLD and illustrate the usefulness of this approach for studying learning disabilities in other areas of mathematics.

The cognitive and neuropsychological studies have revealed several sources of the poor early learning of arithmetic by children with MLD or other low-achieving children (Geary, 2004; Jordan et al., 2003; Ostad, 1998). The first involves delayed adoption of efficient counting procedures for problem solving and is manifest as frequent reliance on finger counting, infrequent use of the counting-on procedure, and frequent counting errors (Geary, 1990). The reliance on finger counting and the frequent counting errors are related to below-average working memory capacity. The delayed adoption of counting-on is related to a poor conceptual understanding of some counting concepts (Geary et al., 2004) and may also reflect a poor understanding of number and quantity per se (Butterworth & Reigosa, 2007). Many children with MLD eventually develop normal procedural competencies for solving simple arithmetic problems, although they usually do so several years after their peers.

A second source of the low achievement of these children involves difficulties in the learning or retrieving of basic facts (Jordan & Montani, 1997; Russell & Ginsburg, 1984). This is not to say these children never correctly retrieve answers, but rather that they correctly retrieve basic facts less often; at times, they also generate different pattern of retrieval errors. Although not conclusive, evidence to date suggests two potential sources of these difficulties. The first involves the formation of long-term memory representations of basic facts, and the second involve interference during the retrieval process (Barrouillet, Fayol, & Lathuliére, 1997; Geary, Hamson, & Hoard, 2000); interference is related to attentional and inhibitory control mechanisms of the central executive component of working memory. Whatever the source, short-term longitudinal and cross-sectional studies suggest that the difficulty in learning or retrieving basic facts is more persistent than the procedural delay (Jordan et al., 2003).

The central executive component of working memory has also been implicated in the procedural delays of children with MLD (e.g., Geary et al., 2007; McLean & Hitch, 1999; Swanson, 1993; Swanson & Sachse-Lee, 2001), and their deficits in this core cognitive competency will almost certainly result in delayed learning in novel and complex mathematical topics. The two other core components of working memory—the phonological loop and visuospatial sketch pad—may also contribute to MLD but in more circumscribed ways.

Butterworth and colleagues, however, have proposed that a poor "number sense" is the core deficit for children with MLD (Butterworth & Reigosa, 2007; Landerl et al., 2003). Number sense is defined in terms of the competencies that are evident in infants and young children and do not require formal schooling. These would involve, as an example, the ability to quickly subitize, or determine with a quick glance without counting that the quantity represented by ●● is less than that represented by ●●●. Deficits in these very fundamental areas would impede the learning of arithmetic in school. There is evidence consistent with the view that children with MLD have deficits in such areas, independent of any deficits in the central executive (e.g., Koontz & Berch, 1996; Jordan et al., 2003; Landerl et al.).

The studies of number sense deficits tend to include children often with average cognitive ability but with lower achievement, as evidenced by test scores, than is typical for research on MLD. It is possible that there are multiple forms of MLD. Central executive deficits would result in a broad range of deficits in mathematics and other areas. Deficits in the number sense system—potentially involving the intraparietal sulcus (see section on Brain Sciences and Mathematics Learning)—would be associated primarily with difficulties understanding quantity and, of course, with all of the mathematics dependent on this knowledge.

Much less is known about MLDs in relation to learning other areas of arithmetic, and very little is known about the specific deficits associated with learning fractions, estimation, geometry, or algebra. The work that has been conducted suggests that children with MLD, and often more general learning disabilities, have difficulties with arithmetic algorithms (Russell & Ginsburg, 1984), quantitative estimation (Hanich, Jordan, Kaplan, & Dick, 2001), rationale numbers (Mazzocco & Devlin, in press), and with algebraic equations and word problems (Hutchinson, 1993; Ives, 2007). Further studies about learning disabilities and learning processes in these and related areas of mathematics are needed, as are studies of the underlying cognitive mechanisms (e.g., central executive component of working memory and basic number knowledge) and brain systems (e.g., areas of the prefrontal cortex that support working memory, and areas of the parietal cortex that support number-related processes and representations; see section on Brain Sciences and Mathematical Learning).

The Task Group also notes that many students with MLD have comorbid reading disabilities or attentional difficulties. Whereas it is known that children with such multiple deficits have more difficulty learning in many areas of mathematics than do children with MLD and no other deficits, the sources of the comorbidity are not well understood.

4. Gifted Students

There are only a few cognitive studies of the sources of the accelerated learning of mathematically gifted students, but those that have been conducted suggest an enhanced ability to remember and process numerical and spatial information. Quasi-experimental and longitudinal studies consistently reveal that accelerated and demanding instruction is needed for these students to reach their full potential in mathematics.

In most academic domains, gifted children achieve the same academic milestones as their more typical peers but do so at an earlier age (for reviews and discussion see Benbow & Lubinski, 1996; Siegler & Kotovsky, 1986). On the basis of this general pattern, intellectually or mathematically gifted children are predicted to learn arithmetic, fractions, algebra, and other areas of mathematics at an earlier age and in many cases with less exposure than other children. There are only a handful of cognitive studies of the processes that might underlie this accelerated learning in mathematics, and even in these studies, the criteria used to define giftedness has varied considerably (Dark & Benbow, 1990, 1991; Geary & Brown, 1991; Mills et al., 1993; Robinson et al., 1996; Swanson, 2006). Nonetheless, the results of these studies suggest an enhanced ability to retrieve spatial and numerical (but not verbal) information from long-term memory and an enhanced ability to manipulate these forms of information in working memory; the extent to which these advantages are learned, inherent, or some combination is not known. Cognitive and developmental studies of children who show promise in the learning of mathematics are clearly needed to better understand the sources of their advantage and to better facilitate their long-term mathematical development.

Even in the absence of detailed studies of cognitive processes, other forms of research on academically and mathematically gifted children and adolescents—as defined by performance on achievement and aptitude tests—reveal that acceleration, alone or in combination with curriculum differentiation, is a best practice for serving the academic needs of these students (Colangelo, Assouline, & Gross, 2004; Southern, Jones, & Stanley, 1993). It is an educational option that is most strongly supported by research (Benbow, 1991; Benbow & Stanley, 1996; Colangelo et al.; Kulik & Kulik, 1984). The underlying principle for educating gifted youth is "appropriate developmental placement," or providing students with educational opportunities tailored to their rates of learning and level of competence (Benbow & Stanley; Colangelo et al.). In the words of Stanley (2000), the idea is to teach students "only what they don't already know" (p. 216). Although multiple studies have been conducted on a variety of accelerative options, the Task Group can summarize the results easily: When differences are found, they favor accelerated programs over traditional instruction, regardless of the mode of acceleration (e.g., Swiatek & Benbow, 1991a, 1991b; The benefits of accelerated instruction remain evident, even 50 years later (Cronbach, 1996). Moreover, students who receive accelerated instruction in math are more likely to be pursuing science, technology, engineering, and math (STEM) careers in their mid-30s (Lubinski, Benbow, Shea, Eftekhari-Sanjani, & Halvorson, 2001; Swiatek & Benbow, 1991a, 1991b). In addition, most students express satisfaction with their acceleration in both the short term and long term (Richardson & Benbow, 1990; Swiatek & Benbow, 1992).

5. Conclusions and Recommendations

Research efforts are needed in areas that assess the effectiveness of interventions designed to: 1) reduce the vulnerability of black and Hispanic students to negative stereotypes about their academic abilities, 2) functionally improve working memory capacity, and 3) provide explicit instruction on how to use strategies for effective and efficient problem solving.

More experimental work is needed to specify the underlying processes that link task engagement and self-efficacy, and the mathematics outcomes for black and Hispanic students. Urgently needed are a scaling-up and experimental evaluation of the interventions that have been found to be effective in enhancing engagement and self-efficacy for black and Hispanic students.

Intervention studies of students with MLD are in the early stages and should be a focus of future research efforts. Further research also is needed to identify the sources of MLD and learning difficulties in the areas of fractions, geometry, and algebra.

F. Brain Sciences and Mathematics Learning

Brain sciences research has the potential to contribute to knowledge of mathematical learning and eventually educational practices. Nevertheless, attempts to connect research in the brain sciences to classroom teaching and student learning in mathematics are premature. Instructional programs in mathematics that claim to be based on brain sciences research remain to be validated.

Although it is sometimes suggested that brain research should provide the scientific foundation for children's education in mathematics and in other academic areas, it is too early to directly apply findings from studies of brain processes during mathematical reasoning to classroom teaching and learning. Yet promising research emerging from the field of cognitive neuroscience is permitting investigators to begin forging links between neurobiological functions and mathematical cognition.

Most research making use of brain imaging and related techniques has focused on basic mental representations of number and quantity (Chochon, Cohen, van de Moortele, & Dehaene, 1999; Dehaene et al., 1999; Göbel, Calabria, Farné, & Rossetti, 2006; Halgren, Boujon, Clarke, Wang, & Chauvel, 2002; Kadosh et al., 2005; Pinel, Piazza, Le Bihan, & Dehaene, 2004; Temple & Posner, 1998; Vuilleumier, Ortigue, & Brugger, 2004; Zorzi et al., 2002), with a few studies exploring problem solving in arithmetic (Gruber, Indefrey, Steinmetz, & Kleinschmidt, 2001; Rickard et al., 2000; Rivera et al., 2005) and simple algebra (Anderson, Qin, Sohn, Stenger, & Carter, 2003; Qin et al., 2003; Qin et al., 2004). In most of these studies, researchers have contrasted the brain regions activated when children (or adolescents) and adults solve the same arithmetic or algebra problems (Kawashima et al., 2004; Qin et al., 2003; Qin et al., 2004; Rivera et al.); mapped changes in neural activity associated with practice at arithmetic (Delazer et al., 2003; Pauli et al., 1994); and differentiated the brain regions involved in arithmetic fact retrieval from those recruited for executing complex calculation procedures, such as regrouping in addition (Kong et al., 2005). In other studies, researchers have compared brain activity when the same quantities are presented in different notations (e.g., 8 versus eight; Kadosh, Kadosh, Kaas, Henik, & Goebel, 2007; Piazza, Pinel, Bihan, & Dehaene, 2007).

There is of course some variation across studies in the brain areas engaged when solving different types of mathematical problems—due to differences in experimental procedures and specific math problems presented across studies—but there are also intriguing consistencies. It has been repeatedly found that comparisons of number magnitudes (Pinel et al., 2004; Temple & Posner, 1998), quantitative estimation (Dehaene et al., 1999), use of a mental number line (Vuilleumier et al., 2004; Zorzi et al., 2002), and problem solving in arithmetic and algebra (Chochon et al., 1999; Qin et al., 2003; Rivera et al., 2005) activate several areas of the parietal cortex, including the bilateral intraparietal sulcus and angular gyrus. The intraparietal sulcus is also active when non-human animals engage in numerical activities (Sawamura, Shima, & Tanji, 2002; Thompson, Mayers, Robertson, & Patterson, 1970) and it has been proposed that a segment of this sulcus, particularly in the left hemisphere, may support an inherent number representational system (Dehaene et al., 2003). The evidence bearing on this last proposal, however, is mixed (Piazza et al., 2007; Shuman & Kanwisher, 2004; Simon & Rivera, 2007).

In all, researchers have used brain imaging and related methods to study the brain regions activated when children and adults solve arithmetic and simple algebra problems (Qin et al., 2003; Qin et al., 2004; Rivera et al., 2005), when the same individuals solve arithmetic problems at earlier and later points in learning (Delazer et al., 2003; Delazer et al., 2005), when individuals solve simple or more complex arithmetic problems (Dehaene et al., 1999; Kong et al., 2005), or when people solve arithmetic problems that involve different operations (Isheback et al., 2006). During the early phases of learning in childhood, numerical and arithmetical estimation and arithmetical problem solving generally engage the intraparietal sulcus of both hemispheres (Dehaene et al., 2003), as well as areas of the prefrontal cortex that support aspects of attentional control and working memory manipulations (Delazer et al., 2003; Menon, Rivera, White, Glover, & Reiss, 2000; Pauli et al., 1994). The execution of arithmetical procedures, such as regrouping in complex arithmetic, is also dependent on these prefrontal regions (Kong et al.). The evidence to date indicates that practice of simple (e.g., 2×5) and more complex (e.g., 23×5) arithmetic results in changes in recruitment of the brain regions supporting these competencies; that is, on easier problems, there is a decreased involvement of the prefrontal and perhaps intraparietal regions and increased engagement of the angular gyrus, especially in the left hemisphere (Delazer et al., 2003; Pauli et al.; Rivera et al.; but see Rickard et al., 2000). There is not a sufficient number of studies with children of various ages and grades to draw strong conclusions about schooling and mathematical development, but the research that has been conducted thus far suggests a similar pattern, that is, decreased involvement of the prefrontal/working memory regions and increased involvement of the angular gyrus with increasing grade level and mathematical experience (Rivera et al.).

This summary is an incomplete picture of schooling- and practice-related changes in brain functioning during mathematical learning. For example, Rivera et al.'s (2005) study also implicates other brain regions—such as the hippocampus which supports the formation of declarative memories—involved in the learning of basic arithmetic facts; Qin et al.'s (2003, 2004) studies suggest the parietal cortex in the adolescent brain may be more responsive than the same regions in the adult brain when individuals are learning to solve simple algebraic equations; Sohn et al.'s (2004) study suggests differences in the brain regions that contribute to success at solving algebraic word problems and algebraic equations; and, Ischebeck et al.'s (2006) results suggest that there may be differences in the network of posterior brain regions engaged during the learning of different arithmetical operations. The progress to date indicates that when combined with insights provided by cognitive research, brain imaging and related methodologies can provide unique and essential information on how children and adults learn mathematics. In coming years, these technologies will almost certainly help answer core questions associated with mathematical learning, such as the sources of learning disabilities and the effects of different forms of instruction on the acquisition of declarative, conceptual, and procedural competencies.

1. Conclusions and Recommendations

Brain sciences research has a unique potential for contributing to knowledge of mathematical learning and cognition, and eventually educational practices. Nevertheless, attempts to connect research in the brain sciences to classroom teaching and student learning in mathematics should not be made until instructional programs in mathematics based on brain sciences research are created and validated.

BIBLIOGRAPHY

Acredolo, C., O'Connor, J., Banks, L., & Horobin, K. (1989). Children's ability to make probability estimates: Skills revealed through application of Anderson's functional measurement methodology. *Child Development, 60,* 933–946.

Ahl, V., Moore, C.F., & Dixon, J.A. (1992). Development of intuitive and numerical proportional reasoning. *Cognitive Development, 7,* 81–108.

Alibali, M. W., & Goldin-Meadow, S. (1993). Gesture-speech mismatch and mechanisms of learning: What the hands reveal about a child's state of mind. *Cognitive Psychology, 25,* 468–523.

American Medical Association, Ad Hoc Committee on Health Literacy for the Council on Scientific Affairs. (1999). Health literacy: Report of the Council on Scientific Affairs. *Journal of the American Medical Association, 281*, 552–557.

Ames, C. (1990). Motivation: What teachers need to know. *Teachers College Record, 91,* 409–421.

Ames, C. (1992). Classrooms: goals, structures, and student motivation. *Journal of Educational Psychology, 84*, 261–71.

Ames, C., & Archer, J. (1988). Achievement goals in the classroom: Students' learning strategies and motivation processes. *Journal of Educational Psychology, 80,* 260–267.

Anderson, J.R., Reder, L.M., & Lebiere, C. (1996). Working memory: Activation limitations on retrieval. *Cognitive Psychology, 30*, 221–256.

Anderson, J.R., Qin, Y., Sohn, M.H., Stenger, V.A., & Carter, C.S. (2003). An information-processing model of the BOLD response in symbol manipulation tasks. *Psychonomic Bulletin & Review, 10,* 241–261.

Antell, S.E., & Keating, D.P. (1983). Perception of numerical invariance in neonates. *Child Development, 54*, 695–701.

Ashcraft, M.H. (1982). The development of mental arithmetic: A chronometric approach. *Developmental Review, 2,* 213–236.

Ashcraft, M.H. (2002). Math anxiety: Personal, educational, and cognitive consequences. *Current Directions in Psychological Science, 11*, 181–185.

Ashcraft, M.H., & Kirk, E.P. (2001). The relationships among working memory, math anxiety, and performance. *Journal of Experimental Psychology: General, 130,* 224–237.

Ashcraft, M.H., & Krause, J.A. (2007). Working memory, math performance, and math anxiety. *Psychonomic Bulletin & Review, 14,* 243–248.

Ashcraft, M.H., Krause, J.A., & Hopko, D.R. (2007). Is math anxiety a mathematical learning disability? In D.B. Berch, & M.M.M. Mazzocco (Eds.), *Why is math so hard for some children? The nature and origins of mathematical learning difficulties and disabilities* (pp. 329–348). Baltimore: Paul H. Brookes.

Ayres, P.L. (2001). Systematic mathematical errors and cognitive load. *Contemporary Educational Psychology, 26,* 227–248.

Ayres, P. (2006). Using subjective measures to detect variations of intrinsic cognitive load within problems. *Learning and Instruction, 16*, 389–400.

Baddeley, A.D. (1986). *Working memory*. Oxford: Oxford University Press.

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4,* 417–423.

Badian, N.A. (1983). Dyscalculia and nonverbal disorders of learning. In H. R. Myklebust (Ed.), *Progress in learning disabilities* (Vol. 5, pp. 235–264). New York: Grune and Stratton.

Bahrick, H.P., & Hall, L.K. (1991). Lifetime maintenance of high school mathematics content. *Journal of Experimental Psychology: General, 120,* 22–33.

Baker, D.W. (2006). The meaning and the measure of health literacy. *Journal of General Internal Medicine, 21*(8), 878−883.

Balfanz, R., & Byrnes, V. (2006). Closing the mathematics achievement gap in high-poverty middle schools: Enablers and constraints. *Journal of Education for Students Placed At Risk, 11*, 143–159.

Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist, 28*, 117–148.

Bandura, A. (1997). *Self-Efficacy: The Exercise of Control.* New York: Worth Publishers.

Barbaresi, W.J., Katusic, S.K., Colligan, R.C., Weaver, A.L., & Jacobsen, S.J. (2005). Math learning disorder: Incidence in a population-based birth cohort, 1976–82, Rochester, Minn. *Ambulatory Pediatrics, 5*, 281–289.

Barbey, A.K., & Sloman, S.A. (2007). Base-rate respect: From ecological rationality to dual processes. *Behavioral and Brain Sciences, 30*(3), 241–254.

Baroody, A.J. (1999). The roles of estimation and the commutativity principle in the development of third graders' mental multiplication. *Journal of Experimental Child Psychology, 74*, 157–193.

Baroody, A.J., Feil, Y., & Johnson, A.R. (2007). An alternative reconceptualization of procedural and conceptual knowledge. *Journal for Research in Mathematics Education*, *38*, 115–131.

Baroody, A.J., & Gannon, K.E. (1984). The development of the commutativity principle and economical addition strategies. *Cognition and Instruction, 1*, 321–339.

Baroody, A.J., & Ginsburg, H.P. (1983). The effects of instruction on children's understanding of the "equals" sign. *Elementary School Journal*, 84, 199–212.

Baroody, A.J., Ginsburg, H.P., & Waxman, B. (1983). Children's use of mathematical structure. *Journal for Research in Mathematics Education, 14*, 156–168.

Baroody, A.J., & Lai, M. (2007). Preschoolers' understanding of the addition–subtraction inverse principle: A Taiwanese sample. *Mathematical Thinking and Learning, 9*, 131–171.

Baroody, A.J., Wilkins, J.L.M., & Tiilikainen, S.H. (2003). The development of children's understanding of additive commutativity: From protoquantitative concept to general concept? In A.J. Baroody & A. Dowker (Eds.), *The development of arithmetic concepts and skills: Constructing adaptive expertise* (pp. 127–160). Mahwah, NJ: Lawrence Erlbaum Associates.

Barron, K.E., & Harackiewicz, J.M. (2001). Achievement goals and optimal motivation: Test multiple goals models. *Journal of Personality and Social Psychology, 80,* 706–722.

Barrouillet, P., Fayol, M., & Lathuliére, E. (1997). Selecting between competitors in multiplication tasks: An explanation of the errors produced by adolescents with learning disabilities. *International Journal of Behavioral Development, 21*, 253–275.

Baumeister, R.F. (2005). The cultural animal: Human nature, meaning, and social life. New York: Oxford University Press.

Bassok, M. (1990). Transfer of domain-specific problem-solving procedures. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 522–533.

Bassok, M., Chase, V.M., & Martin, S.A. (1998). Adding apples and oranges: Alignment of semantic and formal knowledge. *Cognitive Psychology, 35,* 99–134.

Bassok, M., & Holyoak, K. (1989). Interdomain transfer between isomorphic topics in algebra and physics. *Journal of Experimental Psychology: Learning,* 153–166.

Battista, M.T. (2007). The development of geometric and spatial thinking. In F.K. Lester, Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (Vol. 2, pp. 843–908). Charlotte, NC: Information Age Publishing.

Behr, M.J., Harel, G., Post, T., & Lesh, R. (1992). Rational number, ratio, and proportion. In D. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 296–333). New York: Macmillan Publishing.

Beilock, S.L., Kulp, C.A., Holt, L.E., & Carr, T.H. (2004). More on the fragility of performance: Choking under pressure in mathematical problem solving. *Journal of Experimental Psychology: General, 133,* 584–600.

Beilock, S.L., Rydell, R.J., & McConnell, A.R. (2007). Stereotype threat and working memory: Mechanisms, alleviation, and spillover. *Journal of Experimental Psychology: General, 136*, 256–276.

Benbow, C.P. (1991). Meeting the needs of gifted students through use of acceleration: A neglected resource. In M.C. Wang, M.C. Reynolds, & H.J. Walberg (Eds.), *Handbook of special education research and practice* (Vol. 4, pp. 23–36). Elmsford, NY: Pergamon Press.

Benbow, C.P., & Lubinski, D. (Eds.). (1996). *Intellectual talent: Psychometric and social issues*. Baltimore, MD: The Johns Hopkins University Press.

Benbow, C.P., & Stanley, J.C. (1983). Sex differences in mathematical reasoning ability: More facts. *Science, 222*, 1029–1031.

Benbow, C.P., & Stanley, J.C. (1996). Inequity in equity: How “equity” can lead to inequity for high-potential students. *Psychology, Public Policy, and Law,* 2, 249–292.

Benjamin, A.S., Bjork, R.A., & Schwartz, B.L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General*, *127*, 55–68.

Bennett, R.E., Gottesman, R.L., Rock, D.A., & Cerullo, F. (1993). Influence of behavior perception and gender on teachers’ judgments of students’ academic skill. *Journal of Educational Psychology, 85*, 347–356.

Ben-Zeev, T. (1995). The nature and origin of rational errors in arithmetic thinking: Induction from examples and prior knowledge. *Cognitive Science, 19,* 341–376.

Ben-Zeev, T., Fein, S., & Inzlicht, M. (2005). Arousal and stereotype threat. *Journal of Experimental Social Psychology, 41*, 174–181.

Berch, D.B. (2005). Making sense of number sense: Implications for children with mathematical disabilities. *Journal of Learning Disabilities, 38* 333–339.

Berkman, N.D., DeWalt, D.A., Pignone, M.P., et al. (2004). Literacy and health outcomes. *Evidence Report/Technology Assessment No. 87*. Rockville, MD: Agency for Healthcare Research and Quality.

Berlyne, D. (1960). *Conflict, arousal and curiosity*. New York: McGraw-Hill.

Birenbaum, M., Kelly, A.E., & Tatsuoka, K.K. (1993). Diagnosing knowledge states in algebra using the rule-space model. *Journal for Research in Mathematics Education, 24*, 442–459.

Birenbaum, M., Kelly, A.E., Tatsuoka, K.K., & Gutvirtz, Y. (1994). Attribute-mastery patterns from rule space as the basis for student models in algebra. *International Journal of Human–Computer Studies, 40*, 497–508.

Birenbaum, M., & Tatsuoka, K.K. (1993). Applying an IRT-Based cognitive diagnostic model to diagnose students' knowledge states in multiplication and division with exponents. *Applied Measurement in Education, 6*, 255–268.

Bjork, R.A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe and A. Shimamura (Eds.), *Metacognition: Knowing about knowing*. (pp.185–205). Cambridge, MA: MIT Press.

Black, W.C., Nease, R.F., & Tosteson, A.N. (1995). Perceptions of breast cancer risk and screening effectiveness in women younger than 50 years of age. *Journal of the National Cancer Institute, 87*, 720–731.

Blackwell, L., Trzesniewski, K., & Dweck, C.S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development*, 78, 246–263.

Blessing, S.B., & Ross, B.H. (1996). Content effects in problem categorization and problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 792–810.

Block, J.H. (1971). *Mastery learning: Theory and Practice*. New York: Holt, Rinehart and Winston.

Bloom, B. (1981). *All our children learning*. New York: McGraw-Hill.

Blöte, A.W., Van der Burg, E., & Klein, A.S. (2001). Students' flexibility in solving two-digit addition and subtraction problems: Instructional effects. *Journal of Educational Psychology, 93*, 627–638.

Booth, J.L., & Siegler, R.S. (in press). Numerical magnitude representations influence arithmetic learning. *Child Development.*

Booth, J.L., & Siegler, R.S. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology in Mathematics, 42*, 189–201.

Borman, G., & Overman, L. (2004). Academic resilience among poor and minority students. *Elementary School Journal, 104*, 177–195.

Bouwmeester, S., Vermunt, J.K., & Sijtsma, K. (2007). Development and individual differences in transitive reasoning: A fuzzy trace theory approach. *Developmental Review, 27*, 41–74.

Boykin, A.W. (1986). The triple quandary and the schooling of Afro-American children. In V. Nelssen (Ed.), *The school achievement of minority children.* Hillsdale, N.J.: Lawrence Erlbaum, 57–92.

Boykin, A.W., & Allen, B. (2004). Cultural integrity and schooling outcomes of African American children from low-income backgrounds. In P. Pufall and R. Unsworth (Eds.), *Rethinking childhood*. (pp. 104–120). New Brunswick: Rutgers University Press.

Boykin, A.W., & Ellison, C. (1995). The multiple ecologies of Black youth socialization: An afrographic analysis. In R. Taylor (Ed.), *African American youth: Their social and economic status*. Westport, CT: Praeger Press, 93–128.

Brainerd, C.J., & Gordon, L.L. (1994). Development of verbatim and gist memory for numbers. *Developmental Psychology, 30,* 163–177.

Brainerd, C.J., & Reyna, V.F. (1990). Inclusion illusions: Fuzzy-trace theory and perceptual salience effects in cognitive development. *Developmental Review, 10*, 365–403.

Brainerd. C.J., & Reyna, V.F. (1993). Memory independence and memory interference in cognitive development. *Psychological Review, 100*, 42–67.

Brainerd, C.J., & Reyna, V.F. (1995). Autosuggestibility in memory development. C*ognitive Psychology, 28*, 65–101.

Brand, B., Glasson, G., & Green, A. (2006). Sociocultural factors influencing students' learning in science and mathematics: An analysis of the perspectives of African American students. *School Science and Mathematics, 106,* 228–240.

Brannon, E.M. (2002). The development of ordinal numerical knowledge in infancy. *Cognition*, 83, 223–240.

Brase, G.L. (2002). "Bugs" built into the system: An evolutionary explanation for developmental difficulties in learning about fractions. *Learning and Individual Differences, 12*, 391–409.

Briars, D., & Siegler, R.S. (1984). A featural analysis of preschoolers' counting knowledge. *Developmental Psychology, 20*, 607–618.

Brophy, J. (2005). Goal theorists should move on from performance goals. *Educational Psychologist, 40*, 167–176.

Bruer, J.T. (2002). Avoiding the pediatrician's error: How neuroscientists can help educators (and themselves). *Nature Neuroscience, 5,* 1031–1033.

Bryant, P., Christie, C., & Rendu, A. (1999). Children's understanding of the relation between addition and subtraction: Inversion, identity, and decomposition. *Journal of Experimental Child Psychology, 74,* 194–212.

Butterworth, B., & Reigosa, V. (2007). Information processing deficits in dyscalculia. In D. B. Berch & M. M. M. Mazzocco (Eds.), *Why is math so hard for some children? The nature and origins of mathematical learning difficulties and disabilities* (pp. 65–81). Baltimore, MD: Paul H. Brookes Publishing Co.

Byrnes, J. (2003). Factors predictive of mathematic achievement in White, Black, & Hispanic 12th-graders. *Journal of Educational Psychology, 95,* 316–326.

Byrnes, J., & Wasik, B. (1991). Role of conceptual knowledge in mathematical procedural learning. *Developmental Psychology, 27*, 777–787.

Cadinu, M., Maass, A., Rosabianca, A., & Kiesner, J. (2005). Why do women underperform under stereotype threat? Evidence for the role of negative thinking. *Psychological Science*, *16*, 572–578.

Cahalan, M.W., Ingels, S.J., Burns, L.J., Planty, M., & Daniel, B. (2006). *United States high school sophomores: A twenty-two year comparison, 1980–2002.* Retrieved from http://nces.ed.gov/pubs2006/2006327.pdf.

Cai, J. (2004). Why do U.S. and Chinese students think differently in mathematical problem solving? Impact of early algebra learning and teachers' beliefs. *Journal of Mathematical Behavior, 23*, 135–167.

California Department of Education. (2006). *Mathematics Framework for California Public Schools: Kindergarten through grade twelve*. California Department of Education.

Campbell, J.I.D. (1999). Division by multiplication. *Memory & Cognition*, *27*, 791–802.

Campbell, J.I.D., & Graham, D.J. (1985). Mental multiplication skill: Structure, process, and acquisition. *Canadian Journal of Psychology, 3*9, 338–366.

Campbell, J.I.D., & Xue, Q. (2001). Cognitive arithmetic across cultures. *Journal of Experimental Psychology: General, 130,* 299–315.

Canobi, K.H., Reeve, R.A., & Pattison, P.E. (1998). The role of conceptual understanding in children's addition problem solving. *Developmental Psychology, 34,* 882–891.

Canobi, K.H., Reeve, R.A., & Pattison, P.E. (2002). Young children's understanding of addition concepts. *Educational Psychology, 22,* 513–532.

Capraro, M.M., & Joffrion, H. (2006). Algebraic equations: Can middle-school students meaningfully translate from words into mathematical symbols? *Reading Psychology, 27*, 147–164.

Carpenter, T.P., Corbitt, M.K., Kepner, H.S., Lindquist, M.M., & Reys, R.E. (1981). *Results from the second mathematics assessment of the National Assessment of Educational Progress*. Washington, DC: National Council of Teachers of Mathematics.

Carpenter, T.P., Coburn, T.G., Reys, R.E., Wilson, J.W., & Corbitt, M.K. (1978). *Results of the First Mathematics Assessment of the National Assessment of Educational Progress.* Reston, VA: National Council of Teachers of Mathematics.

Carpenter, T.P., & Moser, J.M. (1983). The acquisition of addition and subtraction concepts. In R. Lesh & M. Landau (Eds.), *Acquisition of mathematical concepts and processes* (pp. 7–44). New York: Academic Press.

Carpenter, T.P., & Moser, J.M. (1984). The acquisition of addition and subtraction concepts in grades one through three. *Journal for Research in Mathematics Education, 15*, 179–202.

Carraher, D.W., & Schliemann, A.D. (2007). Early algebra and algebraic reasoning. In F.K. Lester, Jr. (Ed.), *Second handbook of research on mathematics teaching and learning: A project of the National Council of Teachers of Mathematics* (Vol. 2, pp. 669–705). Charlotte, NC: Information Age Publishing.

Carroll, W.M. (1994). Using worked examples as an instructional support in the algebra classroom. *Journal of Educational Psychology, 86*, 360–367.

Cary, M., & Carlson, R.A. (1999). External support and the development of problem solving routines. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 1053–1070.

Case, R., & Okamoto, Y. (1996). The role of central conceptual structures in the development of children's thought. *Monographs of the Society for Research in Child Development*, 61 (Serial No. 246).

Case, R. & Sowder, J.T. (1990). The development of computational estimation: A neo-Piagetian analysis. *Cognition and Instruction, 7,* 79–104.

Casey, M.B., Nuttall, R.L., Pezaris, E., & Benbow, C.P. (1995). The influence of spatial ability on gender differences in mathematics college entrance test scores across diverse samples. *Developmental Psychology, 31*, 697–705.

Casey, M.B., Nuttall, R.L., & Pezaris, E. (1997). Mediators of gender differences in mathematics college entrance test scores: A comparison of spatial skills with internalized beliefs and anxieties. *Developmental Psychology, 33*, 669–680.

Casteel, C. (1997). Attitudes of African American and Caucasian eighth grade students about praises, rewards and punishments. *Elementary School Guidance and Counseling, 31,* 262–72.

Ceci, S.J., & Williams, W. (Eds.). (2006). *Why aren't more women in science? Top researchers debate the evidence.* Washington, DC: American Psychological Association.

Cepeda, N.J., Pashler, H., Vul, E., Wixted, J., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, *132*, 354–380.

Charness, N., & Campbell, J.I.D. (1988). Acquiring skill at mental calculation in adulthood: A task decomposition. *Journal of Experimental Psychology: General*, *117*, 115–129.

Chen, X., Rubin, K.H., & Li, D. (1997). Relation between academic achievement and social adjustment: Evidence from Chinese children. *Developmental Psychology, 33*, 518–525.

Chochon, F., Cohen, L., van de Moortele, P.F., & Dehaene, S. (1999). Differential contributions of the left and right inferior parietal lobules to number processing. *Journal of Cognitive Neuroscience, 11*, 617–630.

Clement, J. (1982). Algebra word problem solutions: Thought processes underlying a common misconception. *Journal for Research in Mathematics Education, 13,* 16–30.

Clements, D.H. (2002). Linking research and curriculum development. In L.D. English (Ed.), *Handbook of international research in mathematics education* (pp. 599–630). Mahwah, NJ: Lawrence Erlbaum Associates.

Clements, D.H. (2004). Major themes and recommendations. In D.H. Clements & J. Sarama (Eds.), & A.M. DiBiase (Assoc. Ed.), *Engaging young children in mathematics: Standards for early childhood mathematics education* (pp. 7–72). Mahwah, NJ: Lawrence Erlbaum Associates.

Clements, D.H., & Battista, M.T. (1992). Geometry and spatial reasoning. In D.A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 420–464). New York: Macmillan.

Clements, D.H., & Sarama, J. (2007a). Early childhood mathematics learning. In F. K. Lester, Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (Vol. 1, pp. 461–555). Charlotte, NC: Information Age Publishing.

Clements, D.H., & Sarama, J. (2007b). Effects of a preschool mathematics curriculum: Summative research on the Building Blocks Project. *Journal for Research in Mathematics Education, 38,* 136–163.

Clements, D.H., Swaminathan, S., Hannibal, M.A.Z., & Sarama, J. (1999). Young children's concepts of shape. *Journal for Research in Mathematics Education, 30, 2,* 192–212.

Cohen, G.L., Garcia, J., Apfel, N., & Master, A. (2006). Reducing the racial achievement gap: A social-psychological Intervention. Retrieved on May 10, 2007 from www.sciencemag.org.

Cohen, L.B., & Marks, K.S. (2002). How infants process addition and subtraction events. *Developmental Science, 5,* 186–201.

Colangelo, N., Assouline, S.G., & Gross, M.U.M. (Eds.). (2004). *A nation deceived: How schools hold back America's brightest students*. Iowa City: University of Iowa.

Coleman, S. (2003). Mathematics classroom performance as a function of communal versus individualistic learning contexts. Unpublished Doctoral Dissertation. Howard University, Washington, DC.

Cooney, J.B., & Swanson, H.L. (1990). Individual differences in memory for mathematical story problems: Memory span and problem perception. *Journal of Educational Psychology, 82,* 570–577.

Cooper, G., & Sweller, J. (1987). Effects of schema acquisition and rule automation on mathematical problem-solving transfer. *Journal of Educational Psychology, 79,* 347–362.

Correa, J., Nunes, T., & Bryant, P. (1998). Young children's understanding of division: The relationship between division terms in a noncomputational task. *Journal of Educational Psychology, 90,* 321–329.

Cowan, N., Saults, J.S., & Elliott, E.M. (2002). The search for what is fundamental in the development of working memory. *Advances in Child Development and Behavior, 29,* 1–49.

Cowan, N. (1995). *Attention and memory: An integrated framework*. New York: Oxford University Press.

Cronbach, L. J. (1996). Acceleration among the Terman males: Correlates in midlife and after. In C.P. Benbow & D. Lubinski (Eds.), *Intellectual talent: Psychometric and social issues* (pp. 179–191). Baltimore: Johns Hopkins University Press.

Dark, V.J., & Benbow, C.P. (1990). Enhanced problem translation and short-term memory: Components of mathematical talent. *Journal of Educational Psychology, 82,* 420–429.

Dark, V.J., & Benbow, C.P. (1991). Differential enhancement of working memory with mathematical versus verbal precocity. *Journal of Educational Psychology, 83,* 48–60.

Darke, I. (1982). A review of research related to the topological primacy thesis. *Educational Studies in Mathematics, 13,* 119–142.

De Bock, D., Van Dooren, W., Janssens, D., & Verschaffel, L. (2002). Improper use of linear reasoning: An in-depth study of the nature and the irresistibility of secondary school students' errors. *Educational Studies in Mathematics, 50*, 311–334.

De Bock, D., Verschaffel, L., & Janssens, D. (1998). The predominance of the linear model in secondary school students' solutions of word problems involving length and area of similar plane figures. *Educational Studies in Mathematics, 35*, 65–83.

De Bock, D., Verschaffel, L., & Janssens, D. (2002). The effects of different problem presentations and formulations on the illusion of linearity in secondary school students. *Mathematical Thinking and Learning, 4*, 65–89.

De Bock, D., Verschaffel, L., Janssens, D., Van Dooren, W., & Claes, K. (2003). Do realistic contexts and graphical representations always have a beneficial impact on students' performance? Negative evidence from a study on modeling non-linear geometry problems. *Learning and Instruction, 13*, 441–463.

Dehaene, S. (1997). *The number sense: How the mind creates mathematics.* NY: Oxford University Press.

Dehaene, S., Izard, V., Pica, P., & Spelke, E. (2006). Core knowledge of geometry in an Amazonian indigene group. *Science, 311*, 381–384.

Dehaene, S., Piazza, M., Pinel, P., & Cohen, L. (2003). Three parietal circuits for number processing. *Cognitive Neuropsychology, 20*, 487–506.

Dehaene, S., Spelke, E., Pinel, P., Stanescu, R., & Tsivkin, S. (1999, May 7). Sources of mathematical thinking: Behavioral and brain-imaging evidence. *Science, 284*, 970–974.

Delaney, P.F., Reder, L.M., Staszewski, J.J., & Ritter, F.E. (1998). The strategy-specific nature of improvement: The power law applies by strategy within task. *Psychological Science, 9*, 1–7.

Delazer, M., Domahs, F., Bartha, L., Brenneis, C., Lochy, A., Trieb, T., et al. (2003). Learning complex arithmetic–An fMRI study. *Cognitive Brain Research, 18*, 76–88.

Delazer, M., Ischeback, A., Domahs, F., Zamarian, L., Koppelstätter, F., Siedentopf, C.M., et al. (2005). Learning by strategies and learning by drill: Evidence from an fMRI study. *NeuroImage, 25*, 838–849.

Dembo, Y., Levin, I., & Siegler, R.S. (1997). A comparison of the geometric reasoning of students attending Israeli ultraorthodox and mainstream schools. *Developmental Psychology, 33*, 92–103.

Dempster, F.N. (1992). The rise and fall of the inhibitory mechanism: Toward a unified theory of cognitive development and aging. *Developmental Review, 12*, 45–75.

DeStefano, D., & LeFevre, J.A. (2004). The role of working memory in mental arithmetic. *European Journal of Cognitive Psychology, 16*, 353–386.

Dixon, J.A., & Moore, C.F. (1996). The developmental role of intuitive principles in choosing mathematical strategies. *Developmental Psychology, 32*, 241–253.

Dowker, A. (1997). Young children's addition estimates. *Mathematical Cognition, 3*, 140–153.

Dowker, A. (2003). Younger children's estimates for addition: The zone of partial knowledge and understanding. In A.J. Baroody & A. Dowker (Eds.), *The development of arithmetic concepts and skills: Constructing adaptive expertise* (pp. 243–265). Mahwah, NJ: Erlbaum.

Dowker, A., Flood, A., Griffiths, H., Harriss, L., & Hook, L. (1996). Estimation strategies of four groups. *Mathematical Cognition, 2*, 113–135.

Duncan, G.J., Dowsett, C.J., Clasessens, A., Magnuson, K., Huston, A.C., Klebanov, P., et al. (2007). School readiness and later achievement. *Developmental Psychology, 43*, 1428–1446.

Duschl, R.A., Schweingruber, H.A., & Shouse, A.W. (Eds.). (2007). *Taking science to school: Learning and teaching sciences in grades K–8*. Washington, DC: National Academies Press.

Dweck, C.S. (1975). The role of expectations and attributions in the alleviation of learned helplessness. *Journal of Personality and Social Psychology, 31*, 674–685.

Dweck, C.S. (1999). *Self-Theories: Their role in motivation, personality, and development*. Philadelphia, PA: Psychology Press.

Dweck, C.S., & Leggett, E.L. (1988). A social–cognitive approach to motivation and personality. *Psychological Review, 95*, 256–273.

Eccles, J.S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology, 53*, 109–32.

Eccles, J.S., Wigfield, A., & Schiefele, U. (1998). Motivation to succeed. In Damon, W. & Eisenberg, W. *Social, emotional, and personality development, Volume* 3, *Handbook of child psychology (5th Edition)* (pp. 1017–1095). New York: John Wiley & Sons.

Elliot, A.J., & Harackiewicz, J.M. (1996). Approach and avoidance achievement goals and intrinsic motivation: A mediational analysis. *Journal of personality and social psychology, 70*, 461–475.

Ellis, H.C., Varner, L.J., & Becker, A.S. (1993). Cognition and emotion: Theories, implications, and educational applications. In L.A. Penner, G.M. Batsche, H.M. Knoff, & D.L. Nelson (Eds.), *The challenge of mathematics and science education: Psychology's response* (pp. 83–111). Washington, DC: American Psychological Association.

Engle, R.W., Conway, A.R.A., Tuholski, S.W., & Shisler, R.J. (1995). A resource account of inhibition. *Psychological Science, 6*, 122–125.

English, L.D., & Halford, G.S. (1995). *Mathematics education: Models and processes*. Mahwah, NJ: Lawrence Erlbaum Associates.

Ericsson, K.A., Krampe, R.T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100*, 363–406.

Esteley, C., Villarreal, M., & Alagia, H. (2004). Extending linear models to non-linear contexts: An in-depth study about two university students' mathematical productions. In M. J. Høines & A. B. Fuglestad (Eds.), *Proceedings of the 28th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 2, pp. 343–350). Bergen, Norway.

Estrada, C.A., Martin-Hryniewicz, M., Peek, B.T., Collins, C., & Byrd, J.C. (2004). Literacy and numeracy skills and anticoagulation control. *American Journal of the Medical Sciences, 328*, 88−93.

Fagerlin, A., Zikmund-Fisher, B.J., & Ubel, P.A. (2005). How making a risk estimate can change the feel of that risk: Shifting attitudes toward breast cancer risk in a general public survey. *Patient Education and Counseling, 57,* 294–299.

Ferguson, R. (1998). Teachers' perceptions and expectations and the Black-White test score gap. In C. Jencks & M. Phillips (Eds.), *The Black White Test Score Gap* (pp. 273–317). Washington, DC: Brookings Institution Press.

Ferguson, R. (2003). Teachers' perceptions and expectations and the Black-White test score gap. *Urban Education, 38,* 460–507.

Fischbein, E. (1990). Intuition and information processing in mathematical activity. *International Journal of Educational Research, 14,* 31–50.

Frensch, P.A., & Geary, D.C. (1993). The effects of practice on component processes in complex mental addition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 433–456.

Freudenthal, H. (1983). *Didactical phenomenology of mathematical structures*. Dordrecht: Reidel.

Frye, D., Braisby, N., Lowe, J., Maroudas, C., & Nicholls J. (1989). Young children's understanding of counting and cardinality. *Child Development, 60*, 1158–1171.

Fuchs, L.S., Fuchs, D., Compton, D.L., Schatschneider, C., Powell, S.R., Seethaler, P.M., et al. (2006). The cognitive correlates of third-grade skill in arithmetic, algorithmic computation, and arithmetic word problems. *Journal of Educational Psychology, 98,* 29–43.

Fuchs, L.S., Fuchs, D., Karns, K., Hamlett, C., Katzaroff, M., & Dutka, S. (1997). Effects of task-focused goals on low-achieving students with and without learning disabilities. *American Educational Research Journal, 34*, 513–543.

Fuchs, L., Fuchs, D., Prentice, K., Burch, M., Hamlett, C., Owen, R., et al. (2003). Enhancing third-grade students' mathematical problem solving with self-regulated learning strategies. *Journal of Educational Psychology, 95*, 306–315.

Fuson, K.C. (1982). An analysis of the counting-on solution procedure in addition. In T.P. Carpenter, J.M. Moser, & T.A. Romberg (Eds.), *Addition and subtraction: A cognitive perspective* (pp. 67–81). Hillsdale, NJ: Lawrence Erlbaum Associates.

Fuson, K.C. (1988). *Children's counting and concepts of number*. New York: Springer-Verlag.

Fuson, K.C. (1990). Conceptual structures for multiunit numbers: Implications for learning and teaching multidigit addition, subtraction, and place value. *Cognition and Instruction, 7,* 343–403.

Fuson, K.C., & Briars, D.J. (1990). Using a base-ten blocks learning/teaching approach for first- and second-grade place-value and multidigit addition and subtraction. *Journal for Research in Mathematics Education, 21*, 180–206.

Fuson, K.C., & Kwon, Y. (1992a). Korean children's single-digit addition and subtraction: Numbers structured by ten. *Journal for Research in Mathematics Education, 23,* 148–165.

Fuson, K.C., & Kwon, Y. (1992b). Korean children's understanding of multidigit addition and subtraction. *Child Development, 63*, 491–506.

Fuson, K.C., Smith, S.T., & Lo Cicero, A.M. (1997). Supporting Latino first graders' ten-structured thinking in urban classrooms. *Journal of Research in Mathematics Education*, 28, 738–766.

Fuson, K.C., Stigler, J.W., & Bartsch, K. (1988). Grade placement of addition and subtraction topics in Japan, Mainland China, the Soviet Union, Taiwan, and the United States. *Journal for Research in Mathematics Education, 19*, 449–456.

Gagné, F., & St Père, F. (2002). When IQ is controlled, does motivation still predict achievement? *Intelligence, 30*, 71–100.

Gailliot, M.T., Plant, E.A., Butz, D.A., & Baumeister, R.F. (2007). Increasing self-regulatory strength can reduce the depleting effect of suppressing stereotypes. *Personality and Social Psychology Bulletin, 33, 281–294.*

Gallagher, A.M., & Kaufman, J.C. (Eds.). (2005). *Gender differences in mathematics: An integrative psychological approach*. Cambridge, UK: Cambridge University Press.

Gauvain, M. (1993). Spatial thinking and its development in sociocultural context. *Annals of Child Development, 9*, 67–102.

Geary, D.C. (1990). A componential analysis of an early learning deficit in mathematics. *Journal of Experimental Child Psychology, 49*, 363–383.

Geary, D.C. (1994). *Children's mathematical development: Research and practical implications*. Washington, DC: American Psychological Association.

Geary, D.C. (1996). The problem size effect in mental addition: Developmental and cross-national trends. *Mathematical Cognition, 2*, 63–93.

Geary, D.C. (2004). Mathematics and learning disabilities. *Journal of Learning Disabilities, 37*, 4–15.

Geary, D.C. (2006). Development of mathematical understanding. In D. Kuhl & R.S. Siegler (Vol. Eds.), *Cognition, perception, and language, Vol. 2* (pp. 777–810). W. Damon (Gen. Ed.), *Handbook of child psychology* (6th Ed.). New York: John Wiley & Sons.

Geary, D.C., Bow-Thomas, C.C., Fan, L., & Siegler, R.S. (1993). Even before formal instruction, Chinese children outperform American children in mental addition. *Cognitive Development, 8*, 517–529.

Geary, D.C., Bow-Thomas, C.C., Liu, F., & Siegler, R.S. (1996). Development of arithmetical competencies in Chinese and American children: Influence of age, language, and schooling. *Child Development, 67*, 2022–2044.

Geary, D.C., Bow-Thomas, C.C., & Yao, Y. (1992). Counting knowledge and skill in cognitive addition: A comparison of normal and mathematically disabled children. *Journal of Experimental Child Psychology, 54*, 372–391.

Geary, D.C., & Brown, S.C (1991). Cognitive addition: Strategy choice and speed-of-processing differences in gifted, normal, and mathematically disabled children. *Developmental Psychology, 27*, 398–406.

Geary, D.C., Frensch, P.A., & Wiley, J.G. (1993). Simple and complex mental subtraction: Strategy choice and speed-of-processing differences in younger and older adults. *Psychology and Aging, 8*, 242–256.

Geary, D.C., Hamson, C.O., Chen, G.P., Liu, F., Hoard, M.K., & Salthouse, T.A. (1997). Computational and reasoning abilities in arithmetic: Cross-generational change in China and the United States. *Psychonomic Bulletin & Review, 4*, 425–430.

Geary, D.C., Hamson, C.O., & Hoard, M.K. (2000). Numerical and arithmetical cognition: A longitudinal study of process and concept deficits in children with learning disability. *Journal of Experimental Child Psychology, 77*, 236–263.

Geary, D.C., Hoard, M.K., Byrd-Craven, J., & Desoto, M. (2004). Strategy choices in simple and complex addition: Contributions of working memory and counting knowledge for children with mathematical disability. *Journal of Experimental Child Psychology, 88*, 121–151.

Geary, D.C., Hoard, M.K., Byrd-Craven, J., Nugent, L., & Numtee, C. (2007). Cognitive mechanisms underlying achievement deficits in children with mathematical learning disability. *Child Development*, *78*, 1343–1359.

Geary, D.C., Hoard, M.K., Nugent, L., & Byrd Craven, J. (in press). Development of number line representations in children with mathematical learning disability. *Developmental Neuropsychology, 33*.

Geary, D.C., Salthouse, T.A., Chen, G.P., & Fan, L. (1996). Are East Asian versus American differences in arithmetical ability a recent phenomenon? *Developmental Psychology*, 32, 254–262.

Geary, D.C., Saults, S.J., Liu, F., & Hoard, M.K. (2000). Sex differences in spatial cognition, computational fluency, and arithmetical reasoning. *Journal of Experimental Child Psychology, 77*, 337–353.

Geary, D.C., & Widaman, K.F. (1992). Numerical cognition: On the convergence of componential and psychometric models. *Intelligence, 16*, 47–80.

Geary, D.C., Widaman, K.F., & Little, T.D. (1986). Cognitive addition and multiplication: Evidence for a single memory network. *Memory & Cognition, 14*, 478–487.

Geary, D.C., & Wiley, J.G. (1991). Cognitive addition: Strategy choice and speed-of-processing differences in young and elderly adults. *Psychology and Aging, 6*, 474–483.

Gelman, R. (1991). Epigenetic foundations of knowledge structures: Initial and transcendent constructions. In S. Carey & R. Gelman (Eds.), *The epigenesis of mind: Essays in biology and cognition* (pp. 293–322). Hillsdale, NJ: Lawrence Erlbaum Associates.

Gelman, R., & Gallistel, C.R. (1978). *The child's understanding of number*. Cambridge, MA: Harvard University Press.

Gelman, R., & Meck, E. (1983). Preschoolers' counting: Principles before skill. *Cognition, 13*, 343–359.

Gelman, R., & Williams, E.M. (1998). Enabling constraints for cognitive development and learning: Domain-specificity and epigenesis. In D. Kuhl & R.S. Siegler (Vol. Eds.), *Cognition, perception, and language, Vol. 2* (pp. 575–630). W. Damon (Gen. Ed.), *Handbook of child psychology* (Fifth Ed.). New York: John Wiley & Sons.

Gelman, S.A. (2003). *The essential child: Origins of essentialism in everyday thought*. New York: Oxford University Press.

Gersten, R., & Chard, D. (1999). Number sense: Rethinking arithmetic instruction for students with mathematical disabilities. *The Journal of Special Education, 33*(1), 18–28.

Gersten, R., & Chard, D.J. (2001). Number sense: Rethinking arithmetic instruction for students with mathematical disabilities. Retrieved December 1, 2006, from http://www.ldonline.org/article/5838.

Gilmore, C.K., & Bryant, P. (2006). Individual differences in children's understanding of inversion and arithmetic skill. *British Journal of Educational Psychology, 76*, 309–331.

Ginsburg, A., Cooke, G., Leinwand, S. Noell, J., & Pollock, E. (2005). *Reassessing U.S. international mathematics performance: New findings from the 2003 TIMSS and PISA.* Washington, DC: American Institutes for Research.

Ginsburg, H. (1977). *Children's arithmetic: The learning process*. New York: D. Van Nostrand.

Ginsburg-Block, M., & Fantuzzo, J. (1997). Reciprocal peer tutoring: An analysis of "teacher" and "student" interactions as a function of training and experience. *School Psychology Quarterly, 12*(2), 134–149.

Girotto, V., & Gonzalez, M. (2007). Extensional reasoning about chances. In W. Schaeken, A. Vandierendonck, W. Schroyens, & G. d'Ydewalle (Eds.), *The mental models theory of reasoning: Refinements and extensions* (pp. 151–166). Mahwah, NJ: Lawrence Erlbaum Associates.

Göbel, S.M., Calabria, M., Farné, A., & Rossetti, Y. (2006). Parietal rTMS distorts the mental number line: Simulating 'spatial' neglect in healthy subjects. *Neuropsychologia, 44*, 860–868.

Goldman, S.R., Mertz, D.L., & Pellegrino, J.W. (1989). Individual differences in extended practice functions and solution strategies for basic addition facts. *Journal of Educational Psychology, 81,* 481–496.

Gonzalez, N., Andrade, R., Civil, M., & Moll, L. (2001). Bridging funds of distributed knowledge: Creating zones of practices in mathematics. *Journal of Education for Students Placed At Risk, 6*, 115–132.

Gonzales, P., Guzman, J. C., Partelow, L., Pahlke, E., Jocelyn, L., Kastberg, D., et al. (2004). *Highlights from the Trends in International Mathematics and Science Study: TIMSS 2003.* Retrieved from http://nces.ed.gov/pubs2005/2005005.pdf

Good, C., Aronson, J., & Inzlicht, M. (2003). Improving adolescents' standardized test performance: An intervention to reduce the effects of stereotype threat. *Applied Developmental Psychology, 24,* 645–662.

Gopnik, A., Meltzoff, A., & Kuhl, P. (1999). *The scientist in the crib: What early learning tells us about the mind*. New York: William Monroe and Company, Inc.

Goswami, U. (1989). Relational complexity and the development of analogical reasoning. *Cognitive Development, 4*, 251–268.

Gottfried, A.E., Fleming, J.S., & Gottfried, A.W. (2001). Continuity of academic intrinsic motivation from childhood through late adolescence: A longitudinal study. *Journal of Educational Psychology, 93*, 3–13.

Graham, S., & Golan, S. (1991). Motivational influences on cognition: Task involvement, ego involvement, and depth of information processing. *Journal of Educational Psychology, 83*, 187–94.

Grant, H., & Dweck, C.S. (2003). Clarifying achievement goals and their impact. *Journal of Personality and Social Psychology, 85*, 541–553.

Griffin, S. (2004). Number worlds: A research-based mathematics program for young children. In D.H. Clements & J. Sarama (Eds.) & A.M. DiBiase (Assoc. Ed.), *Engaging young children in mathematics: Standards for early childhood mathematics education* (pp. 325–340). Mahwah, NJ: Lawrence Erlbaum Associates.

Griffin, S.A., Case, R., & Siegler, R.S. (1994). Rightstart: Providing the central conceptual prerequisites for first formal learning of arithmetic to students at risk for school failure. In K. McGilly (Ed.), *Classroom lessons: Integrating cognitive theory and classroom practice* (pp. 25–49). Cambridge, MA: Bradford Books/MIT Press.

Groen, G.J., & Parkman, J.M. (1972). A chronometric analysis of simple addition. *Psychological Review, 79*, 329–343.

Gross-Tsur, V., Manor, O., & Shalev, R.S. (1996). Developmental dyscalculia: Prevalence and demographic features. *Developmental Medicine and Child Neurology, 38*, 25–33.

Gruber, O., Indefrey, P., Steinmetz, H., & Kleinschmidt, A. (2001). Dissociating neural correlates of cognitive components in mental calculation. *Cerebral Cortex*, *11*, 350–359.

Gutman, L.M. (2006). How student and parent goal orientations and classroom goal structures influence the math achievement of African Americans during the high school transition. *Contemporary Educational Psychology, 31*(1), 44–63.

Halff, H.M. (1977). The role of opportunities for recall in learning to retrieve. *American Journal of Psychology, 90*, 383–406.

Halford, G.S. (1993). *Children's understanding: The development of mental models.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Halgren, E., Boujon, C., Clarke, J., Wang, C., & Chauvel, P. (2002). Rapid distributed fronto-parieto-occipital processing stages during working memory in humans. *Cerebral Cortex, 12*, 710–728.

Hall, C., Davis, N., Bolen, L., & Chea, R. (1999). Gender and racial differences in mathematical performance. *Journal of Social Psychology, 139*, 677–689.

Halpern, D., Benbow, C., Geary, D.C., Gur, R., Hyde, J., & Gernsbacher, M.A. (2007). The science of sex differences in science and mathematics. *Psychological Science in the Public Interest, 8*, 1, 1–52.

Hamann, M.S., & Ashcraft, M.H. (1986). Textbook presentations of the basic addition facts. *Cognition and Instruction, 3,* 173–192.

Hanich, L.B., Jordan, N.C., Kaplan, D., & Dick, J. (2001). Performance across different areas of mathematical cognition in children with learning difficulties. *Journal of Educational Psychology, 93*, 615–626.

Harackiewicz, J.M., Barron, K.E., Pintrich, P.R., Elliott, A.J., & Thrash, T.M. (2002). Revision of achievement goal theory: Necessary and illuminating. *Journal of Educational Psychology, 94*, 638–45.

Hatano, G. (1990). Toward the cultural psychology of mathematical cognition. Commentary on Stevenson, H.W., Lee, S.Y, Chen, C., Stigler, J. W., Hsu, C.C., & Kitamura, S. Contexts of achievement: A study of American, Chinese, and Japanese children. *Monographs of the Society for Research in Child Development, 55,* 1–2, Serial No. 221.

Hatano, G., Amaiwa, S., & Inagaki, K. (1996). "Buggy algorithms" as attractive variants. *Journal of Mathematical Behavior, 15*, 285–302.

Hauser-Cram, P., Sirin, S., & Stipek, D. (2003). When teachers' and parents' values differ: Teachers' ratings of academic competence in children from low-income families. *Journal of Educational Psychology*, 95, 813–820.

Hecht, S.A. (1998). Toward an information processing account of individual differences in fraction skills. *Journal of Educational Psychology, 90*, 545–559.

Hecht, S.A., Close, L., & Santisi, M. (2003). Sources of individual differences in fraction skills. *Journal of Experimental Child Psychology, 86*, 277–302.

Hecht, S.A., Vagi, K.J., & Torgesen, J.K. (2007). Fraction skills and proportional reasoning. In D.B. Berch, & M.M.M. Mazzocco (Eds.), *Why is math so hard for some children? The nature and origins of mathematical learning difficulties and disabilities* (pp. 121–132). Baltimore, MD: Paul H. Brookes Publishing Co.

Hedges, L.V., & Nowell, A. (1995). Sex differences in mental scores, variability, and numbers of high-scoring individuals. *Science*, *269*, 41–45.

Hiele, P.M. van (1986). *Structure and insight: A theory of mathematics education*. Orlando, FL: Academic Press.

Heller, P.M., Post, T.R., Behr, M., & Lesh, R. (1990). Qualitative and numerical reasoning about fractions and rates by seventh- and eighth-grade students. *Journal for Research in Mathematics Education, 21,* 388–402.

Hembree, R. (1990). The nature, effect, and relief of mathematics anxiety. *Journal for Research in Mathematics Education*, *21*, 33–46.

Hembree, R. (1992). Experiments and relational studies in problem solving: A meta-analysis. *Journal for Research in Mathematics Education, 23,* 242–273.

Hiebert, J., & Wearne, D. (1986). Procedures over concepts: The acquisition of decimal number knowledge. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: The case of mathematics* (pp. 199–223). Hillsdale, NJ: Lawrence Erlbaum Associates.

Hiebert, J., & Wearne, D. (1996). Instruction, understanding, and skill in multidigit addition and subtraction. *Cognition and Instruction, 14*, 251–283.

Hinsley, D.A., Hayes, J.R., & Simon, H.A. (1977). From words to equations: Meaning and representation in algebra word problems. In M.A. Just & P.A. Carpenter (Eds.), *Cognitive processes in comprehension* (pp. 89–106). New York: John Wiley & Sons.

Hitch, G.J. (1978). The role of short-term working memory in mental arithmetic. *Cognitive Psychology*, *10*, 302–323.

Hoemann, H.W., & Ross, B.M. (1982). Children's concepts of chance and probability. In C.J. Brainerd (Ed.), *Children's logical and mathematical cognition: Progress in cognitive development research*. New York: Springer-Verlag.

Hoffer, Thomas B., Venkataraman, L., Hedberg, E.C., & Shagle, S. (2007). *Final Report on the National Survey of Algebra Teachers for the National Math Panel*. Chicago: National Opinion Research Center at the University of Chicago.

Holyoak, K.J., & Thagard, P. (Eds.). (1995). *Mental leaps: Analogy in creative thought.* Boston: MIT Press.

Holyoak, K.J., & Thagard, P. (1997). The analogical mind. *American Psychologist, 52*, 35–44.

Hope, J., & Owens, O. (1987). An analysis of the difficulty of learning fractions. *Focus on Learning Problems in Mathematics, 9*, 25–39.

Hughes, S.A. (2003). An early gap in black-white mathematics achievement: Holding school and home accountable in an affluent city school district. *Urban Review***,** 35, 297–322.

Hunt, J.M.V. (1965). Intrinsic motivation and its role in psychological development. In D. Levine (Ed.), *Nebraska Symposium on Motivation* (Vol. 13, pp. 189–282). Lincoln: University of Nebraska Press.

Hunt, R.R., & Ellis, H.C. (2004). *Fundamentals of cognitive psychology (7th ed.)*. New York: McGraw-Hill.

Hunting, R.P., & Davis, G.E. (Eds.). (1991). *Early fraction learning*. New York: Springer-Verlag.

Huntsinger, C.S., Jose, P.E., Fong-Ruey, L., & Wei-Di, C. (1997). Cultural differences in early mathematics learning: A comparison of Euro-American, Chinese-American, and Taiwan-Chinese families. *International Journal of Behavioral Development, 21*, 371–388.

Hurley, E.A., Boykin, A.W., & Allen, B.A. (2005). Communal vs. individual learning of a math-estimation task: African American children and the culture of learning contexts. *Journal of Psychology: Interdisciplinary & Applied, 139*, 513–527.

Hutchinson, N.L. (1993). Effects of cognitive strategy instruction on algebraic problem solving of adolescents with learning disabilities. *Learning Disability Quarterly, 16*, 34–63.

Huttenlocher, J., Jordan, N.C., & Levine, S.C. (1994). A mental model for early arithmetic. *Journal of Experimental Psychology: General, 123*, 284–296.

Hutton, L., A., & Levitt, E. (1987). An academic approach to the remediation of mathematics anxiety. In R. Schwarzer, H.M. van der Ploeg, & C.D. Spielberger (Eds.), *Advances in test anxiety research* (Vol. 5, pp. 207–211). Berwyn, PA: Swets North America Inc.

Hyde, J.S., Fennema, E., & Lamon, S.J. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin, 107,* 139–155.

Ilg, F., & Ames, L.B. (1951). Developmental trends in arithmetic. *Journal of Genetic Psychology, 79*, 3–28.

Institute of Medicine. (2004). *Health literacy: A prescription to end confusion.* Washington, DC: National Academies Press.

Ischeback, A., Zamarian, L., Siedentopf, C., Koppelstätter, F., Benke, T., Felber, S., et al. (2006). How specifically do we learn? Imaging the learning of multiplication and subtraction. *NeuroImage, 30*, 1365–1375.

Ives, B. (2007). Graphic organizers applied to secondary algebra instruction for students with learning disabilities. *Learning Disabilities Research & Practice, 22*, 109–117.

Jacobs, J.E., & Potenza, M. (1991). The use of judgment heuristics to make social and object decisions: A developmental perspective. *Child Development, 62*, 166–178.

Jansen, A.R., Marriott, K., & Yelland, G.W. (2003). Comprehension of algebraic expressions by experienced users of mathematics. *Quarterly Journal of Experimental Psychology, 56A*, 3–30.

Jansen, A.R., Marriott, K., & Yelland, G.W. (2007). Parsing algebraic expressions by experienced users of mathematics. *European Journal of Experimental Psychology, 19, 286*–320.

Jencks, C., & Phillips, M. (1998). *The Black-White Test Score Gap.* Washington, DC: Brookings Institution Press.

Jeong, Y., Levine, S.C. & Huttenlocher, J. (2007). The development of proportional reasoning: Effect of continuous vs. discrete quantities. *Journal of Cognition and Development, 8*, 277–356.

Johanning, D.I. (2004). Supporting the development of algebraic thinking in middle school: A closer look at students' information strategies. *Journal of Mathematical Behavior, 23,* 371–388.

Johnson, E.S. (1984). Sex differences in problem solving. *Journal of Educational Psychology, 76*, 1359–1371.

Jones, K.K., & Byrnes, J.P. (2006). Characteristics of students who benefit from high-quality mathematics instruction. *Contemporary Educational Psychology, 31*, 328–343.

Jordan, N.C., Hanich, L.B., & Kaplan, D. (2003). Arithmetic fact mastery in young children: A longitudinal investigation. *Journal of Experimental Child Psychology, 85*, 103–119.

Jordan, N.C., Kaplan, D., Olah, L., & Locuniak, M.N. (2006). Number sense growth in kindergarten: A longitudinal investigation of children at risk for mathematics difficulties. *Child Development, 77*, 153–175.

Jordan, N.C., & Montani, T.O. (1997). Cognitive arithmetic and problem solving: A comparison of children with specific and general mathematics difficulties. *Journal of Learning Disabilities, 30*, 624–634.

Jussim, L., Eccles, J., & Madon, S. (1996) Social perception, social stereotypes, and teacher expectations: Accuracy and the quest for the powerful self fulfilling prophecy. *Advances in Experimental Social Psychology, 28,* 281–387.

Kadosh, R.C., Henik, A., Rubinsten, O., Mohr, H., Dori, H., van de Van, V., et al. (2005). Are numbers special? The comparison systems of the human brain investigated by fMRI. *Neuropsychologia, 43*, 1238–1248.

Kadosh, R.C., Kadosh, K.C., Kaas, A., Henik, A., & Goebel, R. (2007). Notation-dependent and–independent representations of numbers in the parietal lobes. *Neuron, 53*, 307–314.

Kahn, L., & Civil. M. (2001). Unearthing the mathematics of a classroom garden. In A. McIntyre., A. Rosebery & N. Gonzalez,. (Eds.), *Classroom diversity: Connecting curriculum to students' lives* (pp. 37–50). Portsmouth, NH: Heinemann.

Kail, R. (2003). Information processing and memory. In M. Bornstein, L. Davidson, C. Keyes & K. Moore (Eds.), *Crosscurrents in contemporary psychology* (pp. 269–279). Mahwah, NJ: Lawrence Erlbaum.

Kalyuga, S., Chandler, P., Tuovinen, J., & Sweller, J. (2001). When problem solving is superior to studying worked examples. *Journal of Educational Psychology, 93*, 579–588.

Kalyuga, S., & Sweller, J. (2004). Measuring knowledge to optimize cognitive load factors during instruction. *Journal of Educational Psychology, 96*, 558–568.

Kao, Y.S., & Anderson, J. R. (2006, July). *What are they thinking? Decomposing a complex task*. Poster presented at the 28th Annual Conference of the Cognitive Science Society, Vancouver, British Columbia, Canada.

Kao, Y., & Anderson, J. (2007). *Spatial ability and geometry achievement: A Pilot study*. Poster presented at the Institute of Education Sciences Conference, Washington, DC.

Kao, Y.S., Roll, I., & Koedinger, K.R. (2007, August). *Sources of difficulty in multistep geometry area problem solving*. Poster presented at the 29th Annual Meeting of the Cognitive Science Society, Nashville, TN.

Kapadia, R. (1974). A critical examination of Piaget-Inhelder's view on topology. *Educational Studies in Mathematics, 5*, 419–424.

Kawashima, R., Taira, M., Okita, K., Inoue, K., Tajima, N., Yoshida, H., et al. (2004). A functional MRI study of simple arithmetic—a comparison between children and adults. *Cognitive Brain Research, 18,* 227–233.

Keller, J. (2007). Stereotype threats in classroom settings: The interactive effect of domain identification, task difficulty & stereotype threat on female students' math performance. *British Journal of Educational Psychology, 77*, 323–338.

Kerkman, D. D., & Siegler, R. S. (1993). Individual differences and adaptive flexibility in lower-income children's strategy choices. *Learning and Individual Differences, 5*, 113–136.

Kerkman, D.D., & Wright, J.C. (1988). An exegesis of two theories of compensation development: Sequential decision theory and information integration theory. *Developmental Review, 8*, 323–360.

Kieren, T.E. (1988). Personal knowledge of rational numbers: Its intuitive and formal development. In J. Hiebert & M. Behr (Eds.), *Number concepts and operations in the middle grades* (pp. 162–181). Reston, VA: National Council of Teachers of Mathematics.

Kintsch, W. (1968) Recognition and free recall of organized lists. *Journal of Experimental Psychology, 78*, 481–487.

Kintsch, W., & Greeno, J.G. (1985). Understanding and solving arithmetic word problems. *Psychological Review, 92*, 109–129.

Kirshner, D. (1989). The visual syntax of algebra. *Journal for Research in Mathematics Education, 20*, 274–287.

Klapp, S.T., Boches, C.A., Trabert, M.L., & Logan, G.D. (1991). Automatizing alphabet arithmetic: II. Are there practice effects after automaticity is achieved? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 196–209.

Klein, A., & Bisanz, J. (2000). Preschoolers doing arithmetic: The concepts are willing by the working memory is weak. *Canadian Journal of Experimental Psychology, 54*, 105–115.

Klein, A., & Starkey, P. (2004). Fostering preschool children's mathematical knowledge: Findings from the Berkeley math readiness project. In D.H. Clements, & J. Sarama (Eds.) & A. M. DiBiase (Assoc. Ed.), *Engaging young children in mathematics: Standards for early childhood mathematics education* (pp. 343–360). Mahwah, NJ: Lawrence Erlbaum Associates.

Knuth, E.J., Stephans, A.C., McNeil, N.M., & Alibali, M.W. (2006). Does understanding the equal sign matter? Evidence from solving equations. *Journal for Research in Mathematics Education, 37*, 297–312.

Kobayashi, T., Hiraki, K., Mugitani, R., & Hasegawa, T. (2004). Baby arithmetic: One object plus one tone. *Cognition, 91*, B23–B34.

Koedinger, K.R., Anderson, J.R., Hadley, W.H., & Mark, M.A. (1997). Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education*, *8*, 30–43.

Koedinger, K.R., & Nathan, M.J. (2004). The real story behind story problems: Effects of representations on quantitative reasoning. *Journal of the Learning Sciences*, *13*, 129–164.

Kong, J., Wang, C., Kwong, K., Vangel, M., Chua, E., & Gollub, R. (2005). The neural substrate of arithmetic operations and procedure complexity. *Cognitive Brain Research, 22*, 397–405.

Koontz, K.L., & Berch, D.B. (1996). Identifying simple numerical stimuli: Processing inefficiencies exhibited by arithmetic learning disabled children. *Mathematical Cognition*, *2*, 1–23.

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory. *Psychological Review*, *103*, 490–517.

Kosc, L. (1974). Developmental dyscalculia. *Journal of Learning Disabilities, 7*, 164–177.

Koshmider, J.W., & Ashcraft, M.H. (1991). The development of children's mental multiplication skills. *Journal of Experimental Child Psychology, 51*, 53–89.

Kuhn, D., & Pease, M. (2006). Do children and adults learn differently? *Journal of Cognition and Development, 7*, 279–293.

Kulik, J.A., & Kulik, C.C. (1984). Effects of accelerated instruction on students. *Review of Educational Research, 54*, 409–425.

Kutner, M., Greenberg, E., & Baer, J. (2006). A first look at the literacy of America's adults in the 21st Century. Retrieved August 1, 2007 from http://nces.ed.gov/NAAL/PDF/2006470.PDF.

Kutner, M., Greenberg, E., Jin, Y., Boyle, B., Hsu, Y., Dunleavy, E., et al. (2007). *Literacy in everyday life: Results from the 2003 National Assessment of Adult Literacy*. NCES 2007–480. Retrieved August 1, 2007 from http://nces.ed.gov/Pubs2007/2007480.pdf.

Ladson-Billings, G. (1997). It doesn't add up: African American students' mathematics achievement. *Journal for Research in Mathematics Education, 28*, 697–708.

Landerl, K., Bevan, A., & Butterworth, B. (2004). Developmental dyscalculia and basic numerical capacities: A study of 8–9-year-old students. *Cognition, 93,* 99–125.

Larkin, J., McDermott, J., Simon, D.P., & Simon, H.A. (1980, June 20). Expert and novice performance in solving physics problems. *Science,* 208, 1335–1342.

Larkin, J.H., & Simon, H.A. (1987). When a diagram is (sometimes) worth ten thousand words. *Cognitive Science, 11,* 65–99.

Laski, E.V., & Siegler, R.S. (2007). Is 27 a big number? Correlational and causal connections among numerical categorization, number line estimation, and numerical magnitude comparison. *Child Development, 78*(6), 1723–1743

Le Corre, M., Van de Walle, G., Brannon, E.M., & Carey, S. (2006). Re-visiting the competence/performance debate in the acquisition of the counting principles. Cognitive Psychology, 52, 130–169.

Lee, K,. Ng, S.F., Ng, E.L., & Lim, Z.Y. (2004). Working memory and literacy as predictors of performance on algebraic word problems. *Journal of Experimental Child Psychology, 89,* 140–158.

LeFevre, J.A., Bisanz, J., Daley, K. E., Buffone, L., Sadesky, G. S., & Greenham, S.L. (1996). Multiple routes to solution of single-digit multiplication problems. *Journal of Experimental Psychology: General, 125,* 284–306.

LeFevre, J., DeStefano, D., Coleman, B., & Shanahan, T. (2005). Mathematical cognition and working memory. In J.I.D. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 361–377). New York: Psychology Press.

LeFevre, J., Greenham, S.L., & Waheed, N. (1993). The development of procedural and conceptual knowledge in computational estimation. *Cognition and Instruction, 11,* 95–132.

LeFevre, J.A., & Liu, J. (1997). The role of experience in numerical skill: Multiplication performance in adults from Canada and China. *Mathematical Cognition, 3,* 31–62.

LeFevre, J.A., & Morris, J. (1999). More on the relation between division and multiplication in simple arithmetic: Evidence for mediation of division solutions via multiplication. *Memory & Cognition, 27,* 803–812.

LeFevre, J.A., Smith-Chant, B. L., Fast, L., Skwarchuk, S.L., Sargla, E., Arnup, J.S., Penner-Wilger, M., Bisanz, J., & Kamawar, D. (2006). What counts as knowing? The development of conceptual and procedural knowledge of counting from kindergarten through Grade 2. *Journal of Experimental Child Psychology, 93,* 285–303.

Leighton, R., (Ed.). (2006). *Classic Feynman: All the adventures of a curious character*. New York: Norton.

Lemaire, P., & Lecacheur, M. (2002). Children's strategies in computational estimation. *Journal of Experimental Child Psychology, 82,* 281–304.

Lemaire, P., Lecacheur, M., & Farioli, F. (2000). Children's strategy use in computational estimation. *Canadian Journal of Experimental Psychology, 54,* 141–148.

Lemaire, P., & Siegler, R.S. (1995). Four aspects of strategic change: Contributions to children's learning of multiplication. J*ournal of Experimental Psychology: General, 124,* 83–97.

Lemke, M., Sen, A., Pahlke, E., Partelow, L., Miller, D., Williams, T., et al. (2005). *International Outcomes of Learning in Mathematics Literacy and Problem Solving: PISA 2003 Results from the U.S. Perspective, Highlights December 2004.* NCES 2005-003.

Lepper, M.R., Corpus, H.J., & Iyengar, S.S. (2005). Intrinsic and extrinsic motivational orientations in the classroom: Age differences and academic correlates. *Journal of Educational Psychology, 97*, 184–196.

Lewis, A.B. (1989). Training students to represent arithmetic word problems. *Journal of Educational Psychology, 81*, 521–531.

Lewis, A.B., & Mayer, R.E. (1987). Students' miscomprehension of relational statements in arithmetic word problems. *Journal of Educational Psychology, 79*, 363–371.

Lewis, C. (1981). Skill in algebra. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 85–110). Hillsdale, NJ: Lawrence Erlbaum Associates.

Liben, L. (2002). Spatial development in childhood: Where are we now? In U. Goswami (Ed.), *Blackwell handbook of cognitive development* (pp. 326–348). Oxford: Blackwell Publishing.

Linnenbrink, E. (2005). The dilemma of performance-approach goals: The use of multiple goal contexts to promote students' motivation and learning. *Journal of Educational Psychology, 97*, 197–213.

Lipkus, I.M., Samsa, G., & Rimer, B.K. (2001). General performance on a numeracy scale among highly educated samples. *Medical Decision Making, 21*, 37–44.

Lloyd, F.J., & Reyna, V.F. (2001). A web exercise in evidence-based medicine using cognitive theory. *Journal of General Internal Medicine, 16*, 94–99.

Locuniak, M.N., & Jordan, N.C. (in press). Using kindergarten number sense to predict calculation fluency in second grade. *Journal of Learning Disabilities.*

Logie, R.H., Gilhooly, K.J., & Wynn, V. (1994). Counting on working memory in arithmetic problem solving. *Memory & Cognition, 22*, 395–410.

Long, J., Monoi, S., Harper, B., Knoblauch, D., & Murphy, P. (2007). Academic motivation & achievement among urban adolescents. *Urban Education, 42,* 196–222.

Loveless, T. (2003). Trends in math: The importance of basic skills. *Brookings Review, 21*, 40–43.

Loveless, T., & Coughlan, J. (2004). The arithmetic gap: Can U.S. students add, subtract, multiply, and divide better than previous generations? Does it matter? *Educational Leadership, 61*, 55–59.

Low, R., Over, R., Doolan, L., & Michell, S. (1994). Solution of algebraic word problems following training in identifying necessary and sufficient information within problems. *American Journal of Psychology, 107,* 423–439.

Lubinski, D., Benbow, C.P., Shea, D.L., Efekhari-Sanjani, H., & Halvorson, M.B.J. (2001). Men and women at promise for scientific excellence: Similarity not dissimilarity. *Psychological Science*, 23, 309–317.

Mabbott, D.J., & Bisanz, J. (2003). Developmental change and individual differences in children's multiplication. *Child Development, 74*, 1091–1107.

MacGregor, M., & Stacey, K. (1993). Cognitive models underlying students' formulation of simple linear equations. *Journal for Research in Mathematics Education, 24*, 217–232.

Mack, N.K. (1995). Confounding whole-number and fraction concepts when building on informal knowledge. *Journal for Research in Mathematics Education, 26,* 422–441.

Malloy, C.E., & Jones, M.G. (1998). An investigation of African American students' mathematical problem solving. *Journal for Research in Mathematics Education 29*, 143–163.

Martin, S.A., & Bassok, M. (2005). Effects of semantic cues on mathematical modeling: Evidence from word-problem solving and equation construction tasks. *Memory & Cognition, 33*, 471–478.

Mauro, D.G., LeFevre, J.A., & Morris, J. (2003). Effects of problem format on division and multiplication performance: Division facts are mediated via multiplication-based representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 163–170.

Mayer, R.E. (1981). Frequency norms and structural analysis of algebra story problems into families, categories, and templates. *Instructional Science, 10*, 135–175.

Mayer, R.E. (1982). Memory for algebra story problems. *Journal of Educational Psychology, 74*, 199–216.

Mayfield, K.H., & Chase, P.N. (2002). The effects of cumulative practice on mathematics problem solving. *Journal of Applied Behavior Analysis, 35*, 105–123.

Mazzocco, M.M.M., & Devlin, K.T. (in press). Parts and 'holes': Gaps in rational number sense among children with vs. without mathematical learning disabilities. *Developmental Science*.

McCleary, J. (2006). *A first course in topology: Continuity and dimension*. Providence, RI: American Mathematical Society.

McKinney, J.D., & Speece, D.L. (1986). Academic consequences and longitudinal stability of behavioral subtypes of learning disabled children. J*ournal of Educational Psychology, 78*, 365–372.

McLean, J.F., & Hitch, G.J. (1999). Working memory impairments in children with specific arithmetic learning difficulties. *Journal of Experimental Child Psychology, 74*, 240–260.

McNeil, N.M., & Alibali, M.W. (2005). Knowledge change as a function of mathematical experience: All contexts are not created equal. *Journal of Cognition and Development, 6*, 285–306.

McNeil, N.M., Grandau, L., Knuth, E.J., Alibali, M.W., Stephans, A.C., Hattikudur, S., et al. (2006). Middle-school students' understanding of the equal sign: The books they read can't help. *Cognition and Instruction, 24*, 367–385.

Meece, J.L., Anderman, E.M., & Anderman, L.H. (2006). Classroom goal structure, student motivation, and academic achievement. *Annual Review of Psychology, 57*, 487–503.

Menon, C., Rivera, S.M., White, C.D., Glover, G.H., & Reiss, A.L. (2000). Dissociating prefrontal and parietal cortex activation during arithmetic processing. *NeuroImage, 12*, 357–365.

Metcalfe, J. (2002). Is study time allocated selectively to a region of proximal learning? *Journal of Experimental Psychology: General, 131*, 349–363.

Midgley, C., Kaplan, A., Middleton, M., Maehr, M.L., & Urban, T. (1998). The development and validation of scales assessing students' goal orientations. *Contemporary Educational Psychology, 23*, 113–31.

Miller, K. (1984). Child as the measurer of all things: Measurement of procedures and the development of quantitative concepts. In C. Sophian (Ed.), *Origins of cognitive skills: The eighteenth annual Carnegie symposium on cognition* (pp. 193–228). Hillsdale, NJ: Lawrence Erlbaum Associates.

Miller, K.F., Kelly, M., & Zhou, X. (2005). Learning mathematics in China and the United States: Cross-cultural insights into the nature and course of preschool mathematical development. In J.I.D. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 163–177). New York: Psychology Press.

Miller, K.F., & Paredes, D.R. (1990). Starting to add worse: Effects of learning to multiply on children's addition. *Cognition, 37*, 213–242.

Miller, K., Perlmutter, M., & Keating, D. (1984). Cognitive arithmetic: Comparison of operations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 46–60.

Miller, K.F., Smith, C.M., Zhu, J., & Zhang, H. (1995). Preschool origins of cross-national differences in mathematical competence: The role of number-naming systems. *Psychological Science, 6*, 56–60.

Mills, C.J., Ablard, K.E., & Stumpf, H. (1993). Gender differences in academically talented young students' mathematical reasoning: Patterns across age and subskills. *Journal of Educational Psychology, 85*, 340–346.

Mitchell, J.H., Hawkins, E.F., Stancavage, F.B., & Dossey, J.A. (1999). Estimation skills, mathematics-in-context, and advanced skills in mathematics. *Education Statistics Quarterly, 2*, Issue 1.

Miura, I.T., Okamoto, Y., Kim, C.C., Steere, M., & Fayol, M. (1993). First graders' cognitive representation of number and understanding of place value: Cross-national comparisons—France, Japan, Korea, Sweden, and the United States. *Journal of Educational Psychology, 85*, 24–30.

Miura, I.T., Okamoto, Y., Vlahovic-Stetic, V., Kim, C.C., & Han, J.H. (1999). Language supports for children's understanding of numerical fraction: Cross-national comparisons. *Journal of Experimental Child Psychology, 74*, 356–365.

Mix, K.S., Levine, S.C., & Huttenlocher, J. (1999). Early fraction calculation ability. *Developmental Psychology, 35*, 164–74.

Modestou, M., Gagatsis, A., & Pitta-Pantazi, D. (2004). Students' improper proportional reasoning: The case of area and volume of rectangular figures. In M.J. Høines, & A.B. Fuglestad (Eds.), *Proceedings of the 28th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 3, pp. 345–352). Bergen, Norway.

Moll, L., Amanti, C., & Neff, D. (2005). Funds of knowledge for teaching: Using a qualitative approach to connect homes and classrooms. In N. Gonzalez, L. Moll & C. Amanti (Eds.), *Funds of knowledge: Theorizing practices in households communities and classrooms* (pp. 71–87). Mahwah, NJ: L. Erlbaum.

Mooney, E.S., & Thornton, C.A. (1999). Mathematics attribution differences by ethnicity and socioeconomic status. *Journal of Education for Students Placed at Risk, 4,* 321–332.

Moore, C.F., Dixon, J.A., & Haines, B.A. (1991). Components of understanding in proportional reasoning: A fuzzy set representation of developmental progressions. *Child Development, 62*, 441–459.

Morales, R.V., Shute, V.J., & Pellegrino, J.W. (1985). Developmental differences in understanding and solving simple mathematics word problems. *Cognition and Instruction, 2*, 41–57.

Moss, J. (2005). Pipes, tubes, and beakers: Teaching rational number. In J. Bransford, & S. Donovan (Eds.), *How Children Learn: History Science and Mathematics in the Classroom*. Washington, DC: National Academy Press.

Mullis, V.S., Martin, M., Gonzalez, E.J., Chrostowski, S.J., & Smith, T.A. (2004). *TIMSS 2003 international mathematics report: Findings from IEA's Trends in International Mathematics and Science Study at the fourth and eighth grades*. Retrieved August 1, 2007, from http://timss.bc.edu/PDF/t03_download/T03INTLMATRPT.pdf.

Muraven, M., Baumeister, R.F., & Tice, D.M. (1999). Longitudinal improvement of self-regulation through practice: Building self-control through repeated exercise. *Journal of Social Psychology, 139*, 446–457.

Murphy, M.M., Mazzocco, M.M.M., Hanich, L.B., & Early, M.C. (2007). Cognitive characteristics of children with mathematics learning disability (MLD) vary as a function of the cutoff criterion used to define MLD. *Journal of Learning Disabilities, 40*, 458–478.

National Council of Teachers of Mathematics. (1980). *An agenda for action: Recommendations for school mathematics of the 1980s*. Reston, VA: Author.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Retrieved August 1, 2007 from http://standards.nctm.org/document/index.htm.

National Council of Teachers of Mathematics. (2006). *Curriculum focal points for prekindergarten through grade 8 mathematics: A quest for coherence*. Reston, VA: National Council of Teachers of Mathematics. Retrieved August 1, 2007 from http://www.nctm.org/standards/focalpoints.aspx?id=298.

National Science Board Commission on 21st Century Education in Science, Technology, Engineering, and Mathematics (STEM). Retrieved September 29, 2006 from http://www.nsf.gov/nsb/stem/commission_index.jsp.

Navarro, R., Flores, L., & Worthington, R. (2007). Mexican American middle school students goal intentions in mathematics & science: A test of social cognitive theory. *Journal of Counseling Psychology, 54*, 320–335.

Nelson, T.O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 26, pp. 125–173). New York: Academic Press.

Nettles, M. (2000, October). The status and consequences of admissions test performance for the nation's demographically diverse population of aspiring students. *Paper presented at the Fulbright's Educational Experts Seminar*. Ann Arbor, MI.

Newcombe, N.S., & Huttenlocher, J. (2000). *Making space: The development of spatial representation and reasoning*. Cambridge, MA: MIT Press.

Newcombe, N.S., & Huttenlocher, J. (2006). Development of spatial cognition. In D. Kuhl, & R.S. Siegler (Vol. Eds.), *Cognition, perception, and language,* Vol. 2 (pp. 734–776). In W. Damon (Gen. Ed.), *Handbook of child psychology* (6th Ed.). New York: John Wiley & Sons.

Newell, A., & Rosenbloom, P.S. (1981). Mechanisms of skill acquisition and the power law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1–55). Hillsdale, NJ: Lawrence Erlbaum Associates.

Newell, A., & Simon, H.A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Nicholls, J.G. (1984). Achievement motivation: Conceptions of ability, subjective experience, task choice, and performance. *Psychological Review, 91,* 328–346.

Nowell, A., & Hedges, L.V. (1998). Trends in gender differences in academic achievement from 1960 to 1994: An analysis of differences in mean, variance, and extreme scores. *Sex Roles, 39,* 21–43.

Nunes, T., Schliemann, A.D., & Carraher, D.W. (1993). *Street mathematics and school mathematics.* Cambridge, UK: Cambridge University Press.

Ohlsson, S., & Rees, E. (1991). The function of conceptual understanding in the learning of arithmetic procedures. *Cognition and Instruction, 8,* 103–179.

Opfer, J., & Siegler, R.S. (2007). Representational change and children's numerical estimation. *Cognitive Psychology, 55*, 169–195.

Opfer, J.E., Thompson, C.A., & DeVries, J. (2007). Why children make better estimates of fractional magnitude than adults. In D.S. McNamara & G. Trafton (Eds.), *Proceedings of the XIX Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum Associates.

Osborne**,** J.W. (2007). Linking stereotype threat and anxiety. *Educational Psychology, 27,* 135–154.

Ostad, S.A. (1998). Developmental differences in solving simple arithmetic word problems and simple number-fact problems: A comparison of mathematically normal and mathematically disabled children. *Mathematical Cognition, 4*, 1–19.

Paik, J.H., & Mix, K.S. (2003). U.S. and Korean children's comprehension of fraction names: A reexamination of cross-national differences. *Child Development, 74*, 144–154.

Pashler, H.E., (1999). *The psychology of attention*. Cambridge, MA: MIT Press.

Pashler, H., Rohrer, D., Cepeda, N.J., & Carpenter, S.K. (2007). Enhancing learning and retarding forgetting: Choices and consequences. *Psychonomic Bulletin & Review, 14,* 187–193.

Patrick, H., Kaplan, A., & Ryan, A.M. (2007). Early adolescents' perceptions of the classroom social environment, motivational beliefs, and engagement. *Journal of Educational Psychology, 99,* 83–98.

Pauli, P., Lutzenberger, W., Rau, H., Birbaumer, N., Rickard, T.C., Yaroush, R.A., et al. (1994) Brain potentials during mental arithmetic: Effects of extensive practice and problem difficulty. *Cognitive Brain Research, 2*, 21–29.

Paull, D.R. (1972). *The ability to estimate in mathematics*. [unpublished doctoral dissertation]. Columbia University.

Pawley, D., Ayres, P., Cooper, M., & Sweller, J. (2005). Translating words into equations: A cognitive load theory approach. *Educational Psychology, 25*, 75–97.

Perry, T., Steele, C., & Hillard, A. (2003). *Young, gifted and Black: Promoting high achievement among African American students.* Boston: Beacon Press.

Petitto, A.L. (1990). Development of numberline and measurement concepts. *Cognition and Instruction, 7*, 55–78.

Piaget, J., & Garcia, R. (1989). *Psychogenesis and the history of science*. New York: Columbia University Press.

Piaget, J., & Inhelder, B. (1967). *The child's conception of space* (F.J. Langdon & J.L. Lunzer, Trans.). New York: Norton. (Original work published 1948).

Piazza, M., Pinel, P., Bihan, D.L., & Dehaene, S. (2007). A magnitude code common to numerosities and number symbols in human intraparietal cortex. *Neuron, 53*, 293–305.

Pinel, P., Piazza, D., Le Bihan, D., & Dehaene, S. (2004). Distributed and overlapping cerebral representations of number, size, and luminance during comparative judgments. *Neuron, 41*, 983–993.

Platt, J.R. (1964). Strong inference. *Science, 146*, 347–353.

Pratt, M.W., & Savoy-Levine, K.M. (1998). Contingent tutoring of long-division skills in fourth and fifth graders: Experimental tests of some hypotheses about scaffolding. *Journal of Applied Developmental Psychology, 19*, 287–304.

Pressley, M., & Woloshyn, V. (1995). *Cognitive strategy instruction that really improves Children's academic performance*. Cambridge MA: Brookline Books.

Qin, Y., Carter, C.S., Silk, E.M., Stenger, V.A., Fissell, K., Goode, A., et al. (2004). The change of the brain activation patterns as children learn algebra equation solving. *Proceedings of the National Academy of Sciences USA, 101*, 5686–5691.

Qin, Y., Sohn, M.H., Anderson, J.R., Stenger, V.A., Fissell, K., Goode, A., et al. (2003). Predicting the practice effects on the blood oxygenation level-dependent (BOLD) function of fMRI in a symbolic manipulation task. *Proceedings of the National Academy of Sciences USA, 100*, 4951–4956.

Ramani, G.B., & Siegler, R.S. (2008). Promoting broad and stable improvements in low-income children's numerical knowledge through playing number board games. *Child Development, 79*, 375–394.

Randhawa, B.S., Beamer, J.E., & Lundberg, I. (1993). Role of mathematics self-efficacy in the structural model of mathematics achievement. *Journal of Educational Psychology, 85*, 41–48.

Ranney, M. (1987). The role of structural context in perception: Syntax in the recognition of algebraic expressions. *Memory & Cognition, 15*, 29–41.

Rasmussen, C., Ho, E., & Bisanz, J. (2003). Use of mathematical principle of inversion in young children. *Journal of Experimental Child Psychology, 85*, 89–102.

Reder, L.M. (1987). Strategy selection in question answering. *Cognitive Psychology 19*, 90–138.

Reed, S.K. (1987). A structure-mapping model for word problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 124–139.

Reed, S.K. (1993). A schema-based theory of transfer. In D.K. Detterman & R.J. Sternberg (Eds.), *Transfer on trial: Intelligence, cognition and instruction* (pp. 39–67). Norwood, NJ: Ablex.

Reed, S.K., & Bolstad, C.A. (1991). Use of examples and procedures in problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 753–766.

Reed, S.K., Dempster, A., & Ettinger, M. (1985). Usefulness of analogous solutions for solving algebra word problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*, 106–125.

Reed, S.K., Willis, D., & Guarino, J. (1994). Selecting examples for solving word problems. *Journal of Educational Psychology, 86*, 380–388.

Rehder, B. (1999). Detecting unsolvable algebra word problems*. Journal of Educational Psychology*, 91, 669–683.

Resnick, L.B. (1992). From protoquantities to operators: Building mathematical competence on a foundation of everyday knowledge. In G. Leinhardt, R. Putnam & R.A. Hattrup (Eds.), *Analysis of arithmetic for mathematics teaching* (pp. 373–425). Hillsdale, NJ: Lawrence Erlbaum Associates.

Resnick, L., & Ford, W. (1981). *The psychology of mathematics for instructio*n. Hillsdale, NJ: Lawrence Erlbaum Associates.

Resnick, L.B. & Resnick, D.P. (1992). Assessing the thinking curriculum: New tools for educational reform. In B. Gifford & M. O'Connor (Eds.), *Changing assessments: Alternative views of aptitude, achievement, and instruction* (pp. 37–75). Norwell, MA: Kluwer Academic Publishers.

Reyna, V.F. (2004). How people make decisions that involve risk: A dual-processes approach. *Current Directions in Psychological Science, 13*, 60–66.

Reyna, V.F., & Brainerd, C.J. (1991). Fuzzy-trace theory and children's acquisition of scientific and mathematical concepts. *Learning and Individual Differences, 3,* 27–60.

Reyna, V.F., & Brainerd, C.J. (1993). Fuzzy memory and mathematics in the classroom. In R. Logie & G. Davies (Eds.), *Memory in Everyday Life: Advances in Psychology* (pp. 91–119. Amsterdam: North-Holland.

Reyna, V.F., & Brainerd, C.J. (1994). The origins of probability judgment: A review of data and theories. In G. Wright & P. Ayton (Eds.), *Subjective Probability* (pp. 239–272). New York: Wiley.

Reyna, V.F., & Brainerd, C.J. (2007). The importance of mathematics in health and human judgment: Numeracy, risk communication, and medical decision making. *Learning and Individual Differences*, *17*, 147–159.

Reyna, V.F., & Brainerd, C.J. (2008). Numeracy, ratio bias, and denominator neglect in judgments of risk and probability. *Learning and Individual Differences, 18,* 93–111.

Reyna, V.F., & Ellis, S.C. (1994). Fuzzy-trace memory and framing effects in children's risky decision making. *Psychological Science, 5*, 275–279.

Reyna, V.F., & Mills, B.A. (2007). Converging evidence supports fuzzy-trace theory's nested sets hypothesis (but not the frequency hypothesis). *Behavioral and Brain Sciences, 30*(3), 278–280.

Reys, R.E. (1984). Mental computation and estimation: Past, present, and future. *The Elementary School Journal, 84*, 547–557.

Reys, R.E., Reys, B. J., Nohda, N., & Emori, H. (1995). Mental computation performance and strategy use of Japanese students in Grades 2, 4, 6, and 8. *Journal for Research in Mathematics Education, 26*, 304–326.

Reys, R.E., Reys, B.J., Nohda, N., Ishida, J., Yoshikawa, S., & Shimizu, K. (1991). Computational estimation performance and strategies used by fifth- and eighth-grade Japanese students. *Journal for Research in Mathematics Education, 22*, 39–58.

Reys, R.E., Rybolt, J.F., Bestgen, B.J., & Wyatt, J.W. (1982). Processes used by good computational estimators. *Journal for Research in Mathematics Education, 13*, 183–201.

Rhymer, K.N., Henington, C., Skinner, C.H., & Looby, E.J. (1999). The effects of explicit timing on mathematics performance in second-grade Caucasian and African American students. *School Psychology Quarterly, 14*, 397–407.

Richardson, T.M., & Benbow, C.P. (1990). Long-term effects of acceleration on social and emotional adjustment of mathematically precocious youth. *Journal of Educational Psychology, 82*, 464–470.

Rickard, T.C. (1997). Bending the power law: A CMPL theory of strategy shifts and the automatization of cognitive skills. *Journal of Experimental Psychology: General, 126*, 288–311.

Rickard, T.C. (2004). Strategy execution in cognitive skill learning: An item-level test of candidate models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 65–82.

Rickard, T.C. (2005). A revised identical elements models of arithmetic fact representation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 250–257.

Rickard, T.C., Healy, A.F., & Bourne, L.E., Jr. (1994). On the cognitive structure of basic arithmetic skills: Operation, order, and symbol transfer effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 1139–1153.

Rickard, T.C., Romero, S.G., Basso, G., Wharton, C., Flitman, S., & Grafman, J. (2000). The calculating brain: An fMRI study. *Neuropsychologia, 38*, 325–335.

Rieber, R.W., & Carton, A.S. (Eds.). (1987). *The collected works of L.S. Vygotsky: Volume 1: Problems of general psychology*. New York: Plenum Press.

Riley, M.S., Greeno, J.G., & Heller, J.I. (1983). Development of children's problem-solving ability in arithmetic. In H. P. Ginsburg (Ed.), *The development of mathematical thinking* (pp. 153–196). New York: Academic Press.

Ritter, S., Anderson, J.R., Koedinger, K.R., & Corbett, A. (2007). Cognitive tutor: Applied research in mathematics education. *Psychonomic Bulletin & Review, 14*, 249–255.

Rittle-Johnson, B. & Alibali, M.W. (1999). Conceptual and procedural knowledge of mathematics: Does one lead to the other? *Journal of Educational Psychology, 91*, 175–189.

Rittle-Johnson, B., & Siegler, R.S. (1998). The relation between conceptual and procedural knowledge in learning mathematics: A review. In C. Dolan (Ed.), *The development of mathematical skills* (pp. 75–110). Hove East Sussex, UK: Psychology Press.

Rittle-Johnson, B., Siegler, R.S., & Alibali, M.W. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of Educational Psychology, 93,* 346–362.

Rivera, S.M., Reiss, A.L., Eckert, M.A., & Menon, V. (2005). Developmental changes in mental arithmetic: Evidence for increased functional specialization in the left inferior parietal cortex. *Cerebral Cortex, 15,* 1779–1790.

Robinson, K.M., Arbuthnott, K.D., & Gibbons, K.A. (2002). Adults' representations of division facts: A consequence of learning history? *Canadian Journal of Experimental Psychology, 56,* 302–309.

Robinson, K.M., Arbuthnott, K.D., Rose, D., McCarron, M.C., Globa, C.A., & Phonexay, S.D. (2006). Stability and change in children's division strategies. *Journal of Experimental Child Psychology, 93,* 224–238.

Robinson, K.M., & Ninowski, J.E. (2003). Adults' understanding of inversion concepts: How does performance on addition and subtraction inversion problems compare to performance on multiplication and division problems? *Canadian Journal of Experimental Psychology, 57*, 321–330.

Robinson, K.M., Ninowski, J.E., & Gray, M.L. (2006). Children's understanding of the arithmetic concepts of inversion and associativity. *Journal of Experimental Child Psychology, 94*, 349–362.

Robinson, N.M., Abbott, R.D., Berninger, V.W., & Busse, J. (1996). The structure of abilities in math-precocious young children: Gender similarities and differences. *Journal of Educational Psychology, 88*, 341–352.

Roediger, H.L., & Karpicke, J.D. (2006a). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17,* 249–255.

Roediger, H.L., & Karpicke, J.D. (2006b). The power of testing memory: Basic research and implicational for educational practice. *Perspectives on Psychological Science, 1,* 181–210.

Rohrbeck, C.A., Fantuzzo, J.W., Ginsburg-Block, M.D., & Miller, T.R. (2003). Peer-assisted learning interventions with elementary school students: A meta-analytic review. *Journal of Educational Psychology, 95*, 240–257.

Rohrer, D., & Taylor, K. (2006). The effects of overlearning and distributed practice on the retention of mathematics knowledge. *Applied Cognitive Psychology, 20,* 1209.

Rousselle, L., Palmers, E., & Noel, M. (2004). Magnitude comparison in preschoolers: What counts? Influence of perceptual variables. J*ournal of Experimental Child Psychology, 87*, 57–84.

Royer, J.M., Tronsky, L.N., Chan, Y., Jackson, S.J., & Marchant III, H. (1999). Math-fact retrieval as the cognitive mechanism underlying gender differences in math test performance. *Contemporary Educational Psychology*, *24*, 181–266.

Runquist, W.N. (1983). Some effects of remembering on forgetting. *Memory & Cognition. 11*, 641–650.

Russell, R.L., & Ginsburg, H.P. (1984). Cognitive analysis of children's mathematical difficulties. *Cognition and Instruction, 1*, 217–244.

Ryan, K., & Ryan, A. (2005) Psychological processes underlying stereotype threat and standardized math test performance. *Educational Psychologist*, *40*, 53–63.

Sadler, P.M., & Tai, R.H. (2007). The two high-school pillars supporting college science. *Science*, *317*, 457–458.

Sarama, J. (2004). Technology in early childhood mathematics: Building blocks as an innovative technology-based curriculum. In D.H. Clements & J. Sarama (Eds.) & A.M. DiBiase (Assoc. Ed.), *Engaging young children in mathematics: Standards for early childhood mathematics education* (pp. 361–376). Mahwah, NJ: Lawrence Erlbaum Associates.

Sarama, J., & Clements, D. (2004). Building blocks for early childhood mathematic. *Early Childhood Research Quarterly*, *19,* 1, 181–189.

Sawamura, H., Shima, K., & Tanji, J. (2002, February 21). Numerical representation for action in the parietal cortex of the monkey. *Nature, 415*, 918–922.

Saxton, M., & Cakir, K. (2006). Counting-on, trading and partitioning: Effects of training and prior knowledge on performance on base-10 tasks. *Child Development*, *77*, 767–785.

Schacter, D.L., Wagner, A.D., & Buckner, R.L. (2000). Memory systems of 1999. In E. Tulving & F.I.M. Craik (Eds.), *The Oxford handbook of memory* (pp. 627–644). New York: Oxford University Press.

Schaeffer, B., Eggleston, V.H., & Scott, J.L. (1974). Number development in young children. *Cognitive Psychology, 6*, 357–379.

Schaie, K.W. (1996). *Intellectual development in adulthood: The Seattle longitudinal study*. New York: Cambridge University Press.

Schlottman, A. (2001). Children's probability intuitions: Understanding the expected value of complex gambles. *Child Development, 72*, 103–122.

Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality & Social Psychology, 85*, 440–52.

Schmidt, W. (2003). *Too little to late: American high schools in an international context.* In D. Ravitch (Ed.), *Brookings papers on educational policy* (pp. 253–278). Washington DC: Brookings Institution Press.

Schneider, W., & Shiffrin, R.M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review, 84*, 1–66.

Schoenfeld, A.H. (1985). *Mathematical problem solving*. San Diego, CA: Academic Press.

Schultz, G.F. (1993). Socioeconomic advantage and achievement motivation: Important mediators of academic performance in minority children in urban schools. *Urban Review*, *25*, 221–232.

Schunk, D.H. (1996). Goal and self-evaluative influences during children's cognitive skill learning. *American Educational Research Journal, 33*(2), 359–382.

Schwartz, L.M., Woloshin, S., Black, W.C., &Welch, H.G. (1997). The role of numeracy in understanding the benefit of screening mammography. *Annals of Internal Medicine, 127*(11), 966−972.

Schyns, P.G., Bonnar, L., & Gosselin, F. (2002). Show me the features! Understanding recognition from the use of visual information. *Psychological Science, 13*, 402–409.

Sebrechts, M.M., Enright, M., Bennett, R.E., & Martin, K. (1996). Using algebra word problems to assess quantitative ability: Attributes, strategies, and errors. *Cognition and Instruction, 14*, 285–343.

Seo, K.H., & Ginsburg, H.P. (2003). "You've got to carefully read the math sentence…" Classroom context and children's interpretations of the equals signs. In A. Baroody & A. Dowker (Eds.), *The development of arithmetic concepts and skills: Constructing adaptive expertise* (pp. 161–187). Mahwah, NJ: Lawrence Erlbaum Associates.

Shalev, R.S., Manor, O., & Gross-Tsur, V. (2005). Developmental dyscalculia: A prospective six-year follow-up. *Developmental Medicine & Child Neurology, 47*, 121–125.

Shiffrin, R.M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, *84*, 127–190.

Shrager, J., & Siegler, R.S. (1998). SCADS: A model of children's strategy choices and strategy discoveries. *Psychological Science, 9*, 405–410.

Shuman, M., & Kanwisher, N. (2004). Numerical magnitude and the human parietal lobe: Tests of representational generality and domain specificity. *Neuron, 44*, 557–569.

Siegler, R.S. (1981). Developmental sequences within and between concepts. *Monographs of the Society for Research in Child Development, 46* (Whole No 189).

Siegler, R.S. (1983). Five generalizations about cognitive development. *American Psychologist*, *38*, 263–277.

Siegler, R.S. (1987). The perils of averaging data over strategies: An example from children's addition. *Journal of Experimental Psychology: General*, *116*, 250–264.

Siegler, R.S. (1988a). Individual differences in strategy choices: Good students, not-so-good students, and perfectionists. *Child Development, 59*, 833–851.

Siegler, R.S. (1988b). Strategy choice procedures and the development of multiplication skill. *Journal of Experimental Psychology: General, 117*, 258–275.

Siegler, R.S. (1989). Hazards of mental chronometry: An example from children's subtraction. *Journal of Educational Psychology, 81*, 497–506.

Siegler, R.S. (1991). *Children's thinking*. Englewood Cliffs, NJ: Prentice-Hall.

Siegler, R.S. (1996). *Emerging minds: The process of change in children's thinking*. New York: Oxford University Press.

Siegler, R.S., & Araya, R. (2005). A computational model of conscious and unconscious strategy discovery. *Advances in Child Development and Behavior*, *33*, 1–44.

Siegler, R.S., & Booth, J.L. (2004). Development of numerical estimation in young children. *Child Development*, *75*, 428–444.

Siegler, R.S., & Booth, J.L. (2005). Development of numerical representation. In J.I.D. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 197–212). New York: Psychology Press.

Siegler, R.S., & Jenkins, E. (1989). *How children discover new strategies*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Siegler, R.S., & Kotovsky, K. (1986). Two levels of intellectual precociousness: Shall ever the twain meet? In R.J. Sternberg and J.E. Davidson (Eds.), *Conceptions of giftedness* (pp. 417–435). Cambridge, England: Cambridge University Press.

Siegler, R.S., & Opfer, J. (2003). The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychological Science*, *14*, 237–243.

Siegler, R.S., & Ramani, G.B. (in press). Playing board games promotes low-income children's numerical development. *Developmental Science*.

Siegler, R.S., & Robinson, M. (1982). The development of numerical understandings. In H. W. Reese & L. P. Lipsitt (Eds.), *Advances in child development and behavior: Vol. 16* (pp. 242–312). New York: Academic Press.

Siegler, R.S., & Shrager, J. (1984). Strategy choices in addition and subtraction: How do children know what to do? In C. Sophian (Ed.), *The origins of cognitive skills* (pp. 229–293). Hillsdale, NJ: Lawrence Erlbaum Associates.

Siegler, R.S., & Stern, E. (1998). Conscious and unconscious strategy discoveries: A microgenetic analysis. *Journal of Experimental Psychology: General*, *127*, 377–397.

Silver, E.A. (1986). Using conceptual and procedural knowledge: A focus on relationships. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: The case of mathematics* (pp. 181–198). Hillsdale, NJ: Lawrence Erlbaum Associates.

Simon, T.J., & Rivera, S.M. (2007). Neuroanatomical approaches to the study of mathematical ability and disability. In D.B. Berch & M.M.M. Mazzocco (Eds.), *Why is Math So Hard for Some Children? The Nature and Origins of Mathematical Learning Difficulties and Disabilities* (pp. 283–305). Baltimore, MD: Paul H. Brookes Publishing Co.

Singer, J.A., Kohn, A.S., & Resnick, L.B. (1997). Knowing about proportions in different contexts. In T. Nunes & P. Bryant (Eds.), *Learning and teaching mathematics* (pp. 115–132). East Sussex, UK: Psychology Press.

Singer-Freeman, K., & Goswami, U. (2001). Does half a pizza equal half a box of chocolates? Proportional matching in an analogy paradigm. *Cognitive Development*, *16*, 811–829.

Sirin, S., & Rogers-Sirin, L. (2004). Exploring school engagement of middle class African American adolescents. *Youth & Society*, *35*, 323–340.

Sleeman, D. (1984). An attempt to understand students' understanding of basic algebra. *Cognitive Science*, *8*, 387–412.

Sleeman, D., Kelly, A.E., Martinak, R., Ward, R.D., & Moore, J.L. (1989). Studies of diagnosis and remediation with high school algebra students. *Cognitive Science, 13*, 551–568.

Sloutsky, V.M., Kaminski, J.A., & Heckler, A.F. (2005). The advantage of simple symbols for learning and transfer. *Psychonomic Bulletin & Review, 12*, 508–513.

Smith, C.L., Solomon, G.E., & Carey, S. (2005). Never getting to zero: Elementary school students' understanding of the infinite divisibility of number and matter. *Cognitive Psychology, 51*(2), 101–140.

Smith, H.D. (1999). Use of the anchoring and adjustment heuristic by children. *Current Psychology: Developmental, Learning, Personality, Social, 18*, 294–300.

Smith**,** J.L., Sansone, C., & White, P.H. (2007). The stereotyped task engagement process: The role of interest and achievement motivation. *Journal of Educational Psychology, 99*, 99–114.

Smith, J.P. (1995). Competent reasoning with rational numbers. *Cognition and Instruction, 13,* 3–50.

Sohn, M.H., Goode, A., Koedinger, K.R., Stenger, V.A., Fissell, K., Carter, C.S., & Anderson, J.R. (2004). Behavioral equivalence, but not neural equivalence–neural evidence of alternative strategies in mathematical thinking. *Nature Neuroscience, 7*, 1193–1194.

Song, M.J., & Ginsburg, H.P. (1987). The development of informal and formal mathematical thinking in Korean and U. S. children. *Child Development, 58*, 1286–1296.

Sophian, C. (1997). Beyond competence: The significance of performance for conceptual development. *Cognitive Development, 12*, 281–303.

Sophian, C. (2000). Perceptions of proportionality in young children: Matching spatial ratios. *Cognition*, 75, 145–170.

Sophian, C. (2007). *The origins of mathematical knowledge in childhood. Studies in mathematical thinking and learning*. New York: Lawrence Erlbaum Associates.

Sophian, C., Garyantes, D., & Chang, C. (1997). When three is less than two: Early developments in children's understanding of fractional quantities. *Developmental Psychology, 33*, 731–744.

Sophian, C., Harley, H., & Martin, C.S.M. (1995). Relational and representational aspects of early number development. *Cognition and Instruction, 13,* 253–268.

Sophian, C., & Kailihiwa, C. (1998). Units of counting: Developmental changes. *Cognitive Development, 13,* 561–585.

Sophian, C., & McCorgray, P. (1994). Part-whole knowledge and early arithmetic problem solving. *Cognition and Instruction, 12*, 3–33.

Sophian, C., & Wood, A. (1997). Proportional reasoning in young children: The parts and the whole of it. *Journal of Educational Psychology, 8*9, 309–317.

Southern, W.T., Jones, E.D., & Stanley, J.C. (1993). Acceleration and enrichment: The context and development of program options. In K.A. Heller, F.J. Monks, & H. Passow (Eds.), *International handbook of research and development of giftedness and talent* (pp. 387–409). New York: Pergamon press.

Sowder, J., & Wheeler, M. (1989). The development of concepts and strategies used in computational estimation. *Journal for Research in Mathematics Education, 20*, 130–146.

Spinillo, A.G., Bryant, P. (1991). Children's proportional judgments: The importance of "half." *Child Development*, *62*, 427–440.

Spinillo, A.G., & Bryant, P.E. (1999). Proportional reasoning in young children: Part-part comparisons about continuous and discontinuous quantity. *Mathematical Cognition, 5*, 181–197.

Squire, S., & Bryant, P. (2003). Children's models of division. *Journal of Experimental Child Psychology*, 18, 355–376.

Squire, S., Davies, C., & Bryant, P. (2004). Does the cue help? Children's understanding of multiplicative concepts in different problem contexts. *British Journal of Educational Psychology*, *74*, 515–532.

Stafylidou, S., & Vosniadou, S. (2004). The development of students' understanding of the numerical value of fractions. *Learning and Instruction, 14,* 503–518.

Stanley, J.C. (2000). Helping students learn only what they don't already know. *Psychology, Public Policy, and Law*, *6*, 216–222.

Star, J.R. (2005). Re-"conceptualizing" procedural knowledge. *Journal for Research in Mathematics Education, 36*, 404–411.

Star, J.R. (2007). Foregrounding procedural knowledge. *Journal for Research in Mathematics Education*, *38*, 132–135.

Starch, D. (1911). Transfer of training in arithmetical operations. *Journal of Educational Psychology, 2*, 306–311.

Starkey, P. (1992). The early development of numerical reasoning. *Cognition, 43*, 93–126.

Starkey, P., Klein, A., Chang, I., Dong, Q., Pang, L., & Shou, Y. (1999, April). Environmental supports for young children's mathematical development in China and the United States. *Paper presented at the biennial meeting of the Society for Research in Child Development*, Albuquerque, NM.

Starkey, P., Klein, A., & Wakeley, A. (2004). Enhancing young children's mathematical knowledge through a pre-kindergarten mathematics intervention. *Early Childhood Research Quarterly 19*, 99–120.

Steel, S., & Funnell, E. (2001). Learning multiplication facts: A study of children taught by discovery methods in England. *Journal of Experimental Child Psychology*, *79*, 37–55.

Steele, C. (1992, April). Race and the schooling of Black Americans. *Atlantic Monthly, 269*, 68–78.

Steele, C.M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist, 52,* 613–629.

Steele, C.M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, 69, 797–811.

Steele, C.M., & Aronson, J. (1998). Stereotype threat and the test performance of academically successful African Americans. In Jencks & Phillips (Eds.), *The black-white test score gap* (pp. 401–427). Washington, DC: Brookings Institution Press.

Sternberg, R.J. (Ed.). (1999). *Handbook of creativity*. Cambridge: Cambridge University Press.

Sternberg, R.J., (2006). Recognizing neglected strengths. *Educational Leadership,* 30–35.

Stevens, T., Olivarez, A., & Hamman, D. (2006). The role of cognition motivation, and emotion in explaining the mathematics achievement gap between Hispanic and White students. *Hispanic Journal of Behavioral Sciences, 28,* 161–186.

Stevens, T., Olivarez, A., Lan, W., & Tallent-Runnels, M.K. (2004). Role of mathematics self-efficacy and motivation in mathematics performance across ethnicity. *Journal of Educational Research, 97,* 208–221.

Stevenson, H.W., Chen, C., & Utal, D.H. (1990). Beliefs and achievement: A Study of Black, White, and Hispanic Children. *Child Development, 61, 508–523.*

Stevenson, H.W., Lee, S.Y., Chen, C., Lummis, M., Stigler, J., Fan, L., et al. (1990). Mathematics achievement of children in China and the United States. *Child Development, 61,* 1053–1066.

Stevenson, H.W., Lee, S.Y., Chen, C., Stigler, J.W., Hsu, C.C., & Kitamura, S. (1990). Contexts of achievement: A study of American, Chinese, and Japanese children. *Monographs of the Society for Research in Child Development, 55* (1–2, Serial No. 221).

Stevenson, H.W., Stigler, J.W., Lee, S.Y., Lucker, G.W., Kitamura, S., & Hsu, C.C. (1985). Cognitive performance and academic achievement of Japanese, Chinese, and American children. *Child Development, 56,* 718–734.

Stewart, E. (2006). Family and individual-level predictors of academic success for African American students: A longitudinal path analysis using national data. *Journal of Black Studies 36,* 597–621.

Strand, S., & Deary, I.J. (2006). Sex differences in cognitive ability test scores: A UK national picture. *British Journal of Educational Psychology, 76,* 463–480.

Surber, C.F., & Haines, B.A. (1987). The growth of proportional reasoning: Methodological issues. In R. Vasta & G. Whitehurst (Eds.), *Annals of Child Development,* 4. Greenwich, CT: JAI Press.

Swanson, H.L. (1993). Working memory in learning disability subgroups. *Journal of Experimental Child Psychology, 56,* 87–114.

Swanson, H.L. (2006). Cognitive processes that underlie mathematical precociousness in young children. *Journal of Experimental Child Psychology, 93,* 239–264.

Swanson, H.L., & Sachse-Lee, C. (2001). Mathematical problem solving and working memory in children with learning disabilities: Both executive and phonological processes are important. *Journal of Experimental Child Psychology, 79,* 294–321.

Sweller, J. (1989). Cognitive technology: Some procedures for facilitating learning and problem solving in mathematics and science. *Journal of Educational Psychology, 81,* 457–466.

Sweller, J., & Cooper, G.A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction, 2,* 59–89.

Sweller, J., Mawer, R.F., & Ward, M.R. (1983). Development of expertise in mathematical problem solving. *Journal of Experimental Psychology*: *General, 112,* 639–661.

Swiatek, M.A., & Benbow, C.P. (1991a). A 10-year longitudinal follow-up of ability-matched accelerated and unaccelerated gifted students. *Journal of Educational Psychology, 83*, 528–538.

Swiatek, M., & Benbow, C.P. (1991b). A 10-year longitudinal follow-up of participation in a fast-paced mathematics course. *Journal for Research in Mathematics Education, 22*, 138–150.

Swiatek, M.A., & Benbow, C.P. (1992). Nonacademic correlates of satisfaction with accelerative programs. *Journal of Youth and Adolescence, 21,* 699–723.

Tatsuoka, K.K., Corter, J.E., & Tatsuoka, C. (2004). Patterns of diagnosed mathematical content and process skills in TIMSS-R across a sample of 20 countries. *American Educational Research Journal, 41*, 901–926.

Temple, E., & Posner, M.I. (1998). Brain mechanisms of quantity are similar in 5-year old children and adults. *Proceedings of the National Academy of Sciences USA, 95,* 7836–7841.

Tenenbaum, H.R., & Ruck, M.D. (2007). Are teachers' expectations different for racial minority than for European American students? A meta-analysis. *Journal of Educational Psychology, 99*, 253–273.

Terlecki, M.S., Newcombe, N.S., & Little, M. (in press). Durable and generalized effects of spatial experience on mental rotation: Gender differences in growth patterns. *Applied Developmental Psychology.*

Thomas, J. (1999). Influences on mathematic learning among African American high school students. ERIC Report #430-806. Washington, DC: U.S. Department of Education.

Thompson, P.W., & Saldanha, L.A. (2003). Fractions and multiplicative reasoning. In J. Kilpatrick, G. Martin & D. Schifter (Eds.), *Research companion to the principles and standards for school mathematics* (pp. 95–114). Reston, VA: National Council of Teachers of Mathematics.

Thompson, R.F., Mayers, K.S., Robertson, R.T., & Patterson, C.J. (1970, April 10). Number coding in association cortex of the cat. *Science, 168*, 271–273.

Thornton, C.A. (1978). Emphasizing thinking strategies in basic fact instruction. *Journal for Research in Mathematics Education, 9*, 214–227.

Thornton, C.A. (1990). Solution strategies: Subtraction number facts. *Educational Studies in Mathematics, 21*, 241–263.

Trabasso, T., Riley, C.A., & Wilson, E.G. (1975). The representation of linear order and spatial strategies in reasoning: A developmental study. In R.J. Falmagne (Ed.), *Reasoning: Representation and process in children and adults* (pp. 201–229). Hillsdale, NJ: Erlbaum.

Tronsky, L.N. (2005). Strategy use, the development of automaticity, and working memory involvement in complex multiplication. *Memory & Cognition, 33*, 927–940.

Tucker, C., Zayco, R., Herman, K., Reinke, W., Trujillo, M., Carraway, K., et al. (2002). Teacher and child valuables as predictors of academic engagement among low-income African American children. *Psychology in Schools, 39*, 477–488.

Tulving, E. & Thomson, D.M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80*, 352–373.

Tuovinen, J.E., & Sweller, J. (1999). A comparison of cognitive load associated with discovery learning and worked examples. *Journal of Educational Psychology, 91*, 334–341.

Tversky, B. (2005). Visuospatial reasoning. In K.J. Holyoak & R.J. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 209–240). New York: Cambridge University Press.

Tversky, B. (2004). Visuospatial reasoning. In K.J. Holyoak & R.J. Morrison (Eds.), *The Cambridge Handbook of Thinking and Reasoning*. New York: Cambridge University Press.

Uesaka, Y., Manalo, E., & Ichikawa, S. (2007). What kinds of perceptions and daily learning behaviors promote students' use of diagrams in mathematics problem solving? *Learning and Instruction, 17*, 322–335.

Underwood, B.J. (1964). Degree of learning and the measurement of forgetting. *Journal of Verbal Learning and Verbal Behavior, 3,* 112–129.

U.S. Department of Education, NCES. (1990). *The state of mathematics achievement, NAEP's 1990 assessment of the nation and the trial assessment of the states* (composite report). Washington, DC: Author.

U.S. Department of Education, NCES. (2001). *Early childhood longitudinal study, kindergarten class of 1998–99: Base year public-use data files user's manual* (NCES 2001–029). Washington, DC: Author.

U.S. Department of Education, NCES. (2003). NAEP questions. Retrieved August 10, 2007, from http://nces.ed.gov/nationsreportcard/itmrls/.

U.S. Department of Education, NCES. (2005a). *NAEP long-term trend performance levels*. Retrieved April 17, 2007 from http://nces.ed.gov/nationsreportcard/ltt/performance-levels.asp.

U.S. Department of Education, NCES. (2005b). *NAEP 2004 trends in academic progress: three decades of student performances in reading and mathematics*. Retrieved April 17, 2007 from http://oces.ed.gov/pubsearch/pubsinfo.asp?pubid=2005464.

U.S. Department of Education, NCES. (2006). ECLS–K longitudinal kindergarten–fifth grade public-use data file and electronic codebook (CD-ROM). (NCES 2006-035) Washington, DC: Author.

U.S. Department of Education, NCES. (2007). The long-term trend performance levels. Retrieved on November 1, 2007 from http://nces.ed.gov/nationsreportcard/itt/performancelevels.asp.

Uttal, D.H. (2003). On the relation between play and symbolic thought: The case of mathematics manipulatives. In O. Saracho & B. Spodek (Eds.), *Contemporary perspectives on play in early childhood education.* Charlotte, NC: Information Age Press.

Vamvakoussi, X., & Vosniadou, S. (2004). Understanding the structure of the set of rational numbers: A conceptual change approach. *Learning and Instruction, 14*, 453–467.

Van Dooren, W., De Bock, D., Depaepe, F., Janssens, D., & Verschaffel, L. (2003). The illusion of linearity: Expanding the evidence towards probabilistic reasoning. *Educational Studies in Mathematics, 53*, 113–138.

Van Dooren, W., De Bock, D., Hessels, A., Janssens, D., & Verschaffel, L. (2005). Not everything is proportional: Effects of age and problem type on propensities for overgeneralization. *Cognition and Instruction*, *23*, 57–86.

Van Dooren, W., De Bock, D., Janssens, D., & Verschaffel, L. (2005). Students' overreliance on linearity: An effect of school-like word problems? In Chick, H.L. & Vincent, J.L. (Eds.), *Proceedings of the 29th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 265–272). Melbourne, PME.

van Hiele, P.M. (1986). *Structure and insight*. Orlando, FL: Academic Press.

VanLehn, K. (1990). *Mind bugs: The origins of procedural misconceptions*. Cambridge, MA: MIT Press.

Varelas, M., & Becker, J. (1997). Children's developing understanding of place value: Semiotic aspects. *Cognition and Instruction, 15,* 265–286.

Vilette, B. (2002). Do young children grasp the inverse relationship between addition and subtraction? Evidence against early arithmetic. *Cognitive Development*, *17*, 1365–1383.

Vuilleumier, P., Ortigue, S., & Brugger, P. (2004). The number space and neglect. *Cortex*, *40*, 399–410.

Walberg, H. (1984). Improving the productivity of America's schools. *Educational Leadership* May, 19–27.

Walker, E.L. (1980). *Psychological complexity and preference: A hedgehog theory of behavior*. Monterey, CA: Brooks/Cole Pub. Company.

Weiner, B. (1990). History of motivational research in education. *Journal of Educational Psychology, 82*, 616–622.

Weiss, B.D., Mays, M.Z., Martz, W., Castro, K.M., DeWalt, D.A., Pignone, M.P., et al. (2005). Quick assessment of literacy in primary care: The newest vital sign. *Annals of Family Medicine, 3*, 514–522.

Wenger, R.H. (1987). Cognitive science and algebra learning. In A.H. Schoenfeld (Ed.), *Cognitive science and mathematics education* (pp. 217–251). Hillsdale, NJ: Lawrence Erlbaum Associates.

Wentzel, K.R. (1991). Relations between social competence and academic achievement in early adolescence. *Child Development*, *62*, 1066–1078.

Wertheimer, M. (1959). *Productive thinking*. Oxford, England: Harper.

Wertsch, J. (1985). *Vygotsky and the social formation of mind.* Cambridge, MA: Harvard University Press.

Wolfe, C.J., Reyna, V.F., & Brainerd, C.J. (2005). Fuzzy-trace theory: Implications for transfer in teaching and learning. In J.P. Mestre (Ed.), *Transfer of learning: From a modern interdisciplinary perspective* (pp. 53–88). Greenwich, CT: Information Age Publishing.

Woloshin, S., Schwartz, L., Byram, S., Fischhoff, B., & Welch, H. (2000). A new scale for measuring perceptions of chance: A validation study. *Medical Decision Making, 20*(3), 298–307.

Wolters, C. (2004). Advancing achievement goal theory: Using goal structure and goal orientations to predict students' motivation, cognition and achievement. *Journal of Educational Psychology, 96*, 236–250.

Wu, H. (April, 2005). *Key Mathematical Ideas in Grades 5–8*. Presented at the annual meeting of the National Council of Teachers of Mathematics, Anaheim.

Wu, H. (2006). How mathematicians can contribute to K–12 mathematics education. In M.S. Sole (Ed.), *Proceedings of the International Congress of Mathematics, Madrid 2006, vol. 3* (pp.1676–1688). Zurich, Switzerland: European Mathematical Society.

Wu, H. (2007). *Fractions, decimals, and rational numbers.* Available http://math.berkeley.edu/~wu/ (March 19, 2008).

Wynn, K. (1992). Addition and subtraction by human infants. *Nature, 358*, 749–750.

Xu, F., & Spelke, E.S. (2000). Large number discrimination in 6-month old infants. *Cognition, 74*, B1–B11.

Young, R.M., & O'Shea, T. (1981). Errors in children's subtraction. *Cognitive Science, 5*, 153–177.

Zhu, X., & Simon, H.A. (1987). Learning mathematics from examples and by doing. *Cognition and Instruction, 4*, 137–166.

Zorzi, M., Priftis, K., & Umiltá, C. (2002, May 9). Neglect disrupts the mental number line. *Nature, 417*, 138.

APPENDIX A: Literature Search Guidelines

The goal of the literature search was to identify experimental, cognitive, or related studies of children's mathematics learning in specific content areas (see key words). These involve measures of children's learning, problem solving, or understanding that are more precisely defined (e.g., trial-by-trial assessment of problem solving strategy) than is typically found with psychometric measures (e.g., achievement tests).

First search. This covered a designated set of core learning, cognition, and developmental journals: *American Educational Research Journal; Child Development; Cognition; Cognition and Instruction; Cognitive Development; Cognitive Psychology; Cognitive Science; Current Directions in Psychological Sciences; Developmental Psychology; Developmental Review; Journal of Cognition and Development; Journal of Education Psychology; Journal of Experimental Child Psychology; Journal of Experimental Psychology, Learning, Memory and Cognition; Journal of Experimental Psychology: General; Journal of Memory and Language; Journal of Personality and Social Psychology; Learning and Individual Differences; Mathematical Cognition; Memory and Cognition; Nature; Psychological Bulletin; Psychological Review; Psychological Science; Review of Educational Research; Science.*

Second search. This covered other English-language, peer-reviewed journals that primarily publish empirical studies and are indexed in PsychInfo and Web of Science (Social Sciences Citation Index).

Criteria for Inclusion

- Published in English.
- Participants are age 3 years to young adult.
- Published in a peer-reviewed empirical journal, or a review of empirical research in books or annual reviews.
- Experimental, quasi-experimental, or correlational methods.

APPENDIX B: Search Terms

Search Terms Used in Literature Review

Five Core topics:

Learning and Cognition of:

whole number arithmetic
fractions
estimation
algebra
geometry

Specific Key Terms:

arithmetic
addition
subtraction
multiplication
division
base-10
fraction
number
number line

commutativity
associativity
place-value
perimeter
area
volume
linear equations
function

mathematical equality
mathematical inequality
ratio
equation
number sense
ordinal
cardinal

variable
set
numerosity
zero
proportion
proportional reasoning
number comparison
exponents
radical

arithmetic word problems
algebra word problems
fractions

algorithm
counting
distributive property
estimation
integers
magnitude comparison
math anxiety
mental arithmetic
natural numbers
numeracy
part-whole relationships
problem-size effect
rational numbers
real numbers
regrouping
subitizing
transcoding

math LD (learning disability)
arithmetic LD
dyscalculia

math race
math ethnicity
math sex
math gender
math socioeconomic status
math sociocultural background
math gifted

APPENDIX C: Sex Differences

The following tables and figures summarize the data on math performance by gender using data available on national samples. Data from the Trends in Math and Science Survey (TIMSS) illustrate the math performance of fourth- and eighth-graders. Data from the National Assessment of Educational Progress (NAEP) Long-Term Trend study illustrate performance between groups over the last 30 years. Data from the High School and Beyond (HS&B:80), National Education Longitudinal Study of 1988 (NELS:88), and Education Longitudinal Study of 2002 (ELS:2002) illustrate the math performance of 10th-grade students. Data from the National Adult Literacy Survey (NALS) and the National Assessment of Adult Literacy (NAAL) survey illustrate the quantitative literacy of adults. Data from the Program for International Student Assessment (PISA) illustrate the mathematics literacy and problem-solving proficiency of 15-year-olds. To facilitate the interpretation of the various scores, a description of the test benchmarks and performance levels associated with each test is provided.

National Assessment of Educational Progress Long-Term Trends: Mathematics Scores

This section presents the long-term trends in NAEP mathematics scores. The goal is to describe the differences in performance between groups over the last 30 years and to describe how their scores have evolved over time. For each reporting group, results are presented in the form of the average scale score for intermittent years from 1978 to 2004 and the percent of students at each achievement level in 1978, 1999, and 2004.

Methodology

All data presented in this section were obtained from the NAEP Data Explorer.[2] The Data Explorer allows users to create tables of results by custom combinations of reporting variables. The results can be reported in terms of mean score, percentage of students at or above performance levels, and score percentile.

The Data Explorer also reports standard errors and can calculate the statistical significance of changes in a variable between years or between variables in the same year. The statistical significance of changes between variables over time (e.g., the score difference between girls and boys in 1978 versus the score difference between girls and boys in 2004) is taken either directly from the *NAEP 2004 Trends in Academic Progress* or estimated using the reported standard error provided by the Data Explorer. Only differences that are statistically significant beyond the 0.05 level are described in the text of this section.

[2] http://nces.ed.gov/nationsreportcard/naepdata.

Average Scale Scores and Performance Levels

The NAEP long-term trend assessments are scored on a 0–500 point scale, but all average scale score charts presented here are ranged from 180–340 for consistency and best visibility of score differences. Charts of average scale scores are reconstructed to resemble the gap charts in *NAEP 2004 Trends in Academic Progress*.

The following text was taken verbatim from the National Center for Education Statistics website, http://nces.ed.gov/nationsreportcard/ltt/performance-levels.asp in April 2007.

> More detailed information about what students know and can do in each subject area can be gained by examining their attainment of specific performance levels in each assessment year. This process of developing the performance-level descriptions is different from that used to develop achievement-level descriptions in the main NAEP reports.
>
> For each of the subject area scales, performance levels were set at 50-point increments from 150 through 350. The five performance levels—150, 200, 250, 300, and 350—were then described in terms of the knowledge and skills likely to be demonstrated by students who reached each level.
>
> A “scale anchoring” process was used to define what it means to score in each of these levels. NAEP’s scale anchoring follows an empirical procedure whereby the scaled assessment results are analyzed to delineate sets of questions that discriminate between adjacent performance levels on the scales. To develop these descriptions, assessment questions were identified that students at a particular performance level were more likely to answer successfully than students at lower levels. The descriptions of what students know and can do at each level are based on these sets of questions.
>
> The guidelines used to select the questions were as follows: Students at a given level must have at least a specified probability of success with the questions (75 % for mathematics, 80 % for reading), while students at the next lower level have a much lower probability of success (that is, the difference in probabilities between adjacent levels must exceed 30 percent). For each curriculum area, subject-matter specialists examined these empirically selected question sets and used their professional judgment to characterize each level. The scale anchoring for mathematics trend reporting was based on the 1986 assessment.

The five performance levels are applicable at all three age groups, but only three performance levels are discussed for each age: levels 150, 200, and 250 for age 9; levels 200, 250, and 300 for age 13; and levels 250, 300, and 350 for age 17. These performance levels are the ones most likely to show significant change within an age across the assessment years and do not include the levels that nearly all or almost *no students attained at a particular age in each year.*

The following description of each mathematics performance level was copied from http://nces.ed.gov/nationsreportcard/ltt/math-descriptions.asp in April 2007.

Level 350: Multistep Problem Solving and Algebra

Students at this level can apply a range of reasoning skills to solve multistep problems. They can solve routine problems involving fractions and percents, recognize properties of basic geometric figures, and work with exponents and square roots. They can solve a variety of two-step problems using variables, identify equivalent algebraic expressions, and solve linear equations and inequalities. They are developing an understanding of functions and coordinate systems.

Level 300: Moderately Complex Procedures and Reasoning

Students at this level are developing an understanding of number systems. They can compute with decimals, simple fractions, and commonly encountered percents. They can identify geometric figures, measure lengths and angles, and calculate areas of rectangles. These students are also able to interpret simple inequalities, evaluate formulas, and solve simple linear equations. They can find averages, make decisions based on information drawn from graphs, and use logical reasoning to solve problems. They are developing the skills to operate with signed numbers, exponents, and square roots.

Level 250: Numerical Operations and Beginning Problem Solving

Students at this level have an initial understanding of the four basic operations. They are able to apply whole number addition and subtraction skills to one-step word problems and money situations. In multiplication, they can find the product of a two-digit and a one-digit number. They can also compare information from graphs and charts, and are developing an ability to analyze simple logical relations.

Level 200: Beginning Skills and Understandings

Students at this level have considerable understanding of two-digit numbers. They can add two-digit numbers but are still developing an ability to regroup in subtraction. They know some basic multiplication and division facts, recognize relations among coins, can read information from charts and graphs, and use simple measurement instruments. They are developing some reasoning skills.

Level 150: Simple Arithmetic Facts

Students at this level know some basic addition and subtraction facts, and most can add two-digit numbers without regrouping. They recognize simple situations in which addition and subtraction apply. They also are developing rudimentary classification skills.

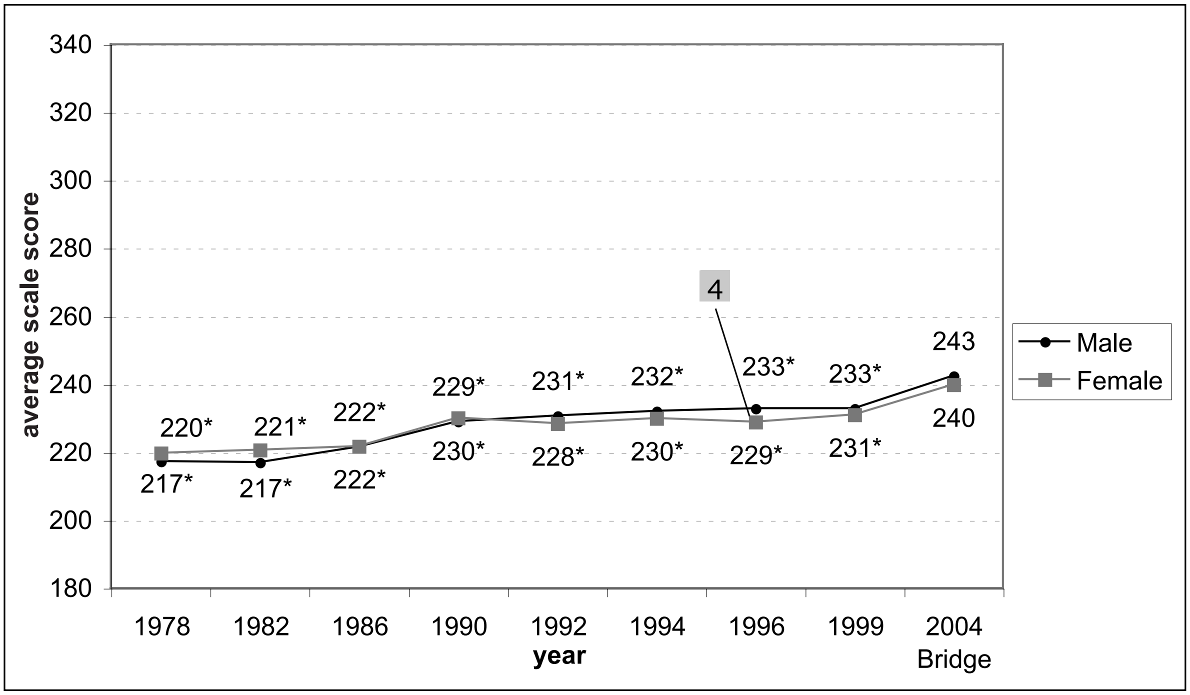
Table C-1: Number of Students in Each NAEP Reporting Group, by Age, Gender, Race/Ethnicity, and Parents' Level of Education: 1978, 1999, and 2004

	Age 9			Age 13			Age 17		
Reporting Group/Year	**1978**	**1999**	**2004**	**1978**	**1999**	**2004**	**1978**	**1999**	**2004**
Total	14800	6000	5200	24200	5900	5700	26800	3800	3800
Male	7400	2940	2548	12100	2950	2736	13132	1824	1824
Female	7400	3060	2652	12100	2950	2964	13668	1976	1976
White	11692	4200	3068	19360	4189	3648	22244	2736	2584
Black	2072	1080	728	3146	885	798	3216	570	456
Hispanic	740	480	988	1452	590	912	1072	380	532
Other			416			342			228
Parents' Level of Education									
Less than high school				2904	354	399	3484	266	342
Graduated high school				7986	1239	1083	8844	76	722
Some education after high school				3388	1003	855	4288	874	836
Graduated college				6292	2832	2679	8576	1824	1786
Unknown				3630	531	684	1340	1140	114

Source: Created by the Institute for Defense Analysis Science and Technology Policy Institute using the NAEP Data Explorer (http://nces.ed.gov/nationsreportcard/naepdata/).

Note: Level of education is parents' level of education and was not collected for 9-year-olds.

Figure C-1: Average NAEP Scale Scores by Gender, Age 9: Intermittent Years From 1978–2004



*Indicates score or gap is significantly different from 2004.
+ 1996 was an exception to general trend of no gender gap in scores at age 9.

Note: Data labels for male (above) and female (below). Between gender score differences (gaps) are shown in shaded boxes only for years in which the gap is statistically significant. Labeled gaps may not reflect labeled scores because of rounding.

“Bridge” refers to updates made to NAEP in 2004. The updates replaced outdated material and accommodated more students with disabilities. In order to maintain the long-term trend, test takers were randomly assigned to either the old test form, called the bridge assessment, or the modified test form. Results from the bridge assessment should be compared to results from assessments prior to 2004, while results from the modified assessment should be compared to assessments given after 2004.

Source: Created by the Institute for Defense Analysis Science and Technology Policy Institute using the NAEP Data Explorer (http://nces.ed.gov/nationsreportcard/naepdata/).

What Is This Indicator?

This indicator presents the average scale score for 9-year-old boys and girls for each assessment since 1978.

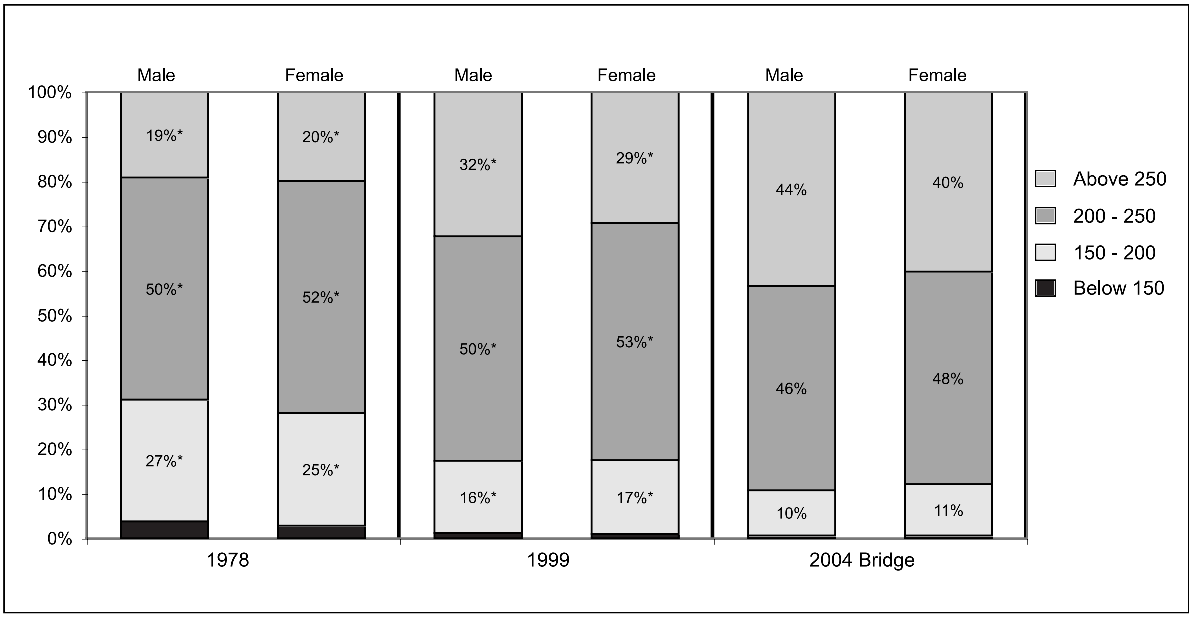
Discussion

- In 2004, the average score for both boys and girls was higher than in any previous assessment.
  - — The average score for 9-year-old boys increased by 10 points between 1999 and 2004, going from 233 in 1999 to 243 in 2004. The average score for boys in 2004 was a 23 point increase from the average score of 220 in 1978.

— The average score for 9-year-old girls increased by 9 points between 1999 and 2004, going from 231 in 1999 to 240 in 2004. The average score for girls in 2004 was a 23-point increase from the average score of 217 in 1978.

- In general, there was no gender gap at age 9. The difference in average score for 9-year-old boys and 9-year-old girls has not been significant in most years.
  — The one exception is 1996, when boys scored 4 points higher than girls on average.

Figure C-2: Percent at NAEP Performance Levels by Gender, 9-Year-Olds: 1978, 1999, and 2004



*Indicates percentage is significantly different from 2004. Differences between categories (genders) may not be statistically significant.

Note: “Bridge” refers to updates made to NAEP in 2004. The updates replaced outdated material and accommodated more students with disabilities. In order to maintain the long-term trend, test takers were randomly assigned to either the old test form, called the bridge assessment, or the modified test form. Results from the bridge assessment should be compared to results from assessments prior to 2004, while results from the modified assessment should be compared to assessments given after 2004.

Source: Created by the Institute for Defense Analysis Science and Technology Policy Institute using the NAEP Data Explorer (http://nces.ed.gov/nationsreportcard/naepdata/).

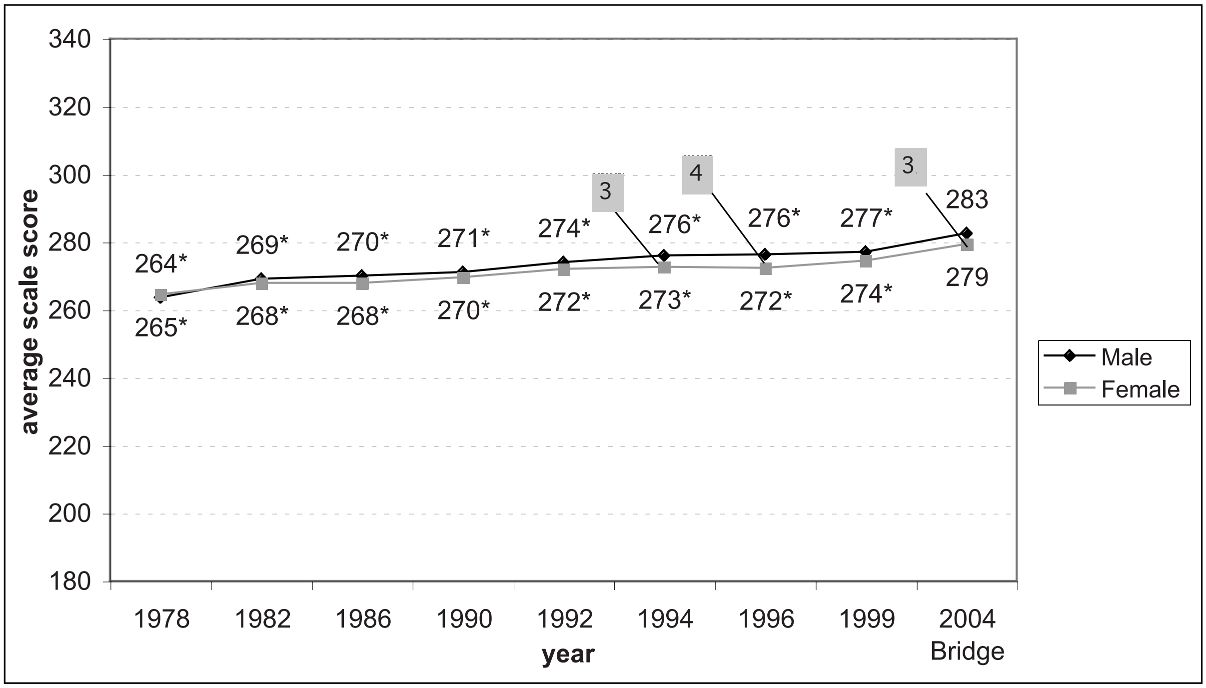
What Is This Indicator?

This indicator presents the percentage of 9-year-olds reaching each performance level by gender. The performance levels reported at age 9 are 150—Simple Arithmetic Facts, 200—Beginning Skills and Understandings, and 250—Numerical Operations and Beginning Problem Solving.

Discussion

- 9-year-old boys and girls reach similar performance levels. The differences in the percent of 9-year-old boys and 9-year-old girls reaching each performance level are not significant.
- The percentage of 9-year-olds reaching the highest achievement level for this age group, at or above 250, has doubled since 1978 and has increased from approximately 30% to 40% between 1999 and 2004.

Figure C-3: Average NAEP Scale Scores by Gender, Age 13: Intermittent Years From 1978–2004



*Indicates score or gap is significantly different from 2004.

Note: Data labels for male (above) and female (below). Between gender score differences (gaps) are shown in shaded boxes only for years in which the gap is statistically significant. Labeled gaps may not reflect labeled scores because of rounding.

“Bridge” refers to updates made to NAEP in 2004. The updates replaced outdated material and accommodated more students with disabilities. In order to maintain the long-term trend, test takers were randomly assigned to either the old test form, called the bridge assessment, or the modified test form. Results from the bridge assessment should be compared to results from assessments prior to 2004, while results from the modified assessment should be compared to assessments given after 2004.

Source: Created by the Institute for Defense Analysis Science and Technology Policy Institute using the NAEP Data Explorer (http://nces.ed.gov/nationsreportcard/naepdata/).

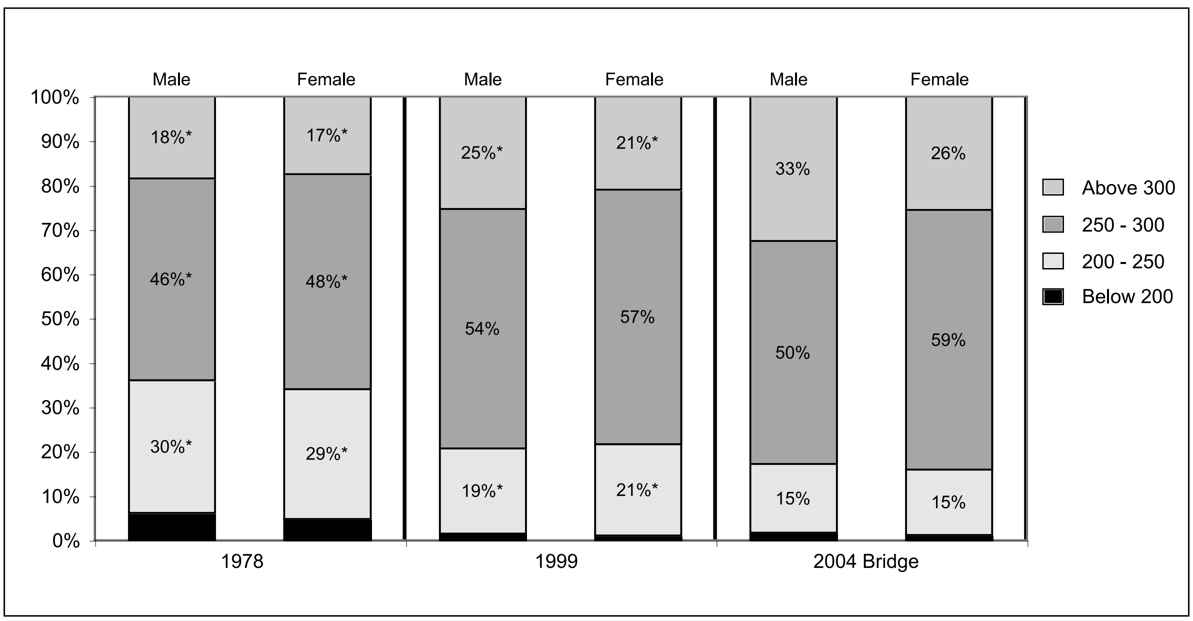
What Is This Indicator?

This indicator presents the average scale score for 13-year-old boys and girls for each assessment since 1978.

Discussion

- In 2004, the average score for both 13-year-old boys and 13-year-old girls was higher than in any previous assessment.
  - — The average score for 13-year-old boys increased by 6 points between 1999 and 2004, going from 277 in 1999 to 283 in 2004. The average score for boys in 2004 was a 19 point increase from the average score of 264 in 1978.
  - — The average score for 13-year-old girls increased by 5 points between 1999 and 2004, going from 274 in 1999 to 279 in 2004. The average score for girls in 2004 was a 14 point increase from the average score of 265 in 1978.
- In general, there was no consistent gender gap at age 13. The difference in average score for 13-year-old boys and 13-year-old girls has not been significant in most years.
  - — In 2004, 1996, and 1994 the average score for boys was 3 to 4 points higher than the average score for girls.

Figure C-4: Percent at NAEP Performance Levels by Gender, 13-Year-Olds: 1978, 1999, and 2004



*Indicates percentage is significantly different from 2004. Differences between categories (genders) may not be statistically significant.

Note: "Bridge" refers to updates made to NAEP in 2004. The updates replaced outdated material and accommodated more students with disabilities. In order to maintain the long-term trend, test takers were randomly assigned to either the old test form, called the bridge assessment, or the modified test form. Results from the bridge assessment should be compared to results from assessments prior to 2004, while results from the modified assessment should be compared to assessments given after 2004.

Source: Created by the Institute for Defense Analysis Science and Technology Policy Institute using the NAEP Data Explorer (http://nces.ed.gov/nationsreportcard/naepdata/).

What Is This Indicator?

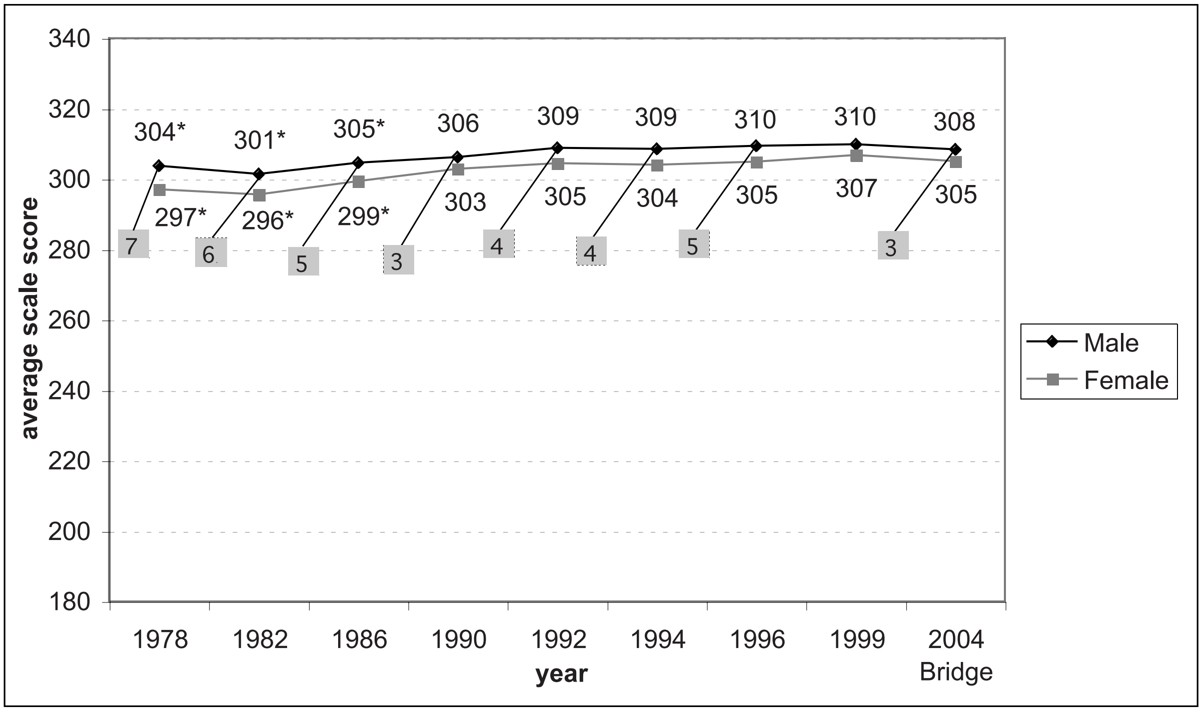
This indicator presents the percentage of 13-year-olds reaching each performance level by gender in 1978, 1999, and 2004. The performance levels reported at age 13 are 200—Beginning Skills and Understandings, 250—Numerical Operations and Beginning Problem Solving, and 300—Moderately Complex Procedures and Reasoning.

Discussion

- In 1999 and 2004, slightly more 13-year-old boys scored at or above 300 than did 13-year-old girls.
    - — In 1999 the gender gap at the 300 level was 4%, with 25% of boys and 21% of girls performing at or above 300.
    - — In 2004 the gender gap at the 300 level was 7%, with 33% of boys and 26% of girls performing at or above 300.
    - — The change in gender gap from 1999 to 2004 was not statistically significant.

- The percentages of boys and girls scoring at or above 300 have increased since 1999 and 1978.
    - — The percentage of 13-year-old boys at or above 300 was 33% in 2004, which was 7% higher than in 1999 and 14% higher than in 1978.
    - — The percentage of 13-year-old girls at of above 300 was 26% in 2004, which was 5% higher than in 1999 and 8% higher than in 1978.

Figure C-5: Average NAEP Scale Scores by Gender, Age 17: Intermittent Years From 1978–2004



*Indicates score or gap is significantly different from 2004.

Note: Data labels for male (above) and female (below). Between gender score differences (gaps) are shown in shaded boxes only for years in which the gap is statistically significant. Labeled gaps may not reflect labeled scores because of rounding.

"Bridge" refers to updates made to NAEP in 2004. The updates replaced outdated material and accommodated more students with disabilities. In order to maintain the long-term trend, test takers were randomly assigned to either the old test form, called the bridge assessment, or the modified test form. Results from the bridge assessment should be compared to results from assessments prior to 2004, while results from the modified assessment should be compared to assessments given after 2004.

Source: Created by the Institute for Defense Analysis Science and Technology Policy Institute using the NAEP Data Explorer (http://nces.ed.gov/nationsreportcard/naepdata/).

What Is This Indicator?

This indicator presents the average scale score for 17-year-old boys and girls for each assessment since 1978.

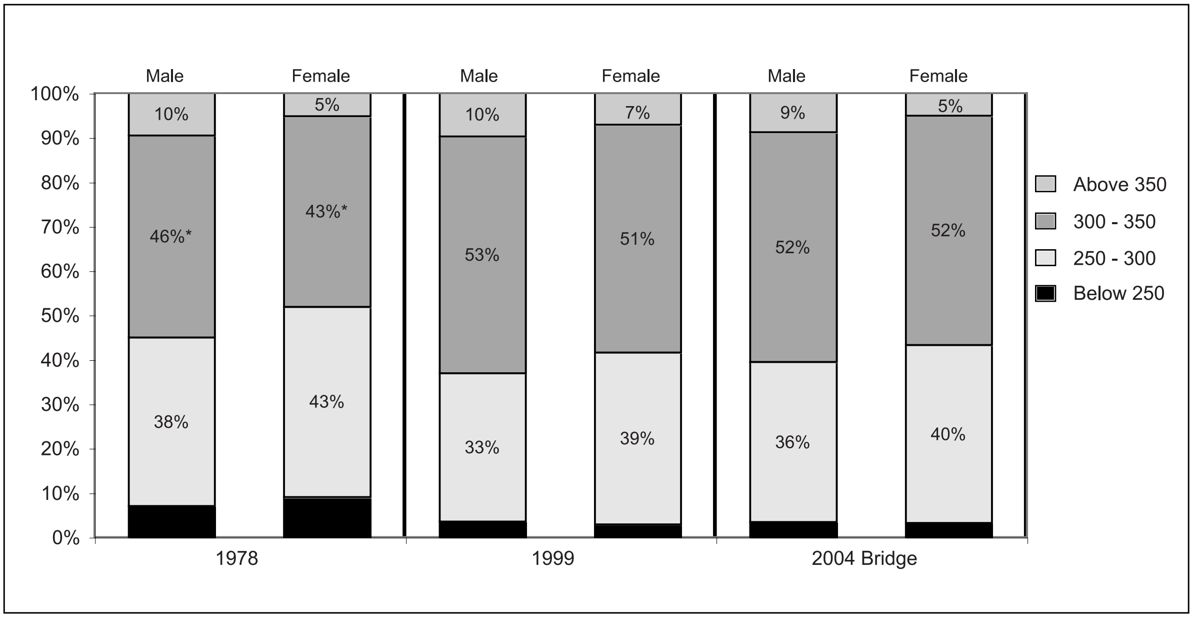
Discussion

- The average score for both girls and boys at age 17 has been flat since 1990, although average scores have increased slightly since 1978.
  - — The average score for 17-year-old boys increased by four points from 304 in 1978 to 308 in 2004.

— The average score for 17-year-old girls increased by eight points from 297 in 1978 to 205 in 2004.

- 17-year-old boys have consistently outscored 17-year-old girls on the long-term mathematics NAEP.
  — The gender gap for 17-year-olds in 2004 was three points and was not significantly different from previous years.

Figure C-6: Percent at NAEP Performance Levels by Gender, 17-Year-Olds: 1978, 1999, and 2004



*Indicates percentage is significantly different from 2004. Differences between categories (genders) may not be statistically significant.

Note: "Bridge" refers to updates made to NAEP in 2004. The updates replaced outdated material and accommodated more students with disabilities. In order to maintain the long-term trend, test takers were randomly assigned to either the old test form, called the bridge assessment, or the modified test form. Results from the bridge assessment should be compared to results from assessments prior to 2004, while results from the modified assessment should be compared to assessments given after 2004.

Source: Created by the Institute for Defense Analysis Science and Technology Policy Institute using the NAEP Data Explorer (http://nces.ed.gov/nationsreportcard/naepdata/).

What Is This Indicator?

This indicator presents the percentage of 17-year-olds reaching each performance level by gender.

The performance levels reported at age 17 are 250—Numerical Operations and Beginning Problem Solving, 300—Moderately Complex Procedures and Reasoning, and 350—Multistep Problem Solving and Algebra.

Discussion

- In 1978 and 2004, slightly more 17-year-old boys scored at or above 350 than did 17-year-old girls, but in 1999 gender differences were not significant.
  - — The gender gap in 1978 was 5%; 10% of boys and 5% of girls scored at or above 350.
  - — The gender gap in 2004 was 4%; 9% of boys and 5% of girls scored at or above 350.
- The percentages of 17-year-old boys and girls at each performance level have, for the most part, not changed significantly between assessments.
  - — The percentage of both girls and boys scoring at the 300 level was higher in 2004 than in 1978. The percentage of boys at the 300 level increased by 9% to 52%, and the percentage of girls at the 300 level increased by 12% to 52%.

Trends in Math and Science Survey: TIMSS

The TIMSS 2003 International Benchmarks of Mathematics Achievement are defined in Mullis et al. (2004, p. 63) as follows.

Grade 8

Advanced International Benchmark – 625

Students can organize information, make generalizations, solve non-routine problems, and draw and justify conclusions from data. They can compute percent change and apply their knowledge of numeric and algebraic concepts and relationships to solve problems. Students can solve simultaneous linear equations and model simple situations algebraically. They can apply their knowledge of measurement and geometry in complex problem situations. They can interpret data from a variety of tables and graphs, including interpolation and extrapolation.

High International Benchmark – 550

Students can apply their understanding and knowledge in a wide variety of relatively complex situations. They can order, relate, and compute fractions and decimals to solve word problems, operate with negative integers, and solve multi-step word problems involving proportions with whole numbers. Students can solve simple algebraic problems including evaluating expressions, solving simultaneous linear equations, and using a formula to determine the value of a variable. Students can find areas and volumes of simple geometric shapes and use knowledge of geometric properties to solve problems. They can solve probability problems and interpret data in a variety of graphs and tables.

Intermediate International Benchmark – 475

Students can apply basic mathematical knowledge in straightforward situations. They can add, subtract, or multiply to solve one-step word problems involving whole numbers and decimals. They can identify representations of common fractions and relative sizes of fractions. They understand simple algebraic relationships and solve linear equations with one variable. They demonstrate understanding of properties of triangles and basic geometric concepts including symmetry and rotation. They recognize basic notions of probability. They can read and interpret graphs, tables, maps, and scales.

Low International Benchmark – 400

Students have some basic mathematical knowledge. (Mullis et al., 2004, p. 62)

Grade 4

Advanced International Benchmark – 625

Students can apply their understanding and knowledge in a wide variety of relatively complex situations. They demonstrate a developing understanding of fractions and decimals, and the relationship between them. They can select appropriate information to solve multi-step word problems involving proportions. They can formulate or select a rule for a relationship. They show understanding of area and can use measurement concepts to solve a variety of problems. They show some understanding of rotation. They can organize, interpret, and represent data to solve problems.

High International Benchmark – 550

Student can apply their knowledge and understanding to solve problems. Students can solve multi-step word problems involving addition, multiplication, and division. They can use their understanding of place value and simple fractions to solve problems. They can identify a number sentence that represents situations. Students show understanding of three-dimensional objects, how shapes can make other shapes, and simple transformation in a plane. They demonstrate a variety of measurement skills and can interpret and use data in tables and graphs to solve problems.

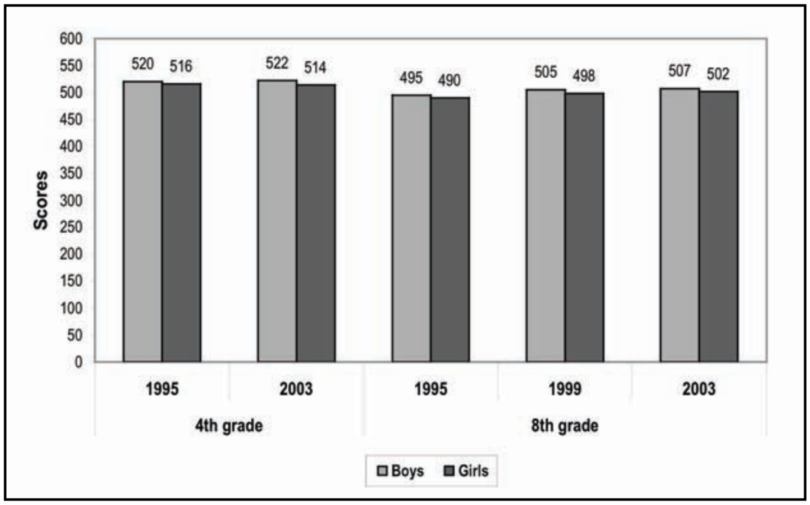
Intermediate International Benchmark – 475

Students can apply basic mathematical knowledge in straightforward situations. They can read, interpret, and use different representations of numbers. They can perform operations with three- and four-digit numbers and decimals. They can extend simple patterns. They are familiar with a range of two-dimensional shapes and read and interpret different representations of the same data.

Low International Benchmark – 400

Students have some basic mathematical knowledge. Students demonstrate an understanding of whole numbers and can do simple computations with them. They demonstrate familiarity with the basic properties of triangles and rectangles. They can read information from simple bar graphs.

Figure C-7: Average TIMSS Mathematical Scale Scores of U.S. 4th- and 8th-Graders, by Sex: Various Years From 1995–2003

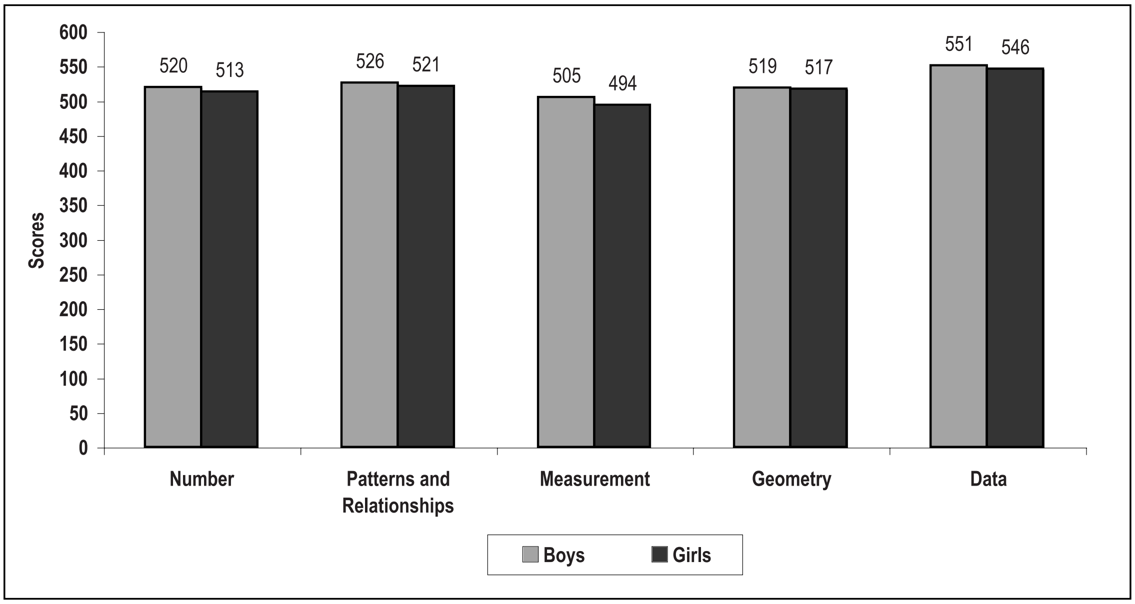


Note: TIMSS international benchmarks: Low 400, Intermediate 475, High 550, Advanced 625

Source: Gonzales et al. (2004), Figures 1 and 2.

Standardized mean difference TIMSS, gender					
	4th grade		**8th grade**		
	1995	*2003*	*1995*	*1999*	*2003*
Boys-Girls	0.05	0.11	0.06	0.08	0.06

Figure C-8: Average TIMSS Mathematical Scale Scores of U.S. 4th-Graders, by Sex, by Content Area: 2003

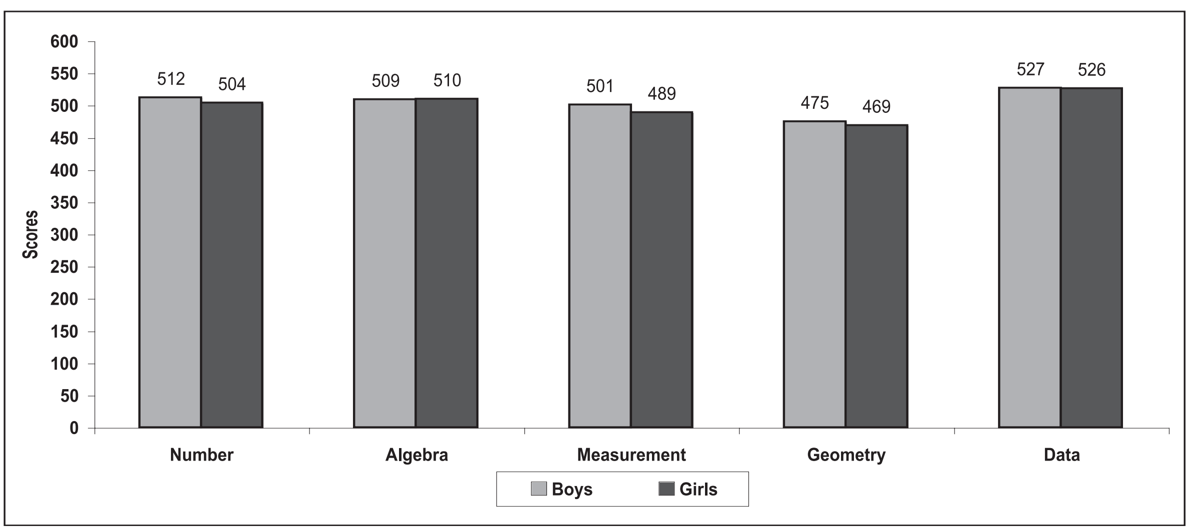


Note: TIMSS international benchmarks: Low 400, Intermediate 475, High 550, Advanced 625

Source: Mullis et al. (2003), Exhibit 3.3.

Standardized mean difference TIMSS, content areas 4th grade					
	Number	**Patterns and Relationships**	**Measurement**	**Geometry**	**Data**
Boys-Girls	0.09	0.06	0.14	0.02	0.06

Figure C-9: Average TIMSS Mathematical Scale Scores of U.S. 8th-Graders, by Sex, by Content Area: 2003



Note: TIMSS international benchmarks: Low 400, Intermediate 475, High 550, Advanced 625

Source: Mullis et al. (2003), Exhibit 3.3.

High School and Beyond of 1980: HS&B:80
National Education Longitudinal Study of 1988: NELS:88
Education Longitudinal Study of 2002: ELS:2002

The scores on the HS&B:80, NELS:88, and ELS:2002 are Item Response Theory (IRT) number-right scores on the NELS:88 1990 58-item scale. IRT estimates achievement based on patterns of correct, incorrect, and unanswered questions. "The IRT-estimated number-right score reflects an estimate of the number of these 58 items that an examinee would have answered correctly if he or she had taken all of the items that appeared on the multiform 1990 NELS:88 mathematics test. The score is the probability of a correct answer on each item, summed over the total mathematics 58-item pool" (Cahalan, Ingels, Burns, Planty, & Daniel, 2006, p.45). These scores are not directly translated into probability-of-proficiency scores. However, five probability-of-proficiency scores in mathematics were estimated for students using performance on clusters of four items each as follows:

Probability of Mastery, Mathematics Levels

1) Simple arithmetical operations on whole numbers, such as simple arithmetic expressions involving multiplication or division of integers;

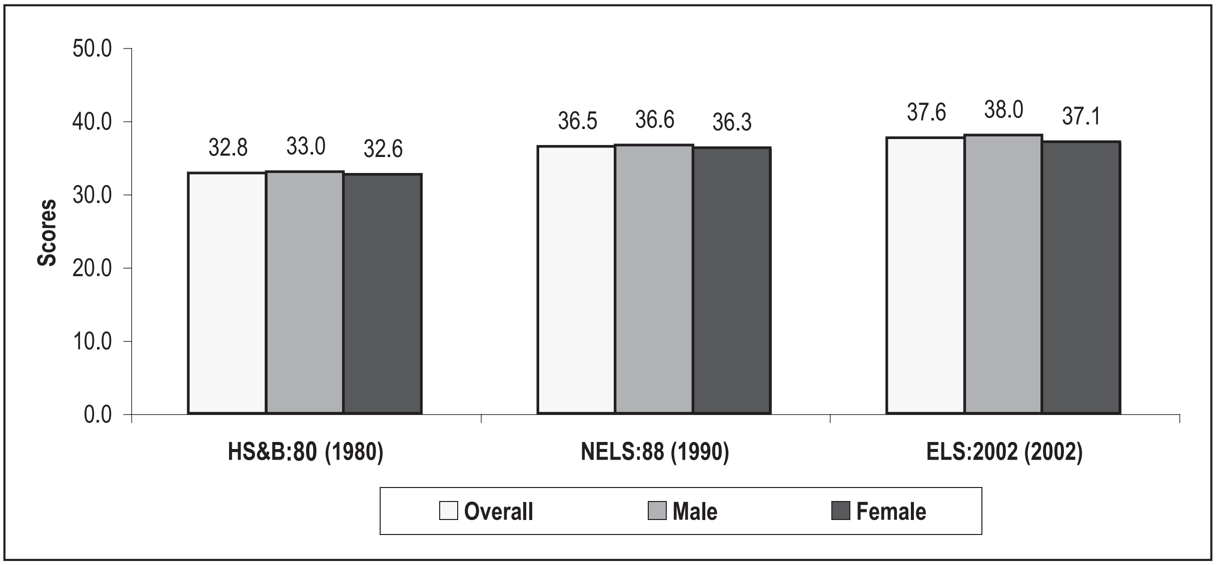
2) Simple operations with decimals, fractions, powers, and roots, such as comparing expressions, given information about exponents;

3) Simple problem solving, requiring the understanding of low-level mathematical concepts, such as simplifying an algebraic expression or comparing the length of line segments illustrated in a diagram;

4) Understanding of intermediate-level mathematical concepts and/or multistep solutions to word problems such as drawing an inference based on an algebraic expression or inequality; and

5) Complex multistep word problems and/or advanced mathematics material such as a two-step problem requiring evaluation of functions. (Cahalan et al., 2006, p. A-28)

Figure C-10: IRT—Estimated Average Math Score (10th-Grade), by Sex (HS&B:80, NELS:88, ELS:2002)



Note: IRT scale score is the estimated number right out of a total of 58.

Source: Cahalan et al. (2006), Tables 18 and 19.

Standardized mean difference sophomores, gender			
	HS&B (1980)	**NELS:88 (1990)**	**ELS:2002 (2002)**
Male-Female	0.03	0.02	0.08

Table C-2: Probability of 10th-Grade Proficiency in Mathematics, by Gender

	NELS:88 (1990)	**ELS:2002 (2002)**
Level 1		
Male	90.7	91.7
Female	90.8	91.6
Level 2		
Male	62.8	68.4
Female	63.3	65.7
Level 3		
Male	44.3	48.0
Female	42.8	44.7
Level 4		
Male	20.2	22.3
Female	17.8	18.5
Level 5		
Male	0.5	1.3
Female	0.3	0.6

Note: Proficiency levels – 1) Simple arithmetical operations with whole numbers; 2) Simple operations with decimals, fractions, powers, and roots; 3) Simple problem solving, requiring the understanding of low-level mathematical concepts; 4) Understanding of intermediate-level mathematical concepts and/or multistep solutions to word problems; and 5) Complex multistep word problems and/or advanced mathematics material.

Source: Cahalan et al., 2006, p. 57.

Program for International Student Assessment: PISA

Mathematics literacy can be classified by proficiency levels, based on scores on the PISA, as follows:

Below level 1 (*less than or equal to 357.77*)

Level 1 (*greater than 357.77 to 420.07*) At Level 1, students can answer questions involving familiar contexts where all relevant information is present and the questions are clearly defined. They are able to identify information and to carry out routine procedures according to direct instructions in explicit situations. They can perform actions that are obvious and follow immediately from the given stimuli.

Level 2 (*greater than 420.07 to 482.38*) At Level 2, students can interpret and recognize situations in contexts that require no more than direct inference. They can extract relevant information from a single source and make use of a single representational mode. Students at this level can employ basic algorithms, formula, procedures, or conventions. They are capable of direct reasoning and making literal interpretations of the results.

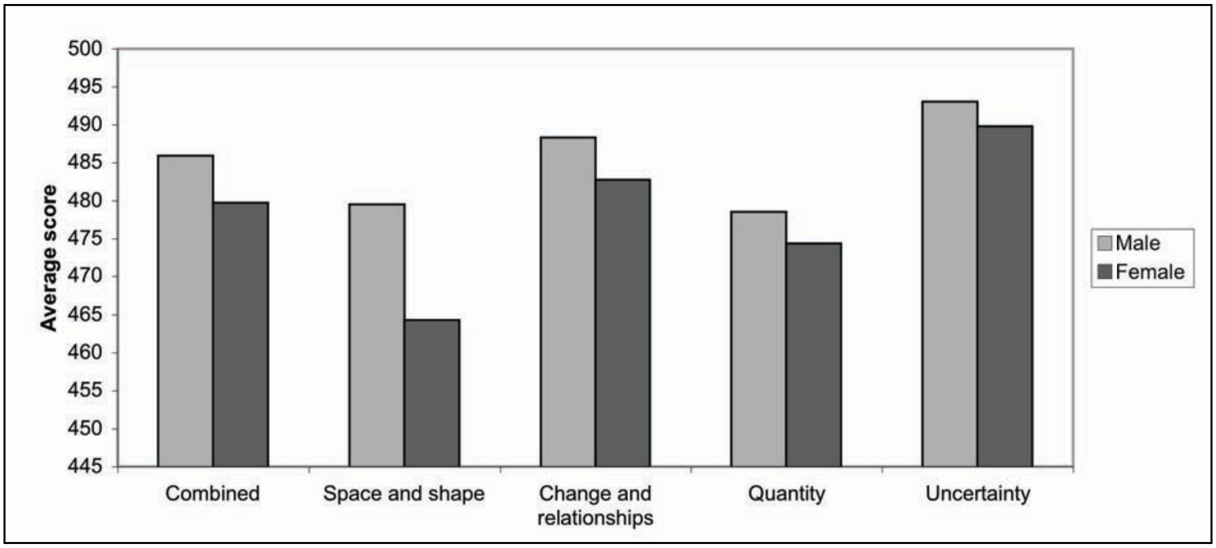
Level 3 (*greater than 482.38 to 544.68*) At Level 3, students can execute clearly described procedures, including those that require sequential decisions. They can select and apply simple problem solving strategies. Students at this level can interpret and use representations based on different information sources and reason directly from them. They can develop short communications reporting their interpretations, results, and reasoning.

Level 4 (*greater than 544.68 to 606.99*) At Level 4, students can work effectively with explicit models for complex concrete situations that may involve constraints or call for making assumptions. They can select and integrate different representations, including symbolic, linking them directly to aspects of real-world situations. Students at this level can utilize well developed skills and reason flexibly, with some insight, in these contexts. They can construct and communicate explanations and arguments based on their interpretations, arguments, and actions.

Level 5 (*greater than 606.99 to 669.3*) At Level 5, students can develop and work with models for complex situations, identifying constraints and specifying assumptions. They can select, compare, and evaluate appropriate problem solving strategies for dealing with complex problems related to these models. Students at this level can work strategically using broad, well-developed thinking and reasoning skills, appropriate linked representations, symbolic and formal characterizations, and insight pertaining to these situations. They can reflect on their actions, and formulate and communicate their interpretations and reasoning.

Level 6 (*greater than 669.3*) At Level 6, students can conceptualize, generalize, and utilize information based on their investigations and modeling of complex problem situations. They can link different information sources and representations and flexibly translate among them. Students at this level are capable of advanced mathematical thinking and reasoning. These students can apply this insight and understandings along with a mastery of symbolic and formal mathematical operations and relationships to develop new approaches and strategies for attacking novel situations. Students at this level can formulate and precisely communicate their actions and reflections regarding their findings, interpretations, arguments, and the appropriateness of these to the original situations (Lemke et al., 2005, p.18).

Figure C-11: Average Mathematics Literacy Scores of U.S. 15-Year-Olds, by Gender: 2003 PISA



Note: Level 1 (greater than 357.77 to 420.07), Level 2 (greater than 420.07 to 482.38), Level 3 (greater than 482.38 to 544.68), Level 4 (greater than 544.68 to 606.99), Level 5 (greater than 606.99 to 669.3), Level 6 (greater than 669.3)

Source: Lemke et al. (2005) Tables B-18 and B-20

Standardized mean difference, 15 year olds, gender*	
	PISA (2003)
Male-Female	0.07

*Standard deviations not provided for subscales

Table C-3: Percentage of U.S. 15-Year-Old Students Scoring at Each Proficiency Level, by Gender: 2003 PISA

	Below level 1	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6
Male	10.5	14.7	23.2	23.1	16.9	8.9	2.8
Female	9.9	16.4	24.6	24.5	16.2	7.2	1.2
Overall	10.2	15.5	23.9	23.8	16.6	8.0	2.0

Source: Lemke et al., 2005, Tables B-19 and B-6

Table C-4: Comparison of U.S. and Organisation for Economic Co-operation and Development (OECD) Countries' Average Scores on 2003 PISA Math Literacy

	U.S. average	OECD average	Number of OECD countries scoring higher than U.S.
Combined	483	500	20
Space and shape	472	496	20
Change and relationships	486	499	18
Quantity	476	501	23
Uncertainty	491	502	16

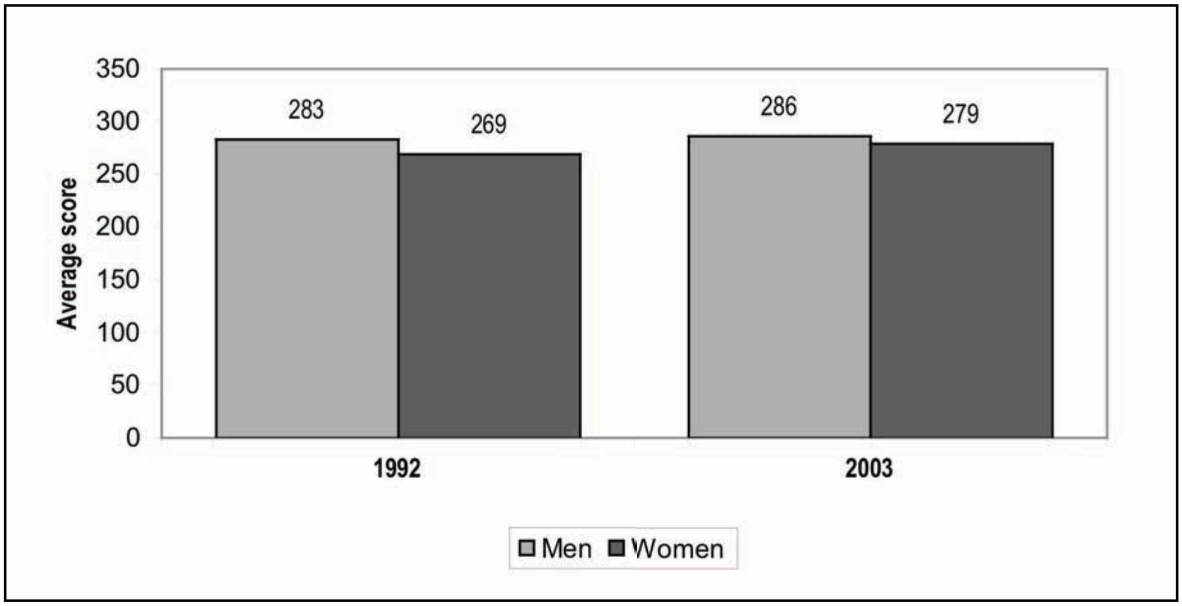
Source: Lemke et al., 2005, Table 2

National Adult Literacy Survey: NALS
National Assessment of Adult Literacy: NAAL

The Committee on Performance Levels for Adult Literacy set performance levels for quantitative literacy as Below Basic, Basic, Intermediate, and Proficient and defined them as follows, based on scores on NALS and NAAL:

> ***Below Basic (0–234)*** indicates no more than the most simple and concrete literacy skills.
>
> *Key abilities*—locating numbers and using them to perform simple *quantitative* operations (primarily addition) when the mathematical information is very concrete and familiar.
>
> ***Basic (235–289)*** indicates skills necessary to perform simple and everyday literacy activities.
>
> *Key abilities*—locating easily identifiable *quantitative* information and using it to solve simple, one-step problems when the arithmetic operation is specified or easily inferred.
>
> ***Intermediate (290–349)*** indicates skills necessary to perform moderately challenging literacy activities.
>
> *Key abilities*—locating less familiar *quantitative* information and using it to solve problems when the arithmetic operation is not specified or easily inferred.
>
> ***Proficient (350–500)*** indicates skills necessary to perform more complex and challenging literacy activities.
>
> *Key abilities*—locating more abstract quantitative information and using it to solve multistep problems when the arithmetic operations are not easily inferred and the problems are more complex (Kutner et al., 2006, p. 3).

Figure C-12: Average Quantitative Literacy Scores of Adults, by Sex: NALS 1992 and NAAL 2003



Note: Literacy levels: Below basic 0–234, Basic 235–289, Intermediate 290–349, Proficient 350–500

Source: Kutner, Greenberg, and Baer (2006), Figure 4.

Standardized mean difference adults, gender		
	1992	**2003**
Male-Female	0.21	0.11

Table C-5: Percentage of Adults in Each Quantitative Literacy Level, by Gender: NALS 1992 and NAAL 2003

		NALS, 1992	**NAAL, 2003**
Below basic	Male	24	21
	Female	28	22
Basic	Male	29	31
	Female	34	35
Intermediate	Male	31	33
	Female	28	32
Proficient	Male	17	16
	Female	9	11

Note: Below Basic (0–234) no more than the most simple and concrete literacy skills; Basic (235–289) skills necessary to perform simple and everyday literacy activities; Intermediate (290–349) skills necessary to perform moderately challenging literacy activities; Proficient (350–500) skills necessary to perform more complex and challenging literacy activities.

Source: Kutner et al., 2007, p. 14

APPENDIX D: Racial/Ethnic Differences

The following tables and figures summarize the data on math performance by Race/Ethnicity using data available on national samples. Data from the National Assessment of Educational Progress (NAEP) Long-Term Trend study illustrate performance between groups over the last 30 years. Data from the Trends in Math and Science Survey (TIMSS) illustrate the math performance of fourth- and eighth-graders. Data from the High School and Beyond (HS&B:80), National Education Longitudinal Study of 1988 (NELS:88), and Education Longitudinal Study of 2002 (ELS:2002) illustrate the math performance of 10th-grade students. Data from the National Adult Literacy Survey (NALS) and the National Assessment of Adult Literacy (NAAL) survey illustrate the quantitative literacy of adults. Data from the Program for International Student Assessment (PISA) illustrate the mathematics literacy and problem-solving proficiency of 15-year-olds. To facilitate the interpretation of the various scores, a description of the test benchmarks and performance levels associated with each test is provided.

National Assessment of Educational Progress Long-Term Trends: Mathematics Scores

This section presents the trends in long-term NAEP mathematics scores. The goal is to describe the differences in performance between groups over the last 30 years and to describe how their scores have evolved over time. For each reporting group, results are presented in the form of the average scale score for 1978–2004 and the percent of students at each achievement level in 1978, 1999, and 2004.

Methodology

All data presented in this section were obtained from the NAEP Data Explorer.[3] The Data Explorer allows users to create tables of results by custom combinations of reporting variables. The results can be reported in terms of mean score, percentage of students at or above performance levels, and score percentile.

The Data Explorer also reports standard errors and can calculate the statistical significance of changes in a variable between years or between variables in the same year. The statistical significance of changes between variables over time (e.g., the score difference between girls and boys in 1978 versus the score difference between girls and boys in 2004) is taken either directly from the *NAEP 2004 Trends in Academic Progress* or estimated using the reported standard error provided by the Data Explorer. Only differences that are statistically significant beyond the 0.05 level are described in the text of this section.

[3] http://nces.ed.gov/nationsreportcard/naepdata/.

Average Scale Scores and Performance Levels

The NAEP long-term trend assessments are scored on a 0–500 point scale, but all average scale score charts presented here are ranged from 180–340 for consistency and best visibility of score differences. Charts of average scale scores are reconstructed to resemble the gap charts in *NAEP 2004 Trends in Academic Progress*.

The following text was taken verbatim from the National Center for Education Statistics website, http://nces.ed.gov/nationsreportcard/ltt/performance-levels.asp in April 2007.

More detailed information about what students know and can do in each subject area can be gained by examining their attainment of specific performance levels in each assessment year. This process of developing the performance-level descriptions is different from that used to develop *achievement-level* descriptions in the main NAEP reports.

For each of the subject area scales, performance levels were set at 50-point increments from 150 through 350. The five performance levels—150, 200, 250, 300, and 350—were then described in terms of the knowledge and skills likely to be demonstrated by students who reached each level.

A "scale anchoring" process was used to define what it means to score in each of these levels. NAEP's scale anchoring follows an empirical procedure whereby the scaled assessment results are analyzed to delineate sets of questions that discriminate between adjacent performance levels on the scales. To develop these descriptions, assessment questions were identified that students at a particular performance level were more likely to answer successfully than students at lower levels. The descriptions of what students know and can do at each level are based on these sets of questions.

The guidelines used to select the questions were as follows: Students at a given level must have at least a specified probability of success with the questions (75% for mathematics, 80 % for reading), while students at the next lower level have a much lower probability of success (that is, the difference in probabilities between adjacent levels must exceed 30%). For each curriculum area, subject-matter specialists examined these empirically selected question sets and used their professional judgment to characterize each level. The scale anchoring for mathematics trend reporting was based on the 1986 assessment.

The five performance levels are applicable at all three age groups, but only three performance levels are discussed for each age: levels 150, 200, and 250 for age 9; levels 200, 250, and 300 for age 13; and levels 250, 300, and 350 for age 17. These performance levels are the ones most likely to show significant change within an age across the assessment years and do not include the levels that nearly all or almost no students attained at a particular age in each year.

The following description of each mathematics performance level was copied from http://nces.ed.gov/nationsreportcard/ltt/math-descriptions.asp in April 2007.

Level 350: Multistep Problem Solving and Algebra

Students at this level can apply a range of reasoning skills to solve multistep problems. They can solve routine problems involving fractions and percents, recognize properties of basic geometric figures, and work with exponents and square roots. They can solve a variety of two-step problems using variables, identify equivalent algebraic expressions, and solve linear equations and inequalities. They are developing an understanding of functions and coordinate systems.

Level 300: Moderately Complex Procedures and Reasoning

Students at this level are developing an understanding of number systems. They can compute with decimals, simple fractions, and commonly encountered percents. They can identify geometric figures, measure lengths and angles, and calculate areas of rectangles. These students are also able to interpret simple inequalities, evaluate formulas, and solve simple linear equations. They can find averages, make decisions based on information drawn from graphs, and use logical reasoning to solve problems. They are developing the skills to operate with signed numbers, exponents, and square roots.

Level 250: Numerical Operations and Beginning Problem Solving

Students at this level have an initial understanding of the four basic operations. They are able to apply whole number addition and subtraction skills to one-step word problems and money situations. In multiplication, they can find the product of a two-digit and a one-digit number. They can also compare information from graphs and charts, and are developing an ability to analyze simple logical relations.

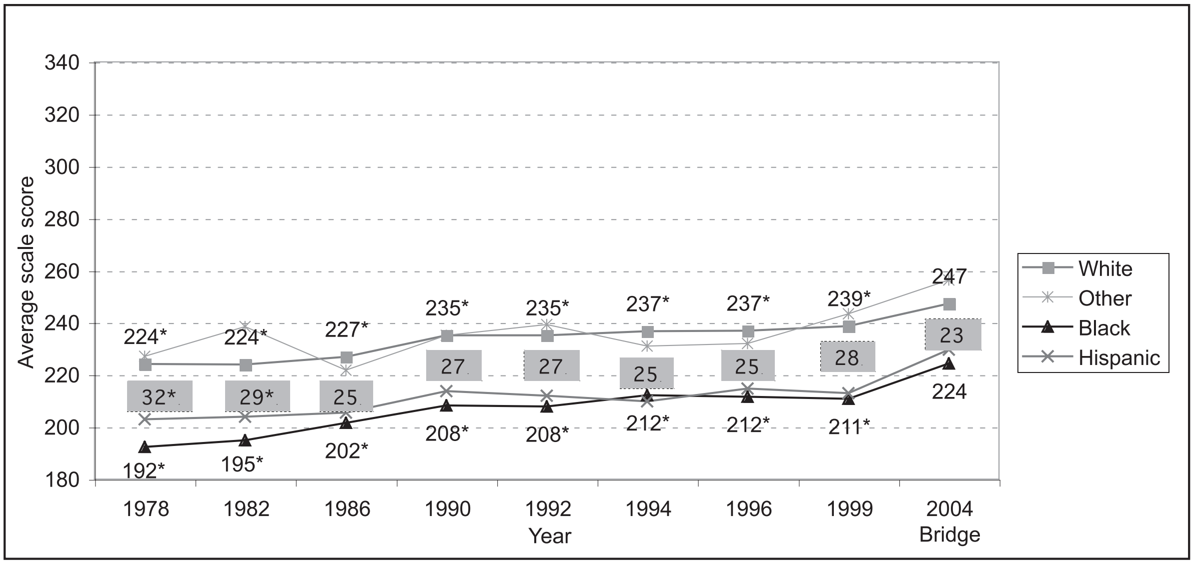
Level 200: Beginning Skills and Understandings

Students at this level have considerable understanding of two-digit numbers. They can add two-digit numbers but are still developing an ability to regroup in subtraction. They know some basic multiplication and division facts, recognize relations among coins, can read information from charts and graphs, and use simple measurement instruments. They are developing some reasoning skills.

Level 150: Simple Arithmetic Facts

Students at this level know some basic addition and subtraction facts, and most can add two-digit numbers without regrouping. They recognize simple situations in which addition and subtraction apply. They also are developing rudimentary classification skills.

Figure D-1: Average NAEP Scale Scores by Race/Ethnicity, Age 9: Intermittent Years From 1978–2004



*Indicates score or gap is significantly different from 2004.

Note: Data labels for white (above) and black (below). Between race/ethnicity score differences (gaps) are shown in shaded boxes only for years in which the gap is statistically significant. Labeled gaps may not reflect labeled scores because of rounding.

"Bridge" refers to updates made to NAEP in 2004. The updates replaced outdated material and accommodated more students with disabilities. In order to maintain the long-term trend, test takers were randomly assigned to either the old test form, called the bridge assessment, or the modified test form. Results from the bridge assessment should be compared to results from assessments prior to 2004, while results from the modified assessment should be compared to assessments given after 2004.

Source: Created by the Institute for Defense Analysis Science and Technology Policy Institute using the NAEP Data Explorer (http://nces.ed.gov/nationsreportcard/naepdata/).

What Is This Indicator?

This indicator presents the average scale score for 9-year-olds by race. "Other" includes Asian/Pacific Islander and American Indian/Alaska Native.

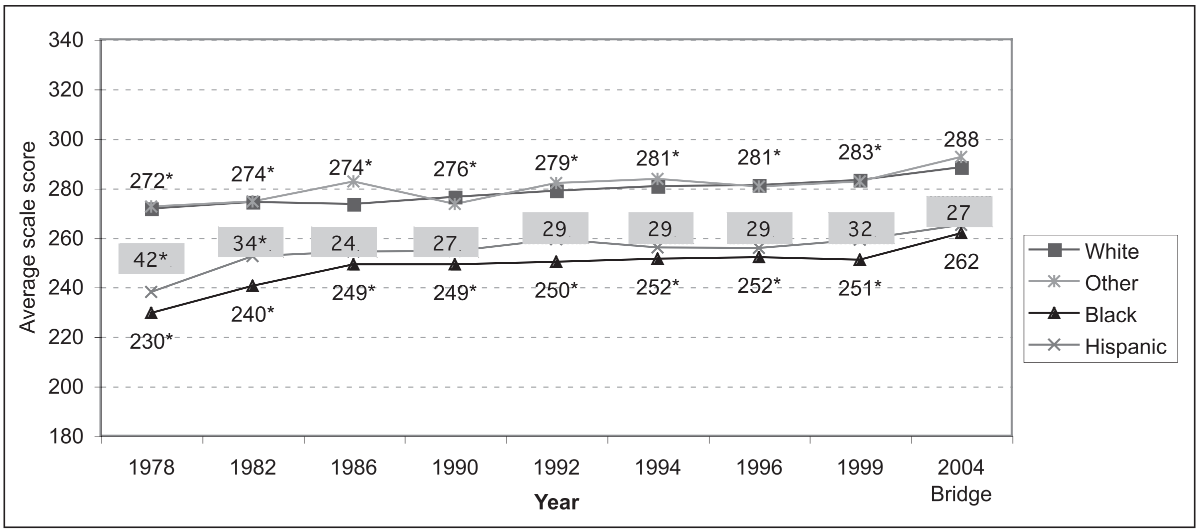
Discussion

- Whites and Other races significantly outscore blacks and Hispanics on the long-term mathematics NAEP at age 9.
- The average score for black students of 224 was 23 points lower than the average score for white students in 2004.
  - — The black-white gap has not changed significantly since 1986.
  - — The black-white gap has closed by 9 points since 1978.
- The average score for Hispanic students of 230 was 18 points lower than the average score for white students in 2004.
  - — The Hispanic-white gap has closed by 8 points since 1999.

— The Hispanic-white gap in 2004 was not significantly different from the gap in 1978.

- Average scores for all racial groups were higher in 2004 than in previous assessment years.
  - — The average score for black students was 224 in 2004, which was a 13 point increase from 1999 and a 32 point increase from 1978.
  - — The average score for Hispanic students was 230 in 2004, which was a 17 point increase from 1999 and a 27 point increase from 1978.
  - — The average score for white students was 247 in 2004, which was an 8 point increase from 1999 and a 23 point increase from 1978.
  - — The average score for Other students was 256 in 2004, which was a 13 point increase from 1999 and a 29 point increase from 1978.

Figure D-2: Average NAEP Scale Scores by Race/Ethnicity, Age 13: Intermittent Years From 1978–2004



*Indicates score or gap is significantly different from 2004.

Note: Data labels for white (above) and black (below). Between race/ethnicity score differences (gaps) are shown in shaded boxes only for years in which the gap is statistically significant. Labeled gaps may not reflect labeled scores because of rounding.

"Bridge" refers to updates made to NAEP in 2004. The updates replaced outdated material and accommodated more students with disabilities. In order to maintain the long-term trend, test takers were randomly assigned to either the old test form, called the bridge assessment, or the modified test form. Results from the bridge assessment should be compared to results from assessments prior to 2004, while results from the modified assessment should be compared to assessments given after 2004.

Source: Created by the Institute for Defense Analysis Science and Technology Policy Institute using the NAEP Data Explorer (http://nces.ed.gov/nationsreportcard/naepdata/).

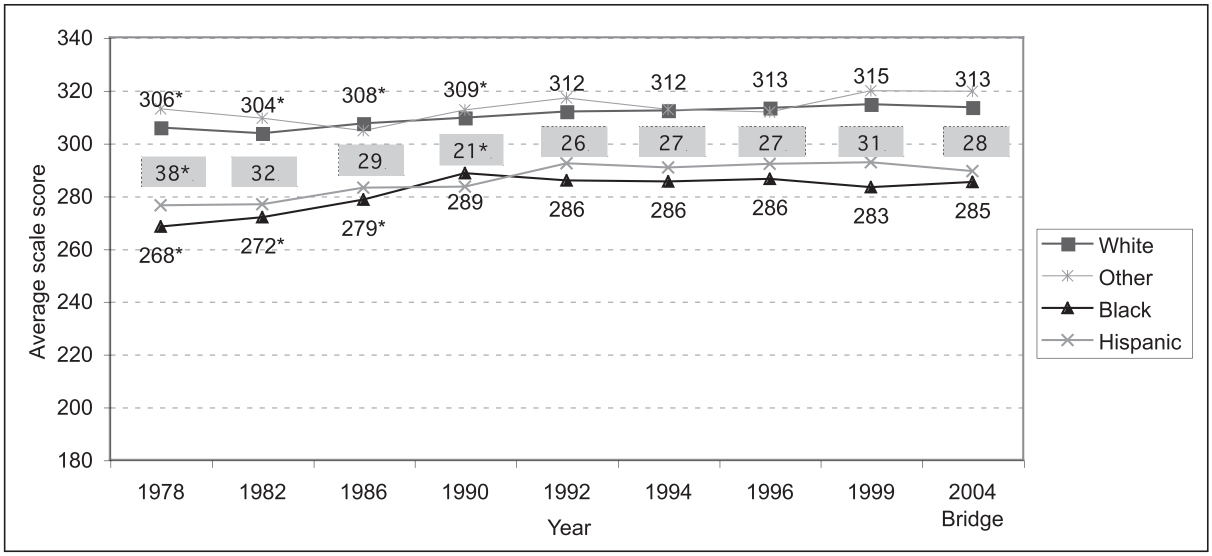
What Is This Indicator?

This indicator presents the average scale score for 13-year-olds by race. “Other” includes Asian/Pacific Islander and American Indian/Alaska Native.

Discussion

- Whites and Other races significantly outscore blacks and Hispanics on the long-term mathematics NAEP at age 13.
- The average score for black students was 27 points lower than the average score for white students in 2004.
    - — The black-white gap has not changed significantly since 1986.
    - — The black-white gap has closed by 15 points since 1978.
- The average score for Hispanic students was 23 points lower than the average score for white students in 2004.
    - — The Hispanic-white gap in 2004 was not significantly different from the gap in 1999.
    - — The Hispanic-white gap has closed by 11 points since 1978.
- Average scores for whites, blacks, and Hispanics were higher in 2004 than in previous assessment years.
    - — The average score for black 13-year-old students was 262 in 2004, which was an 11 point increase from 1999 and a 32 point increase from 1978.
    - — The average score for Hispanic students was 265 in 2004, which was a 6 point increase from 1999 and a 27 point increase from 1978.
    - — The average score for white students was 288 in 2004, which was a 5 point increase from 1999 and a 17 point increase from 1978.
    - — The average score for Other students was 292 in 2004, which was a 20 point increase from 1978, but not significantly different from 1999.

Figure D-3: Average NAEP Scale Scores by Race/Ethnicity, Age 17: Intermittent Years From 1978–2004



*Indicates score or gap is significantly different from 2004.

Note: Data labels for white (above) and black (below). Between race/ethnicity score differences (gaps) are shown in shaded boxes only for years in which the gap is statistically significant. Labeled gaps may not reflect labeled scores because of rounding.

“Bridge” refers to updates made to NAEP in 2004. The updates replaced outdated material and accommodated more students with disabilities. In order to maintain the long-term trend, test takers were randomly assigned to either the old test form, called the bridge assessment, or the modified test form. Results from the bridge assessment should be compared to results from assessments prior to 2004, while results from the modified assessment should be compared to assessments given after 2004.

Source: Created by the Institute for Defense Analysis Science and Technology Policy Institute using the NAEP Data Explorer (http://nces.ed.gov/nationsreportcard/naepdata/).

What Is This Indicator?

This indicator presents the average scale score for 17-year-olds by race. “Other” includes Asian/Pacific Islander and American Indian/Alaska Native.

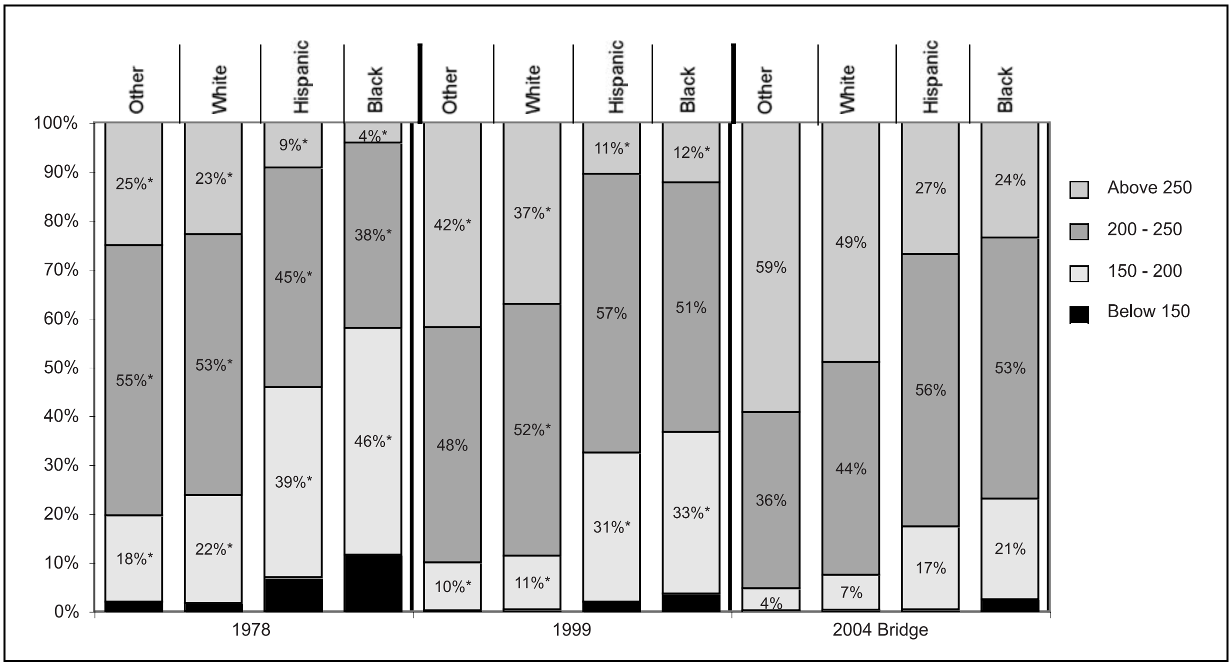
Discussion

- Whites and Other races significantly outscore blacks and Hispanics on the long-term mathematics NAEP at age 17.
- The average score for black students was 28 points lower than the average score for white students in 2004.
    - — The black-white gap has not changed significantly since 1992.
    - — The black-white gap has closed by 10 points since 1978.
- The average score for Hispanic students was 24 points lower than the average score for white students in 2004.
    - — The Hispanic-white gap in 2004 is not significantly different from the gap in 1999.

— The Hispanic-white gap in 2004 is not significantly different from the gap in 1978.

- The average scale scores for whites, blacks, and Hispanics have increased since 1978, but the average scale scores for all races have been flat since 1992.
  — The average score for black students was 285 in 2004, which was a 17 point increase from 1978.
  — The average score for Hispanic students was 289 in 2004, which was a 13 point increase from 1978.
  — The average score for white students was 313 in 2004, which was a 7 point increase from 1978.

Figure D-4: Percent at NAEP Performance Levels by Race/Ethnicity, Age 9: 1978, 1999, and 2004



*Indicates percentage is significantly different from 2004. Differences between categories (racial/ethnic groups) may not be statistically significant.

Note: "Bridge" refers to updates made to NAEP in 2004. The updates replaced outdated material and accommodated more students with disabilities. In order to maintain the long-term trend, test takers were randomly assigned to either the old test form, called the bridge assessment, or the modified test form. Results from the bridge assessment should be compared to results from assessments prior to 2004, while results from the modified assessment should be compared to assessments given after 2004.

Source: Created by the Institute for Defense Analysis Science and Technology Policy Institute using the NAEP Data Explorer (http://nces.ed.gov/nationsreportcard/naepdata/).

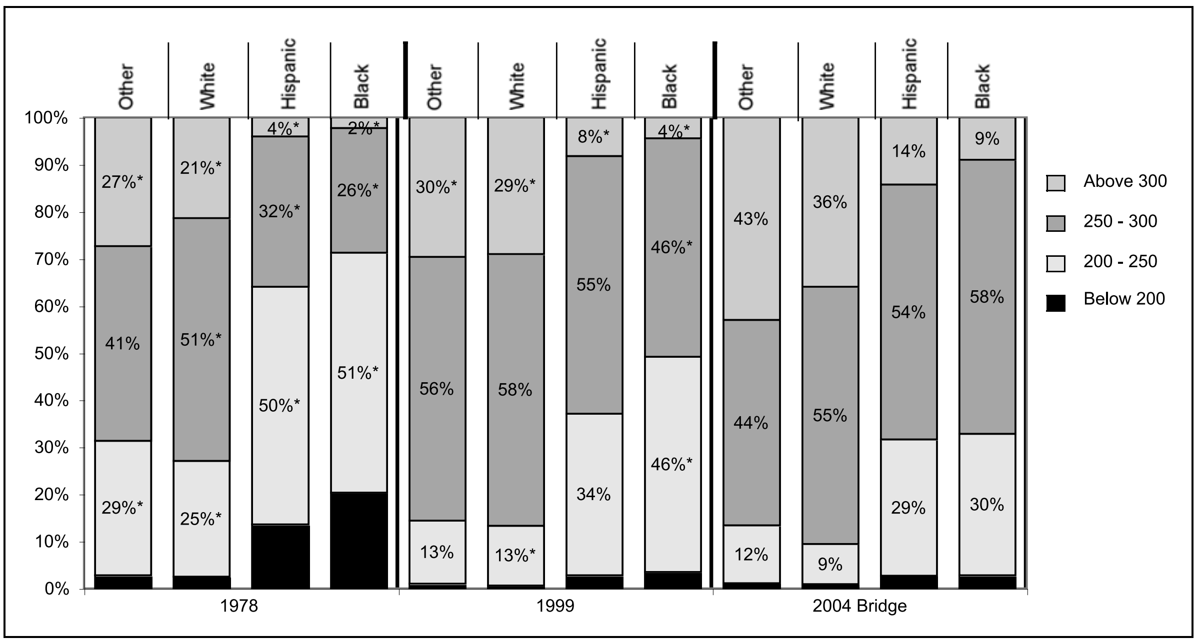
What Is This Indicator?

This indicator presents the percentage of 9-year-olds reaching each performance level by race. The performance levels reported at age 9 are 150—Simple Arithmetic Facts, 200—Beginning Skills and Understandings, and 250—Numerical Operations and Beginning Problem Solving.

Discussion

- More 9-year-old whites and Other races scored at or above the 250 level than blacks and Hispanics.
  - — Differences between whites and Other, and between blacks and Hispanics at the 250 level were generally not significant.
  - — 49% of white students scored at or above 250 in 2004, while only 24%, or half as many black 9-year-olds reached the 250 performance level in 2004.
- The percentages of 9-year-olds of all races at or above the 250 level have increased since the 1999 and the 1978 assessments.
  - — The percentage of black 9-year-olds scoring at or above 250 has increased by a factor of 6 since 1978 and doubled since 1999, going from 4% in 1978 to 12% in 1999 and 24% in 2004.
  - — Meanwhile, the percentage of white 9-year-olds scoring at or above 250 increased from 23% to 49% between 1978 and 2004.
- While blacks and Hispanics have seen large increases in the percent of 9-year-olds in the top performance level, the black-white gap at the 250 performance level has widened slightly since 1978, going from 19% in 1978 to 25% in 1999 and 2004.

Figure D-5: Percent at NAEP Performance Levels by Race/Ethnicity, Age 13: 1978, 1999, and 2004



*Indicates percentage is significantly different from 2004. Differences between categories (racial/ethnic groups) may not be statistically significant.

Note: "Bridge" refers to updates made to NAEP in 2004. The updates replaced outdated material and accommodated more students with disabilities. In order to maintain the long-term trend, test takers were randomly assigned to either the old test form, called the bridge assessment, or the modified test form. Results from the bridge assessment should be compared to results from assessments prior to 2004, while results from the modified assessment should be compared to assessments given after 2004.

Source: Created by the Institute for Defense Analysis Science and Technology Policy Institute using the NAEP Data Explorer (http://nces.ed.gov/nationsreportcard/naepdata/).

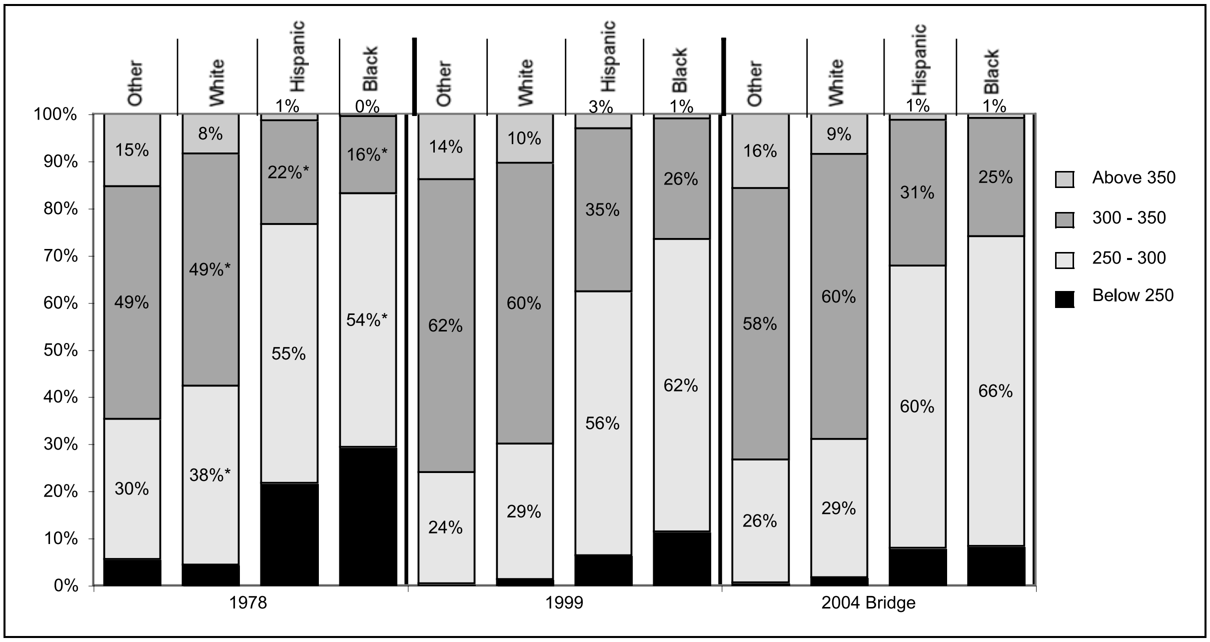
What Is This Indicator?

This indicator presents the percentage of 13-year-olds reaching each performance level by race. The performance levels reported at age 13 are 200—Beginning Skills and Understandings, 250—Numerical Operations and Beginning Problem Solving, and 300—Moderately Complex Procedures and Reasoning.

Discussion

- More 13-year-old whites and Other scored at or above the 300 level than blacks and Hispanics.
  - — Differences between whites and Other and between blacks and Hispanics at the 300 level were generally not significant.
  - — 36% of white students scored at or above 300 in 2004, while only 9% of black 9-year-olds reached the 300 performance level in 2004.
- The percentages of 13-year-olds of all races at or above the 300 level have increased since the 1999 and the 1978 assessments.
- The black-white gap for 13-year-olds at the 300 performance level has widened slightly since 1978, going from 19% in 1978 to 27% in 2004.
  - — During this time period the percentage of black 13-year-olds at the 300 level only increased from 2% in 1978 to 9% in 2004.
  - — Meanwhile, the percentage of white 13-year-olds at the 300 level increased from 21% in 1978 to 36% in 2004.

Figure D-6: Percent at NAEP Performance Levels by Race/Ethnicity, Age 17: 1978, 1999, and 2004



*Indicates percentage is significantly different from 2004. Differences between categories (racial/ethnic groups) may not be statistically significant.

Note: "Bridge" refers to updates made to NAEP in 2004. The updates replaced outdated material and accommodated more students with disabilities. In order to maintain the long-term trend, test takers were randomly assigned to either the old test form, called the bridge assessment, or the modified test form. Results from the bridge assessment should be compared to results from assessments prior to 2004, while results from the modified assessment should be compared to assessments given after 2004.

Source: Created by the Institute for Defense Analysis Science and Technology Policy Institute using the NAEP Data Explorer (http://nces.ed.gov/nationsreportcard/naepdata/).

What Is This Indicator?

This indicator presents the percentage of 17-year-olds reaching each performance level by gender. The performance levels reported at age 17 are 250—Numerical Operations and Beginning Problem Solving, 300—Moderately Complex Procedures and Reasoning, and 350—Multistep Problem Solving and Algebra.

Discussion

- The percentage of 17-year-olds at or above the 350 level has not changed over time, although the percentage at the 300 level has increased for all races since 1978.
- In 2004 and 1978, 8.5% of whites reached the 350 performance level, while less than one percent of black 17-year-olds scored at or above 350. In 1999, the difference was not statistically significant.
- Differences between whites and Other, and blacks and Hispanics are generally not statistically significant for 17-year-olds at the 350 level.

Trends in Math and Science Survey: TIMSS

The TIMSS 2003 International Benchmarks of Mathematics Achievement are defined in Mullis et al. (2004, p. 63) as follows:

Grade 8

Advanced International Benchmark – 625

Students can organize information, make generalizations, solve non-routine problems, and draw and justify conclusions from data. They can compute percent change and apply their knowledge of numeric and algebraic concepts, and relationships to solve problems. Students can solve simultaneous linear equations and model simple situations algebraically. They can apply their knowledge of measurement and geometry in complex problem situations. They can interpret data from a variety of tables and graphs, including interpolation and extrapolation.

High International Benchmark – 550

Students can apply their understanding and knowledge in a wide variety of relatively complex situations. They can order, relate, and compute fractions and decimals to solve word problems, operate with negative integers, and solve multi-step word problems involving proportions with whole numbers. Students can solve simple algebraic problems including evaluating expressions, solving simultaneous linear equations, and using a formula to determine the value of a variable. Students can find areas and volumes of simple geometric shapes and use knowledge of geometric properties to solve problems. They can solve probability problems and interpret data in a variety of graphs and tables.

Intermediate International Benchmark – 475

Students can apply basic mathematical knowledge in straightforward situations. They can add, subtract, or multiply to solve one-step word problems involving whole numbers and decimals. They can identify representations of common fractions and relative sizes of fractions. They understand simple algebraic relationships and solve linear equations with one variable. They demonstrate understanding of properties of triangles and basic geometric concepts including symmetry and rotation. They recognize basic notions of probability. They can read and interpret graphs, tables, maps, and scales.

Low International Benchmark – 400

Students have some basic mathematical knowledge. (Mullis et al., 2004, p.62)

Grade 4

Advanced International Benchmark – 625

Students can apply their understanding and knowledge in a wide variety of relatively complex situations. They demonstrate a developing understanding of fractions and decimals, and the relationship between them. They can select appropriate information to solve multistep word problems involving proportions. They can formulate or select a rule for a relationship. They show understanding of area and can use measurement concepts to solve a variety of problems. They show some understanding of rotation. They can organize, interpret, and represent data to solve problems.

High International Benchmark – 550

Student can apply their knowledge and understanding to solve problems. Students can solve multistep word problems involving addition, multiplication, and division. They can use their understanding of place value and simple fractions to solve problems. They can identify a number sentence that represents situations. Students show understanding of three-dimensional objects, how shapes can make other shapes, and simple transformation in a plane. They demonstrate a variety of measurement skills, and can interpret and use data in tables and graphs to solve problems.

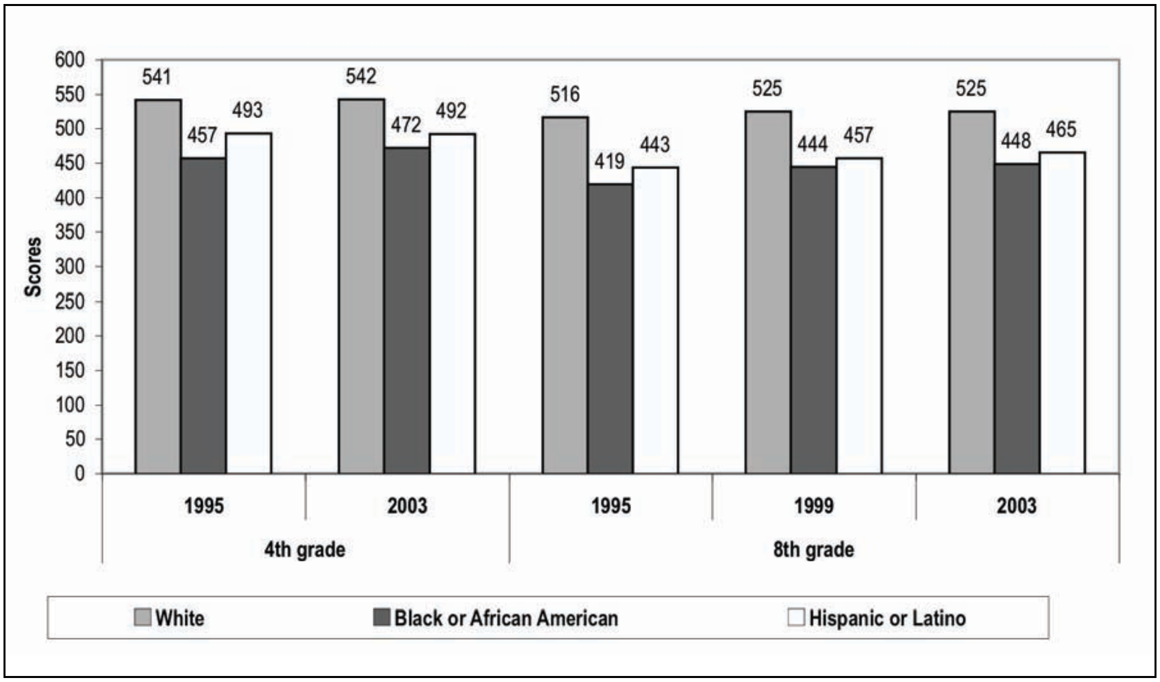
Intermediate International Benchmark – 475

Students can apply basic mathematical knowledge in straightforward situations. They can read, interpret, and use different representations of numbers. They can perform operations with three- and four-digit numbers and decimals. They can extend simple patterns. They are familiar with a range of two-dimensional shapes, and read and interpret different representations of the same data.

Low International Benchmark – 400

Students have some basic mathematical knowledge. Students demonstrate an understanding of whole numbers and can do simple computations with them. They demonstrate familiarity with the basic properties of triangles and rectangles. They can read information from simple bar graphs.

Figure D-7: Average TIMSS Mathematical Scale Scores of U.S. 4th- and 8th-Graders, by Race/Ethnicity: Various Years From 1995–2003



Note: TIMSS international benchmarks: Low 400, Intermediate 475, High 550, Advanced 625

Source: Gonzales et al. (2004), Figures 1 & 2.

Standardized mean difference TIMSS, race/ethnicity					
	4th grade		**8th grade**		
	1995	*2003*	*1995*	*1999*	*2003*
White-Black	1.01	0.92	1.08	0.92	0.96
White-Hispanic	0.57	0.66	0.81	0.77	0.75

High School and Beyond of 1980: HS&B:80
National Education Longitudinal Study of 1988: NELS:88
Education Longitudinal Study of 2002: ELS:2002

The scores on the HS&B:80, NELS:88, and ELS:2002 are Item Response Theory (IRT) scores on the NELS:88 1990 58-item scale. IRT estimates achievement based on patterns of correct, incorrect, and unanswered questions. "The IRT-estimated number-right score reflects an estimate of the number of these 58 items that an examinee would have answered correctly if he or she had taken all of the items that appeared on the multiform 1990 NELS:88 mathematics test. The score is the probability of a correct answer on each item, summed over the total mathematics 58-item pool" (Cahalan et al., 2006, p. 45). These scores are not directly translated into probability of proficiency scores. However, five probability of proficiency scores in mathematics were estimated for students using performance on clusters of four items each as follows:

Probability of Mastery, Mathematics Levels

1) Simple arithmetical operations on whole numbers, such as simple arithmetic expressions involving multiplication or division of integers;

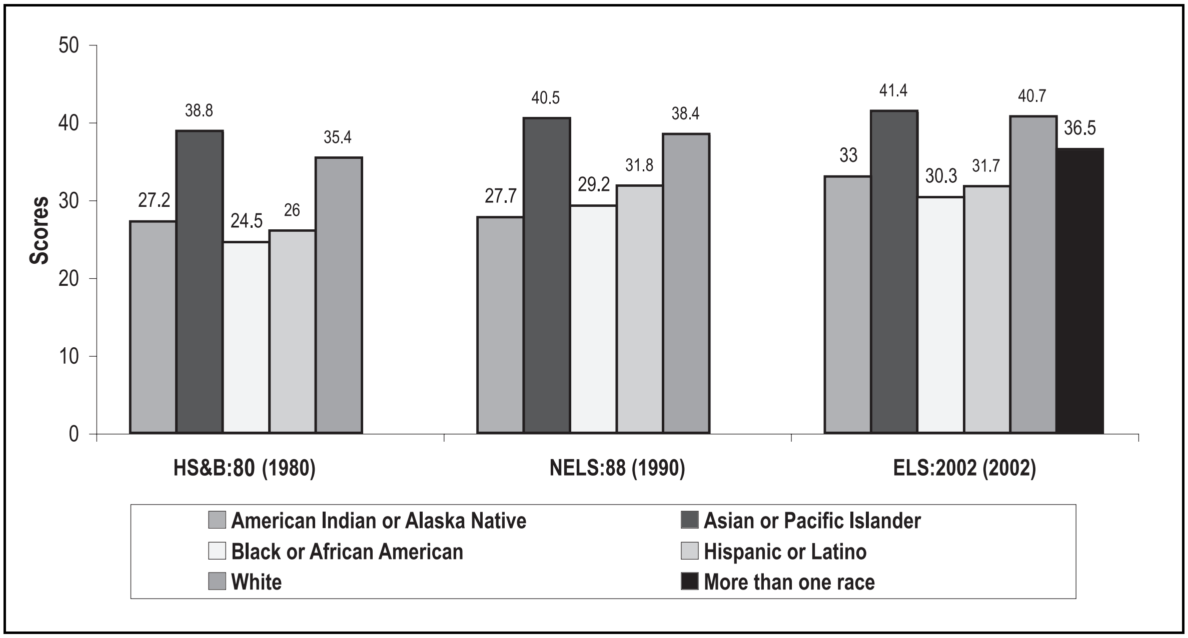
2) Simple operations with decimals, fractions, powers, and roots, such as comparing expressions, given information about exponents;

3) Simple problem solving, requiring the understanding of low-level mathematical concepts, such as simplifying an algebraic expression or comparing the length of line segments illustrated in a diagram;

4) Understanding of intermediate-level mathematical concepts and/or multistep solutions to word problems such as drawing an inference based on an algebraic expression or inequality; and

5) Complex multistep word problems and/or advanced mathematics material such as a two-step problem requiring evaluation of functions. (Cahalan et al., 2006, p. A-28)

Figure D-8: IRT—Estimated Average Math Score (10th-Grade), by Race/Ethnicity (HS&B:80, NELS:88, ELS:2002)



Note: IRT scale score is the estimated number right out of a total of 58.

Source: Cahalan et al. (2006), Tables 18 and 19.

Standardized mean difference sophomores, race/ethnicity			
	HS&B (1980)	**NELS:88 (1990)**	**ELS:2002 (2002)**
White-American Indian	0.71	0.94	0.78
White-Asian	-0.28	-0.18	-0.07
White-Black	1.01	0.82	1.03
White-Hispanic	0.84	0.58	0.84
White-Other			0.39

Table D-1: Probability of 10th-Grade Proficiency in Mathematics by Race/Ethnicity

	NELS:88 (1990)	ELS:2002 (2002)
Level 1		
Asian or Pacific Islander	93.7	95.2
Black or African American	80.8	83.8
Hispanic or Latino	85.0	83.7
White	93.3	95.5
Level 2		
Asian or Pacific Islander	73.7	77.6
Black or African American	38.4	42.3
Hispanic or Latino	44.9	46.9
White	69.6	77.9
Level 3		
Asian or Pacific Islander	57.8	60.2
Black or African American	18.7	19.4
Hispanic or Latino	24.4	25.5
White	50.1	57.9
Level 4		
Asian or Pacific Islander	29.6	31.7
Black or African American	5.2	4.7
Hispanic or Latino	8.0	8.8
White	22.5	27.0
Level 5		
Asian or Pacific Islander	1.2	4.0
Black or African American	less than 0.1	0.1
Hispanic or Latino	0.1	0.3
White	0.5	1.2

Note: Proficiency levels – 1) Simple arithmetical operations with whole numbers; 2) Simple operations with decimals, fractions, powers, and roots; 3) Simple problem solving, requiring the understanding of low-level mathematical concepts; 4) Understanding of intermediate-level mathematical concepts and/or multistep solutions to word problems; and 5) Complex multistep word problems and/or advanced mathematics material.

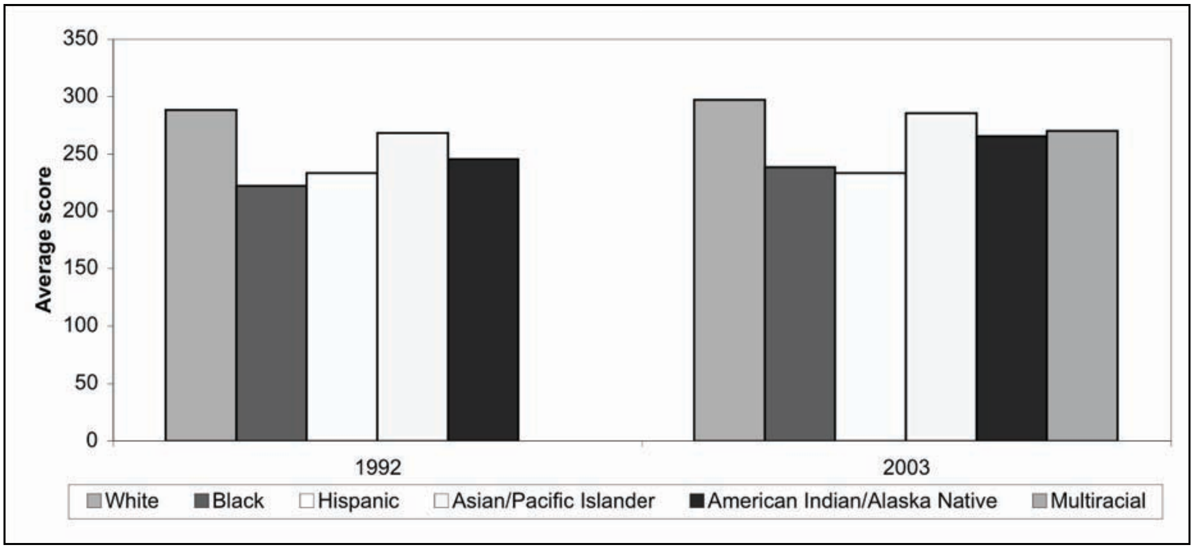
Source: Cahalan et al., 2006, p. 58.

National Adult Literacy Survey: NALS
National Assessment of Adult Literacy: NAAL

The Committee on Performance Levels for Adult Literacy set performance levels for quantitative literacy as Below Basic, Basic, Intermediate, and Proficient and defined them as follows, based on scores on NALS and NAAL:

> ***Below Basic (0–234)*** indicates no more than the most simple and concrete literacy skills.
>
> *Key abilities*—locating numbers and using them to perform simple *quantitative* operations (primarily addition) when the mathematical information is very concrete and familiar.
>
> ***Basic (235–289)*** indicates skills necessary to perform simple and everyday literacy activities.
>
> *Key abilities*—locating easily identifiable *quantitative* information and using it to solve simple, one-step problems when the arithmetic operation is specified or easily inferred.
>
> ***Intermediate (290–349)*** indicates skills necessary to perform moderately challenging literacy activities.
>
> *Key abilities*—locating less familiar *quantitative* information and using it to solve problems when the arithmetic operation is not specified or easily inferred.
>
> ***Proficient (350–500)*** indicates skills necessary to perform more complex and challenging literacy activities.
>
> *Key abilities*—locating more abstract quantitative information and using it to solve multistep problems when the arithmetic operations are not easily inferred and the problems are more complex. (Kutner et al., 2006, p. 3).

Figure D-9: Average Quantitative Literacy Scores of Adults, by Race/Ethnicity: NALS 1992 and NAAL 2003



Note: Literacy levels: Below basic 0-234, Basic 235-289, Intermediate 290-349, Proficient 350-500

Source: Kutner, Greenberg, and Baer (2006), Figure 1: Kutner et al. (2006), Figure 2-6a.

Standardized mean difference adults, race/ethnicity		
	NALS 1992	**NAAL 2003**
White-Black	1.00	0.97
White-Hispanic	0.83	1.05
White-Asian	-0.17	0.08
White-American Indian	0.65	0.52
White-Multiracial		0.44

Table D-2: Percentage of Adults in Each Quantitative Literacy Level, by Race/Ethnicity: NALS 1992 and NAAL 2003

		NALS, 1992	NAAL, 2003
Below basic			
	White	9	7
	Black	30	24
	Hispanic	35	44
	Asian/Pacific Islander	25	14
	American Indian/Alaska Native	17	19
	Multiracial		7
Basic			
	White	25	25
	Black	41	43
	Hispanic	33	30
	Asian/Pacific Islander	30	32
	American Indian/Alaska Native	43	29
	Multiracial		35
Intermediate			
	White	48	51
	Black	27	31
	Hispanic	28	23
	Asian/Pacific Islander	36	42
	American Indian/Alaska Native	35	41
	Multiracial		54
Proficient			
	White	18	17
	Black	2	2
	Hispanic	5	4
	Asian/Pacific Islander	9	12
	American Indian/Alaska Native	5	10
	Multiracial		4

Note: Below Basic (0–234)—no more than the most simple and concrete literacy skills; Basic (235–289)—skills necessary to perform simple and everyday literacy activities; Intermediate (290–349)—skills necessary to perform moderately challenging literacy activities; Proficient (350–500)—skills necessary to perform more complex and challenging literacy activities.

Source: Kutner et al., 2007, p. 16.

Program for International Student Assessment: PISA

Mathematics literacy can be classified by proficiency levels, based on scores on the PISA, as follows:

Below level 1 (*less than or equal to 357.77*)

Level 1 (*greater than 357.77 to 420.07*) At Level 1, students can answer questions involving familiar contexts where all relevant information is present and the questions are clearly defined. They are able to identify information and to carry out routine procedures according to direct instructions in explicit situations. They can perform actions that are obvious and follow immediately from the given stimuli.

Level 2 (*greater than 420.07 to 482.38*) At Level 2, students can interpret and recognize situations in contexts that require no more than direct inference. They can extract relevant information from a single source and make use of a single representational mode. Students at this level can employ basic algorithms, formula, procedures, or conventions. They are capable of direct reasoning and making literal interpretations of the results.

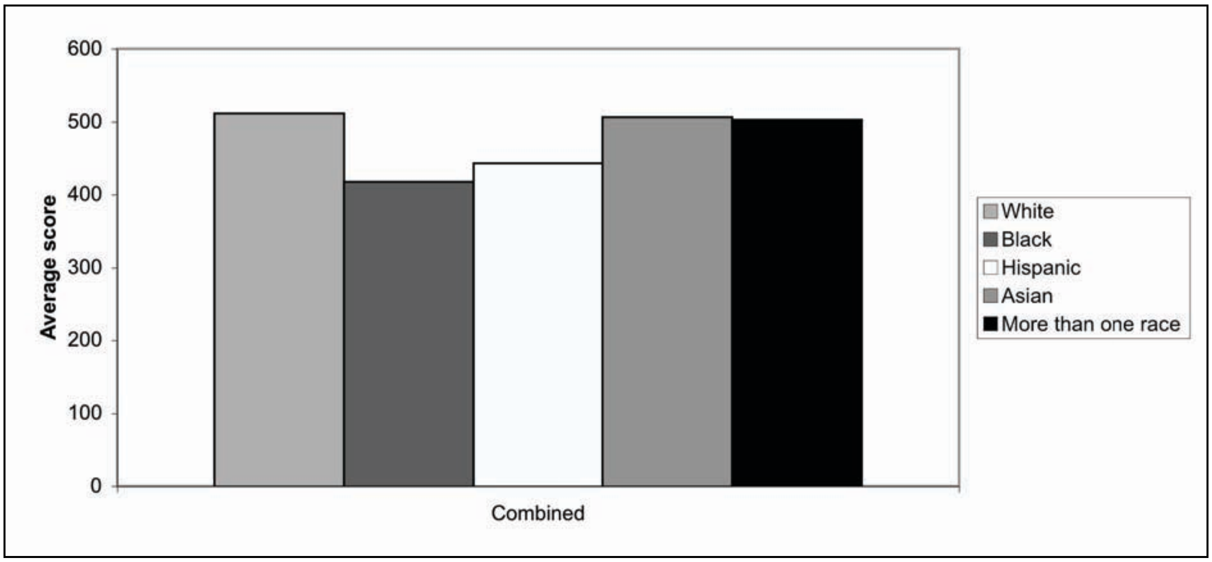
Level 3 (*greater than 482.38 to 544.68*) At Level 3, students can execute clearly described procedures, including those that require sequential decisions. They can select and apply simple problem solving strategies. Students at this level can interpret and use representations based on different information sources and reason directly from them. They can develop short communications reporting their interpretations, results, and reasoning.

Level 4 (*greater than 544.68 to 606.99*) At Level 4, students can work effectively with explicit models for complex concrete situations that may involve constraints or call for making assumptions. They can select and integrate different representations, including symbolic, linking them directly to aspects of real-world situations. Students at this level can utilize well developed skills and reason flexibly, with some insight, in these contexts. They can construct and communicate explanations and arguments based on their interpretations, arguments, and actions.

Level 5 (*greater than 606.99 to 669.3*) At Level 5, students can develop and work with models for complex situations, identifying constraints and specifying assumptions. They can select, compare, and evaluate appropriate problem solving strategies for dealing with complex problems related to these models. Students at this level can work strategically using broad, well-developed thinking and reasoning skills, appropriate linked representations, symbolic and formal characterizations, and insight pertaining to these situations. They can reflect on their actions and formulate and communicate their interpretations and reasoning.

Level 6 (*greater than 669.3*) At Level 6, students can conceptualize, generalize, and utilize information based on their investigations and modeling of complex problem situations. They can link different information sources and representations and flexibly translate among them. Students at this level are capable of advanced mathematical thinking and reasoning. These students can apply this insight and understandings along with a mastery of symbolic and formal mathematical operations and relationships to develop new approaches and strategies for attacking novel situations. Students at this level can formulate and precisely communicate their actions and reflections regarding their findings, interpretations, arguments, and the appropriateness of these to the original situations (Lemke et al., 2005, p.18).

Figure D-10: Average Mathematic Literacy Scores of U.S. 15-Year-Olds, by Race/Ethnicity: 2003 PISA



Note: Level 1 (greater than 357.77 to 420.07), Level 2 (greater than 420.07 to 482.38), Level 3 (greater than 482.38 to 544.68), Level 4 (greater than 544.68 to 606.99), Level 5 (greater than 606.99 to 669.3), Level 6 (greater than 669.3).

Source: Lemke et al. (2005) Tables B-26.

Standardized mean difference, 15-year-olds, race/ethnicity	
White-Black	0.99
White-Hispanic	0.72
White-Asian	0.06
White-Multiracial	0.10

Source: Lemke et al., 2005, Tables B-19 and B-6.

APPENDIX E: Socioeconomic Differences (SES)

The following tables and exhibits summarize the data on math performance by SES using data available on national samples. Data from the National Assessment of Educational Progress (NAEP) Long-Term Trend study illustrate performance between groups over the last 30 years. Data from the Trends in Math and Science Survey (TIMSS) illustrate the math performance of fourth- and eighth-graders. Data from the High School and Beyond (HS&B:80), National Education Longitudinal Study of 1988 (NELS:88), and Education Longitudinal Study of 2002 (ELS:2002) illustrate the math performance of 10th-grade students. Data from the National Adult Literacy Survey (NALS) and the National Assessment of Adult Literacy (NAAL) survey illustrate the quantitative literacy of adults. Data from the Program for International Student Assessment (PISA) illustrate the mathematics literacy and problem-solving proficiency of 15-year-olds. To facilitate the interpretation of the various scores, a description of the test benchmarks and performance levels associated with each test is provided.

National Assessment of Educational Progress Long-Term Trends: Mathematics Scores

This section presents the trends in long-term NAEP mathematics scores. The goal is to describe the differences in performance between groups over the last 30 years and to describe how their scores have evolved over time. For each reporting group, results are presented in the form of the average scale score for 1978–2004 and the percent of students at each achievement level in 1978, 1999, and 2004.

Methodology

All data presented in this section were obtained from the NAEP Data Explorer.[4] The Data Explorer allows users to create tables of results by custom combinations of reporting variables. The results can be reported in terms of mean score, percentage of students at or above performance levels, and score percentile.

The Data Explorer also reports standard errors and can calculate the statistical significance of changes in a variable between years or between variables in the same year. The statistical significance of changes between variables over time (e.g., the score difference between girls and boys in 1978 versus the score difference between girls and boys in 2004) is taken either directly from the *NAEP 2004 Trends in Academic Progress* or estimated using the reported standard error provided by the Data Explorer. Only differences that are statistically significant beyond the 0.05 level are described in the text of this section.

[4] http://nces.ed.gov/nationsreportcard/naepdata/

Average Scale Scores and Performance Levels

The NAEP long-term trend assessments are scored on a 0–500 point scale, but all average scale score charts presented here are ranged from 180–340 for consistency and best visibility of score differences. Charts of average scale scores are reconstructed to resemble the gap charts in *NAEP 2004 Trends in Academic Progress*.

The following text was taken verbatim from the National Center for Education Statistics website, http://nces.ed.gov/nationsreportcard/ltt/performance-levels.asp in April 2007.

More detailed information about what students know and can do in each subject area can be gained by examining their attainment of specific performance levels in each assessment year. This process of developing the performance-level descriptions is different from that used to develop *achievement-level* descriptions in the main NAEP reports.

For each of the subject area scales, performance levels were set at 50-point increments from 150 through 350. The five performance levels—150, 200, 250, 300, and 350—were then described in terms of the knowledge and skills likely to be demonstrated by students who reached each level.

A "scale anchoring" process was used to define what it means to score in each of these levels. NAEP's scale anchoring follows an empirical procedure whereby the scaled assessment results are analyzed to delineate sets of questions that discriminate between adjacent performance levels on the scales. To develop these descriptions, assessment questions were identified that students at a particular performance level were more likely to answer successfully than students at lower levels. The descriptions of what students know and can do at each level are based on these sets of questions.

The guidelines used to select the questions were as follows: Students at a given level must have at least a specified probability of success with the questions (75% for mathematics, 80% for reading), while students at the next lower level have a much lower probability of success (that is, the difference in probabilities between adjacent levels must exceed 30%). For each curriculum area, subject-matter specialists examined these empirically selected question sets and used their professional judgment to characterize each level. The scale anchoring for mathematics trend reporting was based on the 1986 assessment.

The five performance levels are applicable at all three age groups, but only three performance levels are discussed for each age: levels 150, 200, and 250 for age 9; levels 200, 250, and 300 for age 13; and levels 250, 300, and 350 for age 17. These performance levels are the ones most likely to show significant change within an age across the assessment years and do not include the levels that nearly all or almost no students attained at a particular age in each year.

The following description of each mathematics performance level was copied from http://nces.ed.gov/nationsreportcard/ltt/math-descriptions.asp in April 2007.

Level 350: Multistep Problem Solving and Algebra

Students at this level can apply a range of reasoning skills to solve multistep problems. They can solve routine problems involving fractions and percents, recognize properties of basic geometric figures, and work with exponents and square roots. They can solve a variety of two-step problems using variables, identify equivalent algebraic expressions, and solve linear equations and inequalities. They are developing an understanding of functions and coordinate systems.

Level 300: Moderately Complex Procedures and Reasoning

Students at this level are developing an understanding of number systems. They can compute with decimals, simple fractions, and commonly encountered percents. They can identify geometric figures, measure lengths and angles, and calculate areas of rectangles. These students are also able to interpret simple inequalities, evaluate formulas, and solve simple linear equations. They can find averages, make decisions based on information drawn from graphs, and use logical reasoning to solve problems. They are developing the skills to operate with signed numbers, exponents, and square roots.

Level 250: Numerical Operations and Beginning Problem Solving

Students at this level have an initial understanding of the four basic operations. They are able to apply whole number addition and subtraction skills to one-step word problems and money situations. In multiplication, they can find the product of a two-digit and a one-digit number. They can also compare information from graphs and charts, and are developing an ability to analyze simple logical relations.

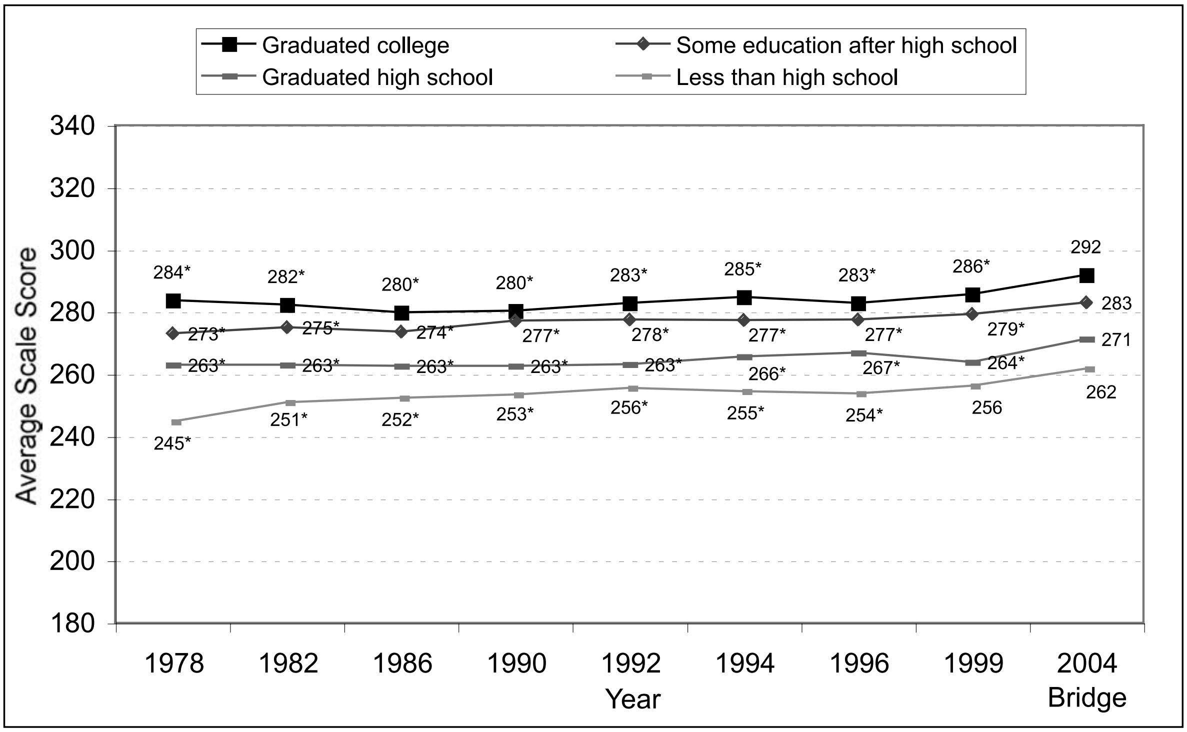
Level 200: Beginning Skills and Understandings

Students at this level have considerable understanding of two-digit numbers. They can add two-digit numbers but are still developing an ability to regroup in subtraction. They know some basic multiplication and division facts, recognize relations among coins, can read information from charts and graphs, and use simple measurement instruments. They are developing some reasoning skills.

Level 150: Simple Arithmetic Facts

Students at this level know some basic addition and subtraction facts, and most can add two-digit numbers without regrouping. They recognize simple situations in which addition and subtraction apply. They also are developing rudimentary classification skills.

Figure E-1: Average NAEP Scale Scores, by Parents' Highest Level of Education, Age 13: Intermittent Years From 1978–2004



*Indicates score is significantly different from 2004.

Note: "Bridge" refers to updates made to NAEP in 2004. The updates replaced outdated material and accommodated more students with disabilities. In order to maintain the long-term trend, test takers were randomly assigned to either the old test form, called the bridge assessment, or the modified test form. Results from the bridge assessment should be compared to results from assessments prior to 2004, while results from the modified assessment should be compared to assessments given after 2004.

Source: Created by the Institute for Defense Analysis Science and Technology Policy Institute using the NAEP Data Explorer (http://nces.ed.gov/nationsreportcard/naepdata/).

What Is This Indicator?

NAEP asks 13- and 17-year-old students to report both of their parents' highest level of education. Parental education level is the background variable on the long-term NAEP that most closely addresses SES. This indicator presents the average scale score of 13-year-old students grouped by the highest level of education attained by either parent.

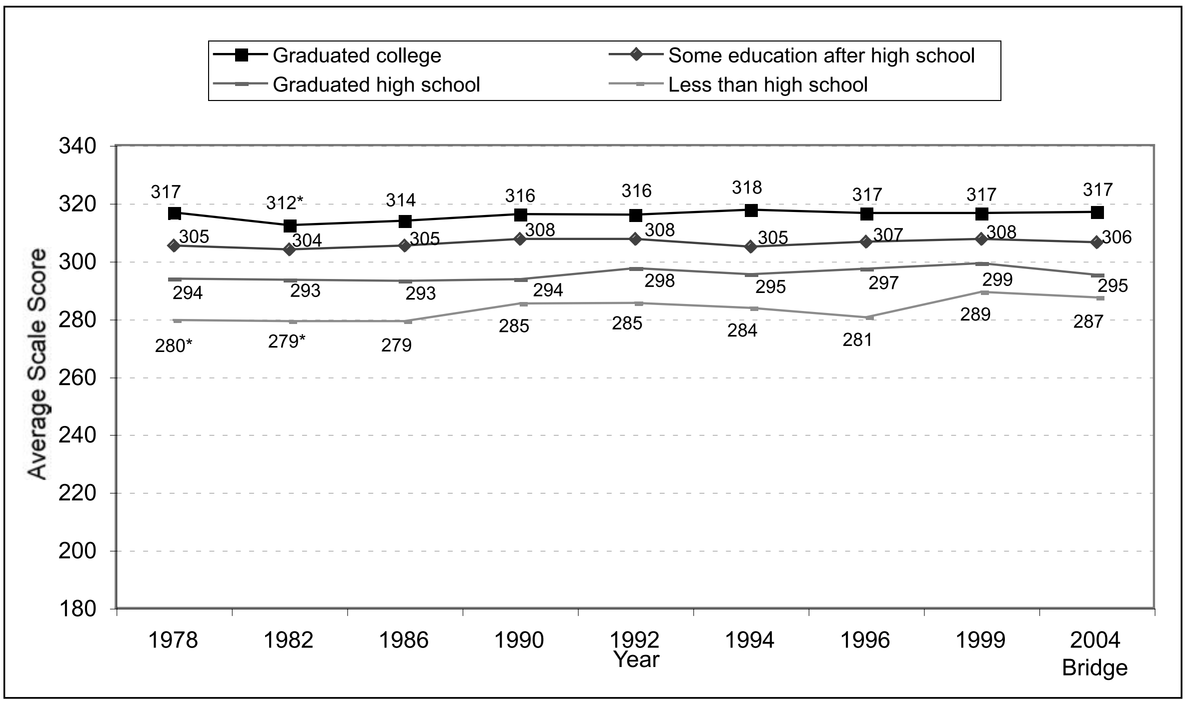
Discussion

- Parents' level of education is directly related to students' average scale score.
- In 2004, 13-year-olds with at least one parent who graduated college scored 30 points higher than students whose parents had less than a high school education. This gap has not changed significantly since 1978.

- In 1978, the gap between 13-year-olds with at least one parent who graduated from college and 13-year-olds whose parents did not complete high school was 39 points. This is a significant difference from 2004.
- The gap between 13-year-olds with at least one parent who graduated from high school and 13-year-olds whose parents did not complete high school has also improved since 1978, decreasing from 18 points in 1978 to 10 points in 2004.

- For 13-year-olds who reported that their parents completed high school, had some education after high school, or completed college, average scores were higher in 2004 than in any previous assessment year.
  - The average score for 13-year-olds whose parents did not finish high school has increased since 1978 but did not change significantly between 1999 and 2004.

Figure E-2: Average NAEP Scale Scores, by Parents' Highest Level of Education, Age 17: Intermittent Years From 1978–2004



*Indicates score is significantly different from 2004

Note: "Bridge" refers to updates made to NAEP in 2004. The updates replaced outdated material and accommodated more students with disabilities. In order to maintain the long-term trend, test takers were randomly assigned to either the old test form, called the bridge assessment, or the modified test form. Results from the bridge assessment should be compared to results from assessments prior to 2004, while results from the modified assessment should be compared to assessments given after 2004.

Source: Created by the Institute for Defense Analysis Science and Technology Policy Institute using the NAEP Data Explorer (http://nces.ed.gov/nationsreportcard/naepdata/).

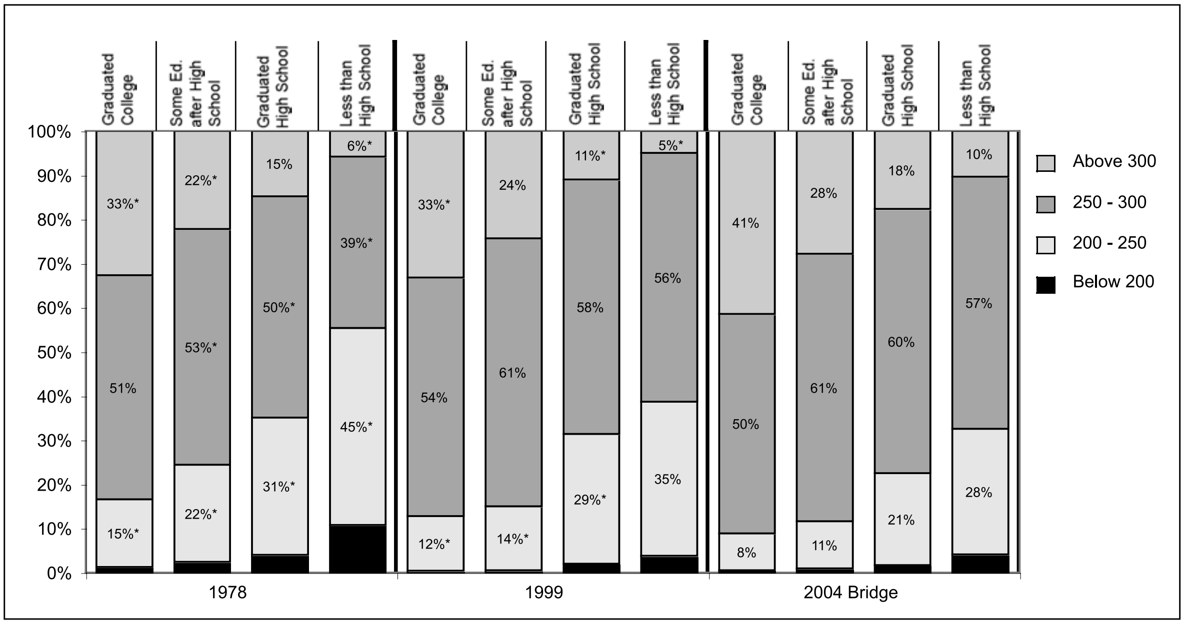
What Is This Indicator?

NAEP asks 13- and 17-year-old students to report both of their parents' highest level of education. Parental education level is the background variable on the long-term NAEP that most closely addresses SES. This indicator presents the average scale score of 17-year-old students grouped by the highest level of education attained by either parent.

Discussion

- Parents' level of education is directly related to students' average scale score.
- In 2004, 17-year-olds with at least one parent who graduated college scored 30 points higher than 17-year-olds whose parents had less than a high school education. This gap has not changed significantly since 1978.
    - — In 1978, the gap between 17-year-olds with at least one parent who graduated from college and 17-year-olds whose parents did not complete high school was 37 points. This is a significant difference from 2004.
    - — The gap between 17-year-olds with at least one parent who graduated from high school and 17-year-olds whose parents did not complete high school has also improved since 1978, decreasing from 14 points in 1978 to 8 points in 2004.
- The average scale score for 17-year-olds at all levels of parental education have generally not changed over the life of the assessment.
    - — The average scale score of 17-year-olds whose parents did not graduate from high school increased from 280 in 1978 to 187 in 2004, but the average scores of all other groups are flat when compared to 1978 and 1999.

Figure E-3: Percent at NAEP Performance Levels, by Parents' Highest Level of Education, Age 13: 1978, 1999, and 2004



*Indicates percentage is significantly different from 2004. Differences between categories may not be statistically significant.

Note: "Bridge" refers to updates made to NAEP in 2004. The updates replaced outdated material and accommodated more students with disabilities. In order to maintain the long-term trend, test takers were randomly assigned to either the old test form, called the bridge assessment, or the modified test form. Results from the bridge assessment should be compared to results from assessments prior to 2004, while results from the modified assessment should be compared to assessments given after 2004.

Source: Created by the Institute for Defense Analysis Science and Technology Policy Institute using the NAEP Data Explorer (http://nces.ed.gov/nationsreportcard/naepdata/).

What Is This Indicator?

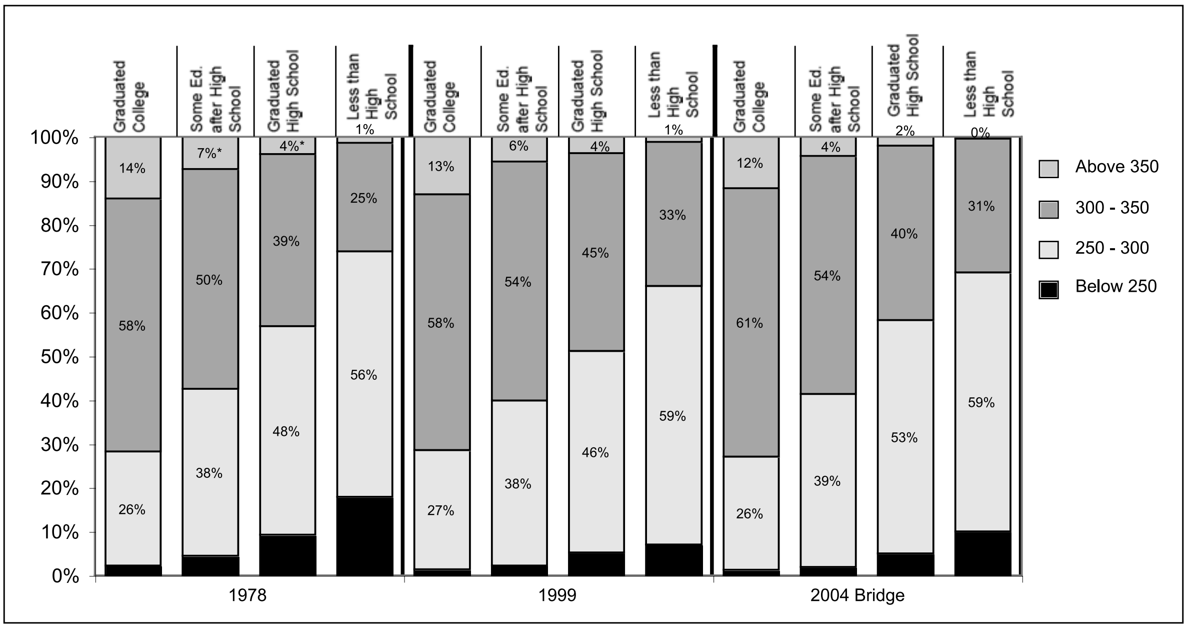
NAEP asks 13- and 17-year-old students to report both of their parents' highest level of education. Parental education level is the background variable on the long-term NAEP that most closely addresses SES. This indicator presents the percentage of 13-year-olds reaching each performance level by parents' highest level of education. The performance levels reported at age 13 are 200—Beginning Skills and Understandings, 250—Numerical Operations and Beginning Problem Solving, and 300—Moderately Complex Procedures and Reasoning.

Discussion

- Higher levels of parental education correlate with a higher percentage of 13-year-olds scoring at or above the 300 level, and a lower percentage of students at or below the 200 level. The effect of parental education on the percentage of 13-year-olds at the 250 level is not significant in most cases.

- While the percentage of students at or above 300 has increased over time for all parental education groups, the gap in achievement between the highest and lowest parental education groups has not changed significantly since 1999 or 1978.
  - — The percentage of 13-year-olds with at least one parent who graduated from college scoring at or above the 300 level increased by 9%, from 33% in 1978 and 1999 to 41% in 2004.
  - — The percentage of 13-year-olds whose parents did not finish high school scoring at or above the 300 level increased by 5%, from 5 to 6% in 1978 and 1999 to 10% in 2004.

Figure E-4: Percent at NAEP Performance Levels, by Parents' Highest Level of Education, Age 17: 1978, 1999, and 2004



*Indicates percentage is significantly different from 2004. Differences between categories may not be statistically significant.

Note: "Bridge" refers to updates made to NAEP in 2004. The updates replaced outdated material and accommodated more students with disabilities. In order to maintain the long-term trend, test takers were randomly assigned to either the old test form, called the bridge assessment, or the modified test form. Results from the bridge assessment should be compared to results from assessments prior to 2004, while results from the modified assessment should be compared to assessments given after 2004.

Source: Created by the Institute for Defense Analysis Science and Technology Policy Institute using the NAEP Data Explorer (http://nces.ed.gov/nationsreportcard/naepdata/).

What Is This Indicator?

NAEP asks 13- and 17-year-old students to report both of their parents' highest level of education. Parental education level is the background variable on the long-term NAEP that most closely addresses SES. This indicator presents the percentage of 17-year-olds reaching each performance level by parents' highest level of education. The performance levels reported at age 17 are 250—Numerical Operations and Beginning Problem Solving, 300—Moderately Complex Procedures and Reasoning, and 350—Multistep Problem Solving and Algebra.

Discussion

- Higher levels of parental education correlate with a higher percentage of 17-year-olds scoring at the 300 and 350 levels, and a lower percentage of students at or below the 250 level.
- The achievement rates of 17-year-olds in all parental education groups and performance levels have not changed since 1999 or 1978.
- Because of the small number of students whose parents did not graduate from high school reaching the 350 level, NAEP does not report statistical significance for comparisons with that subgroup.

Trends in Math and Science Survey: TIMSS

The TIMSS 2003 International Benchmarks of Mathematics Achievement are defined in Mullis et al. (2004, p. 63) as follows:

Grade 8

Advanced International Benchmark – 625

Students can organize information, make generalizations, solve non-routine problems, and draw and justify conclusions from data. They can compute percent change and apply their knowledge of numeric and algebraic concepts and relationships to solve problems. Students can solve simultaneous linear equations and model simple situations algebraically. They can apply their knowledge of measurement and geometry in complex problem situations. They can interpret data from a variety of tables and graphs, including interpolation and extrapolation.

High International Benchmark – 550

Students can apply their understanding and knowledge in a wide variety of relatively complex situations. They can order, relate, and compute fractions and decimals to solve word problems, operate with negative integers, and solve multi-step word problems involving proportions with whole numbers. Students can solve simple algebraic problems including evaluating expressions, solving simultaneous linear equations, and using a formula to determine the value of a variable. Students can find areas and volumes of simple geometric shapes and use knowledge of geometric properties to solve problems. They can solve probability problems and interpret data in a variety of graphs and tables.

Intermediate International Benchmark – 475

Students can apply basic mathematical knowledge in straightforward situations. They can add, subtract, or multiply to solve one-step word problems involving whole numbers and decimals. They can identify representations of common fractions and relative sizes of fractions. They understand simple algebraic relationships and solve linear equations with one variable. They demonstrate understanding of properties of triangles and basic geometric concepts including symmetry and rotation. They recognize basic notions of probability. They can read and interpret graphs, tables, maps, and scales.

Low International Benchmark – 400

Students have some basic mathematical knowledge (Mullis et al., 2004, p.62).

Grade 4

Advanced International Benchmark – 625

Students can apply their understanding and knowledge in a wide variety of relatively complex situations. They demonstrate a developing understanding of fractions and decimals and the relationship between them. They can select appropriate information to solve multistep word problems involving proportions. They can formulate or select a rule for a relationship. They show understanding of area and can use measurement concepts to solve a variety of problems. They show some understanding of rotation. They can organize, interpret, and represent data to solve problems.

High International Benchmark – 550

Student can apply their knowledge and understanding to solve problems. Students can solve multi-step word problems involving addition, multiplication, and division. They can use their understanding of place value and simple fractions to solve problems. They can identify a number sentence that represents situations. Students show understanding of three-dimensional objects, how shapes can make other shapes, and simple transformation in a plane. They demonstrate a variety of measurement skills and can interpret and use data in tables and graphs to solve problems.

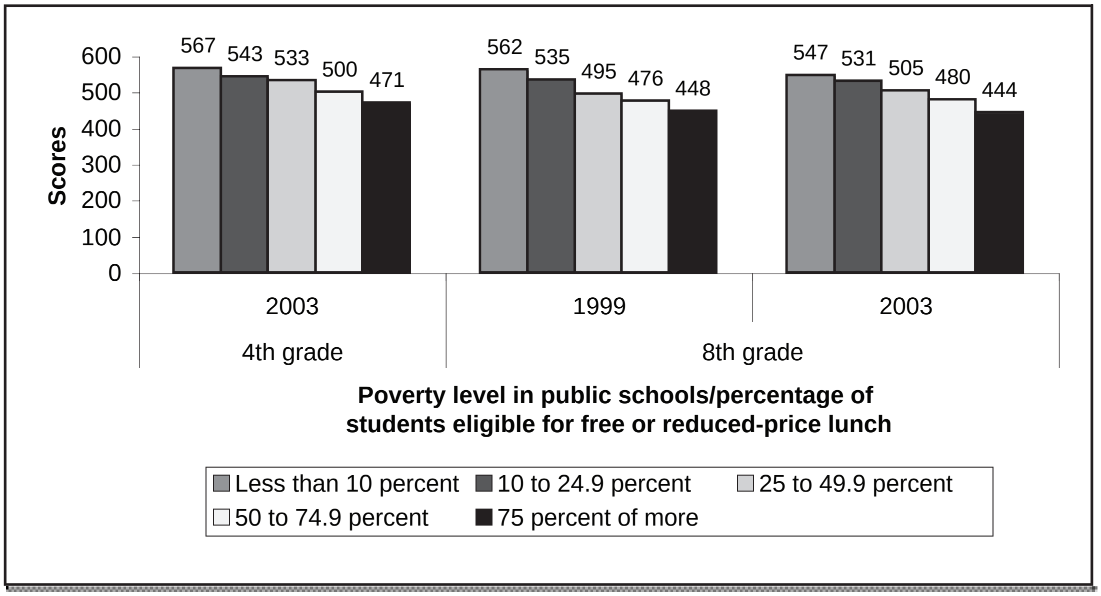
Intermediate International Benchmark – 475

Students can apply basic mathematical knowledge in straightforward situations. They can read, interpret, and use different representations of numbers. They can perform operations with three- and four-digit numbers and decimals. They can extend simple patterns. They are familiar with a range of two-dimensional shapes, and read and interpret different representations of the same data.

Low International Benchmark – 400

Students have some basic mathematical knowledge. Students demonstrate an understanding of whole numbers and can do simple computations with them. They demonstrate familiarity with the basic properties of triangles and rectangles. They can read information from simple bar graphs.

Figure E-5: Average TIMSS Mathematical Scale Scores of U.S. 4th- and 8th-Graders, by School Poverty Level: 1999 and 2003



Source: Gonzales et al., 2004, Tables C8 and C11.

High School and Beyond of 1980: HS&B:80
National Education Longitudinal Study of 1988: NELS:88
Education Longitudinal Study of 2002: ELS:2002

The scores on the HS&B:80, NELS:88, and ELS:2002 are Item Response Theory (IRT) number-right scores on the NELS:88 1990 58-item scale. IRT estimates achievement based on patterns of correct, incorrect, and unanswered questions. "The IRT-estimated number-right score reflects an estimate of the number of these 58 items that an examinee would have answered correctly if he or she had taken all of the items that appeared on the multiform 1990 NELS:88 mathematics test. The score is the probability of a correct answer on each item, summed over the total mathematics 58-item pool" (Cahalan et al., 2006, p.45). These scores are not directly translated into probability of proficiency scores. However, five probability of proficiency scores in mathematics were estimated for students using performance on clusters of four items each as follows:

Probability of Mastery, Mathematics Levels

1) Simple arithmetical operations on whole numbers, such as simple arithmetic expressions involving multiplication or division of integers;

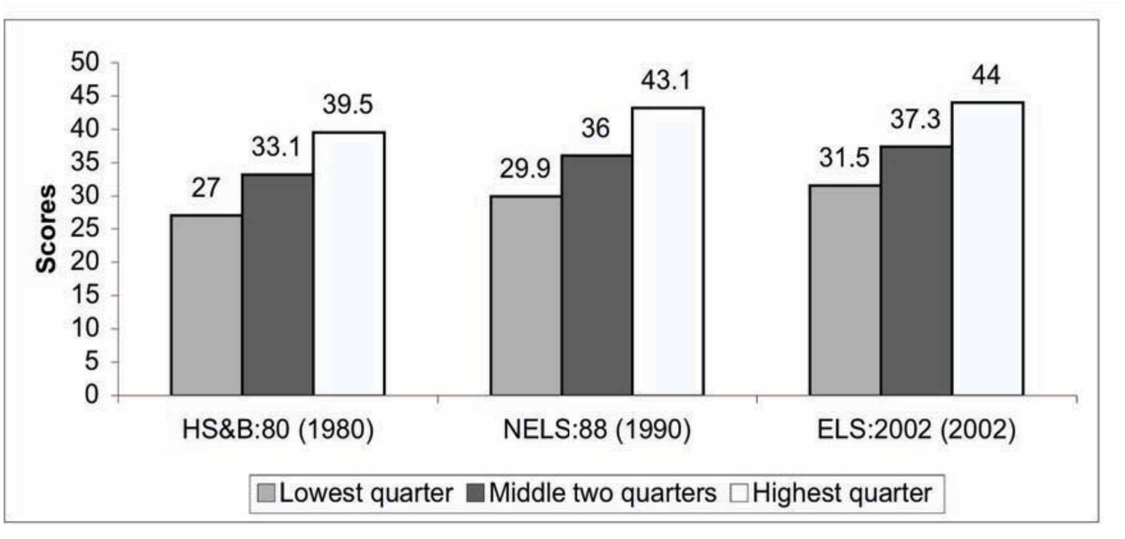
2) Simple operations with decimals, fractions, powers, and roots, such as comparing expressions, given information about exponents;

3) Simple problem solving, requiring the understanding of low-level mathematical concepts, such as simplifying an algebraic expression or comparing the length of line segments illustrated in a diagram;

4) Understanding of intermediate-level mathematical concepts and/or multistep solutions to word problems such as drawing an inference based on an algebraic expression or inequality; and

5) Complex multistep word problems and/or advanced mathematics material such as a two-step problem requiring evaluation of functions. (Cahalan et al., 2006, p. A-28)

Figure E-6: IRT—Estimated Average Math Score (10th Grade), by SES (HS&B:80, NELS:88, ELS:2002)



Source: Cahalan et al. (2006), Tables 18 and 19.

Table E-1: Probability of 10th-Grade Proficiency in Mathematics, by SES

	NELS:88 (1990)	ELS:2002 (2002)
Level 1		
Lowest quarter	83.1	84.5
Middle quarters	91.1	92.5
Highest quarter	97.1	97.1
Level 2		
Lowest quarter	41.3	46.4
Middle quarters	62.6	67.8
Highest quarter	83.3	86.2
Level 3		
Lowest quarter	20.4	25.1
Middle quarters	41.4	44.7
Highest quarter	67.4	70.9
Level 4		
Lowest quarter	5.7	7.6
Middle quarters	15.9	17.7
Highest quarter	36.2	38.7
Level 5		
Lowest quarter	0.1	0.2
Middle quarters	0.2	0.5
Highest quarter	1.0	2.6

Note: Proficiency levels —1) Simple arithmetical operations with whole numbers; 2) Simple operations with decimals, fractions, powers, and roots; 3) Simple problem solving, requiring the understanding of low-level mathematical concepts; 4) Understanding of intermediate-level mathematical concepts and/or multistep solutions to word problems; and 5) Complex multistep word problems and/or advanced mathematics material.

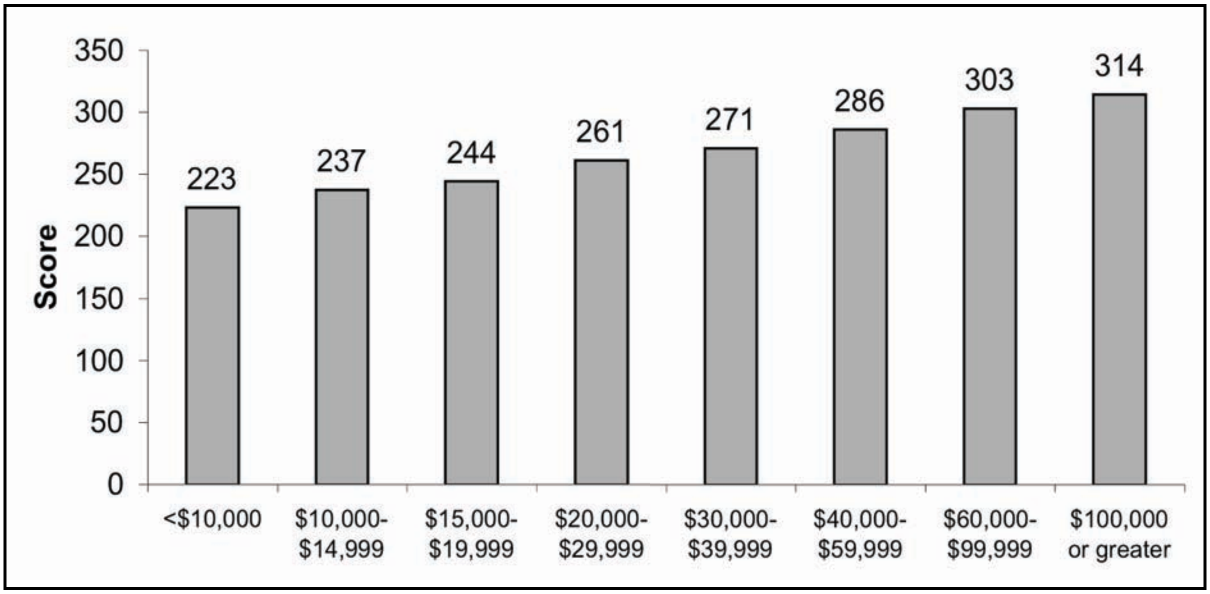
Source: Cahalan et al., 2006, p. 57-58.

National Adult Literacy Survey: NALS
National Assessment of Adult Literacy: NAAL

The Committee on Performance Levels for Adult Literacy set performance levels for quantitative literacy as Below Basic, Basic, Intermediate, and Proficient and defined them as follows, based on scores on NALS and NAAL:

> ***Below Basic (0–234)*** indicates no more than the most simple and concrete literacy skills.
>
> *Key abilities*—locating numbers and using them to perform simple *quantitative* operations (primarily addition) when the mathematical information is very concrete and familiar.
>
> ***Basic (235–289)*** indicates skills necessary to perform simple and everyday literacy activities.
>
> *Key abilities*—locating easily identifiable *quantitative* information and using it to solve simple, one-step problems when the arithmetic operation is specified or easily inferred.
>
> ***Intermediate (290–349)*** indicates skills necessary to perform moderately challenging literacy activities.
>
> *Key abilities*—locating less familiar *quantitative* information and using it to solve problems when the arithmetic operation is not specified or easily inferred.
>
> ***Proficient (350–500)*** indicates skills necessary to perform more complex and challenging literacy activities.
>
> *Key abilities*—locating more abstract quantitative information and using it to solve multistep problems when the arithmetic operations are not easily inferred and the problems are more complex (Kutner et al., 2006, p. 3).

Figure E-7: Average Quantitative Literacy Scores of Adults, by Household Income: NAAL 2003



Note: Literacy levels: Below basic 0–234, Basic 235–289, Intermediate 290–349, Proficient 350–500

Source: Kutner et al. (2007), Figure 2-17.

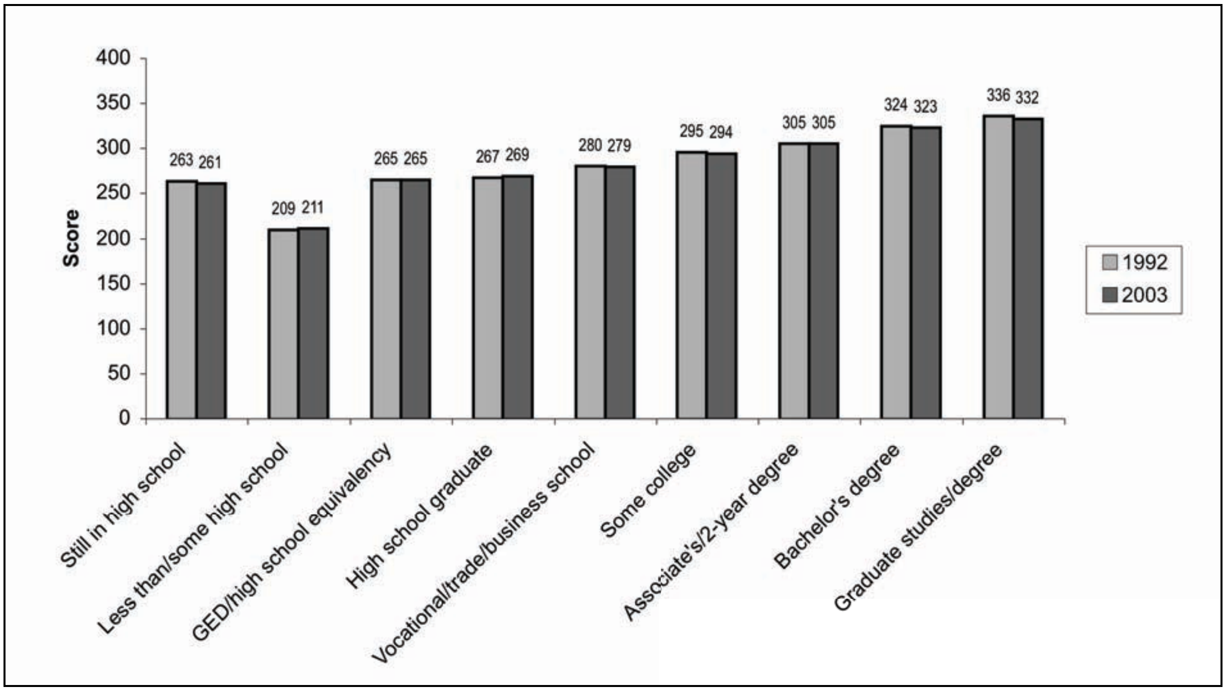
Table E-2: Percentage of Adults in Each Quantitative Literacy Level, by Household Income: NAAL 2003

	<$10,000	$10,000–$14,999	$15,000–$19,999	$20,000–$29,999	$30,000–$39,999	$40,000–$59,999	$60,000–$99,999	$100,000 or greater
Below Basic	26	16	11	16	11	12	7	2
Basic	9	8	6	14	14	21	19	9
Intermediate	4	4	3	10	11	22	28	18
Proficient	2	2	2	5	6	18	37	29

Note: Below Basic (0–234) no more than the most simple and concrete literacy skills; Basic (235–289) skills necessary to perform simple and everyday literacy activities; Intermediate (290–349) skills necessary to perform moderately challenging literacy activities; Proficient (350–500) skills necessary to perform more complex and challenging literacy activities.

Source: Kutner et al. (2007), Table 2-3.

Figure E-8: Average Quantitative Literacy Scores of Adults, by Highest Educational Attainment: NALS 1992 and NAAL 2003



Source: Kutner et al., 2007, Table 3-2.

Table E-3: Percentage of Adults in Each Quantitative Literacy Level, by Highest Education Attainment: NALS 1992 and NAAL 2003

	Still in high school	**Less than/ some high school**	**GED/ high school equivalency**	**High school graduate**	**Vocational/ trade/ business school**	**Some college**	**Associate's/ 2-year degree**	**Bachelor's degree**	**Graduate studies/ degree**
1992	(NALS)								
Below Basic	31	65	25	26	18	11	8	5	2
Basic	37	25	46	41	39	34	29	21	15
Intermediate	27	9	26	29	35	42	45	44	43
Proficient	6	1	3	5	8	13	18	31	39
2003	(NAAL)								
Below Basic	31	64	26	24	18	10	7	4	3
Basic	38	25	43	42	41	36	30	22	18
Intermediate	25	10	28	29	35	43	45	43	43
Proficient	5	1	3	5	6	11	18	31	36

Source: Kutner et al., 2007, Figure 3-1c.

Program for International Student Assessment: PISA

Mathematics literacy can be classified by proficiency levels, based on scores on the PISA, as follows:

Level 1 (*greater than 357.77 to 420.07*) At Level 1, students can answer questions involving familiar contexts where all relevant information is present and the questions are clearly defined. They are able to identify information and to carry out routine procedures according to direct instructions in explicit situations. They can perform actions that are obvious and follow immediately from the given stimuli.

Level 2 (*greater than 420.07 to 482.38*) At Level 2, students can interpret and recognize situations in contexts that require no more than direct inference. They can extract relevant information from a single source and make use of a single representational mode. Students at this level can employ basic algorithms, formula, procedures, or conventions. They are capable of direct reasoning and making literal interpretations of the results.

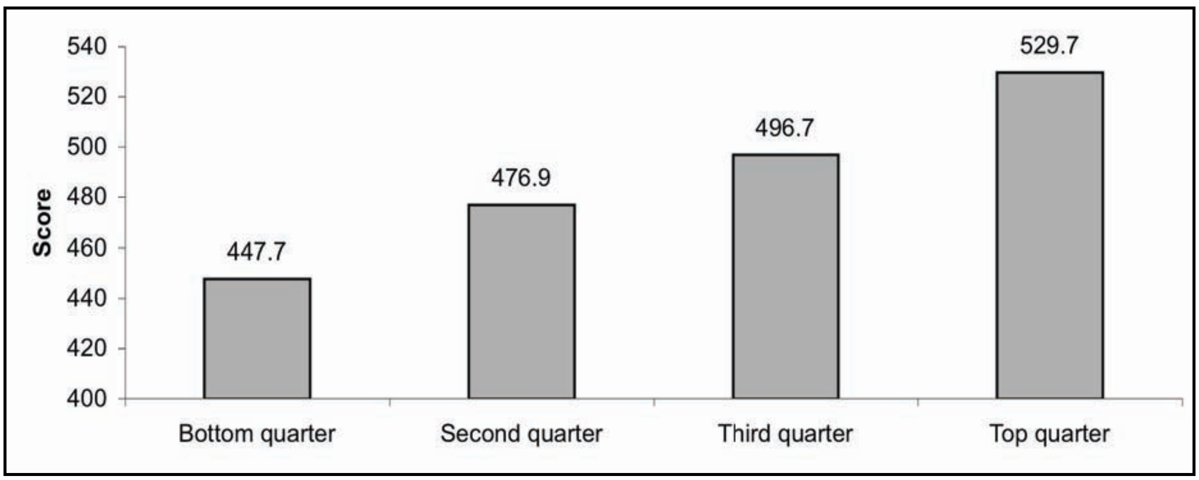
Level 3 (*greater than 482.38 to 544.68*) At Level 3, students can execute clearly described procedures, including those that require sequential decisions. They can select and apply simple problem solving strategies. Students at this level can interpret and use representations based on different information sources and reason directly from them. They can develop short communications reporting their interpretations, results, and reasoning.

Level 4 (*greater than 544.68 to 606.99*) At Level 4, students can work effectively with explicit models for complex concrete situations that may involve constraints or call for making assumptions. They can select and integrate different representations, including symbolic, linking them directly to aspects of real-world situations. Students at this level can utilize well developed skills and reason flexibly, with some insight, in these contexts. They can construct and communicate explanations and arguments based on their interpretations, arguments, and actions.

Level 5 (*greater than 606.99 to 669.3*) At Level 5, students can develop and work with models for complex situations, identifying constraints and specifying assumptions. They can select, compare, and evaluate appropriate problem solving strategies for dealing with complex problems related to these models. Students at this level can work strategically using broad, well-developed thinking and reasoning skills, appropriate linked representations, symbolic and formal characterizations, and insight pertaining to these situations. They can reflect on their actions and formulate and communicate their interpretations and reasoning.

Level 6 (*greater than 669.3*) At Level 6, students can conceptualize, generalize, and utilize information based on their investigations and modeling of complex problem situations. They can link different information sources and representations and flexibly translate among them. Students at this level are capable of advanced mathematical thinking and reasoning. These students can apply this insight and understandings along with a mastery of symbolic and formal mathematical operations and relationships to develop new approaches and strategies for attacking novel situations. Students at this level can formulate and precisely communicate their actions and reflections regarding their findings, interpretations, arguments, and the appropriateness of these to the original situations (Lemke et al., 2005, p.18).

Figure E-9: Average Mathematics Literacy Scores of U.S. 15-Year-Olds, by Quarters on the International Socioeconomic Index: 2003 PISA



Note: Level 1 (greater than 357.77 to 420.07), Level 2 (greater than 420.07 to 482.38), Level 3 (greater than 482.38 to 544.68), Level 4 (greater than 544.68 to 606.99), Level 5 (greater than 606.99 to 669.3), Level 6 (greater than 669.3)

Source: Lemke et al., 2005, Tables B-24.

Chapter 5: Report of the Task Group on Teachers and Teacher Education

Deborah Loewenberg Ball, Chair
James Simons
Hung-Hsi Wu
Raymond Simon, Ex Officio
Grover J. "Russ" Whitehurst, Ex Officio
Jim Yun, U.S. Department of Education Staff

CONTENTS

Tables

Abbreviations

AIMS PK–16	The Alliance for Improvement of Mathematics Skills—PreKindergarten Through Grade 16
ANOVA	Analysis of Variance
BA	Bachelor of Arts
BNSE	Belize National Selection Exam
BSAP	Basic Skills Assessment Program
CCTDM	Classroom Center Teacher Developmental Mathematics
CKT-M	Content Knowledge for Teaching Mathematics
CSAP	Colorado Student Assessment Program
CTBS	Comprehensive Test of Basic Skills
ETS	Education Testing Service
FCAT	Florida Comprehensive Achievement Test
GLS	Generalized Least Squares
GPA	Grade Point Average
GSCE	Graduate School College Education
HLM	Hierarchical Linear Modeling
HS	High School
ICC	Intra-Class Correlation
IMA	Integrated Mathematics Assessment
IP	Incentive Plan
ISD	Independent School District
ITBS	Iowa Test of Basic Skills
KCPE	Kentucky Council on Postsecondary Education
LSAY	Longitudinal Study of American Youth
MA	Master of Arts
MAT	Master of Arts in Teaching
MMI	Maysville Mathematics Initiative
NAEP	National Assessment of Educational Progress
NBPTS	National Board for Professional Teaching Standards
NCATE	National Council for the Accreditation of Teacher Education
NCE	Norm Curve Equivalent
NELS	National Educational Longitudinal Survey
NWEA	Northwest Evaluation Association Study
OLS	Ordinary Least Squares
PFP	Pay for Performance
PK–16	Pre-School Through Grade 16
PPST	Pre-Professional Skills Tests
PS	Propensity Score
RCT	Randomized Controlled Trials
RDD	Regression Discontinuity Design
SAT	Scholastic Aptitude Test
SAT-9	Stanford Achievement Test, Ninth Edition
SC	South Carolina

SSCI	Social Sciences Citation Index
STAR	Student/Teacher Achievement Ratio
SUPP	Supplemental
TAAS	Texas Assessment of Academic Skills
TFA	Teach for America
TIMSS	Trends in Mathematics and Science Study
TRAD	Traditional
WCET	Weighted Common Examinations Total
WWC	What Works Clearinghouse

Executive Summary

Introduction

Teachers are crucial to students' opportunities to learn mathematics, and substantial differences in the mathematics achievement of students are attributable to differences among teachers. Therefore, the National Mathematics Advisory Panel (Panel) was charged with making recommendations, based on the best available scientific evidence, on "the training, selection, placement, and professional development of teachers of mathematics in order to enhance students' learning of mathematics," according to presidential Executive Order 13398. To address this charge, the Panel established a Teachers and Teacher Education Task Group (Task Group), which combed the research on the relationship between teachers' own knowledge and students' achievement, and how effective teachers can best be recruited, prepared, supported, and rewarded.

The four questions that structured the work of the Task Group are:

1) What is the relationship between the depth and quality of teachers' mathematical knowledge and students' mathematics achievement?

2) What is known about programs that help teachers develop the necessary mathematical knowledge for teaching? Which of these programs have been shown to affect instructional practice and student achievement?

3) What types of recruitment and retention strategies are used to attract and retain highly effective teachers of mathematics? How effective are they?

4) What evidence exists for the effectiveness of elementary mathematics specialist teachers with respect to student achievement? What models exist for elementary mathematics specialists and their preparation?

The research base that addresses these questions was found to be uneven. Therefore, in addition to making recommendations based on the best available evidence, the Task Group makes recommendations about the research needed in order to improve practice and policy with respect to teachers.

Methodology

The Task Group identified available scientific evidence, published in peer-reviewed journals and national reports, and categorized the quality of the research studies related to each of the four research questions. Studies were categorized as high quality, moderate quality, or lesser quality based on the appropriateness of the design in answering the specific research question. High-quality studies were those that employed a randomized control trial design or those that addressed the weaknesses of a correlational design through the use of large samples, control variables, multiple specifications, etc. Standardized regression coefficients or standardized mean differences were calculated as appropriate. High-quality studies served as the primary basis for the Task Group's recommendations, although the available evidence varied for each research question. Because of the paucity of rigorous empirical research to answer the Task Group's fourth question, the Task Group provides a description of the programs and models that exist in the United States, some distinctions among the different models, and commentary on the costs and benefits of those different models.

Teachers' Knowledge of Mathematics

Common sense dictates that teachers must know the mathematical content they teach, but defining a precise body of mathematical knowledge that would effectively serve teachers and would guide teacher education, professional development, and policy has proved challenging. Therefore, the Task Group considered the empirical evidence on how the level and scope of teachers' own mathematical knowledge affects the learning of their students. Across the rigorous research studies reviewed on the relationship between teachers' mathematical knowledge and their students' achievement, teacher content knowledge was measured in three different ways: teacher certification, mathematics course work, and tests of teachers' mathematical knowledge. Across all studies, the general results are mixed but overall do confirm the importance of teachers' content knowledge. However, because most studies have relied on proxies for teachers' mathematical knowledge (such as teacher certification or courses taken), existing research does not reveal the specific mathematical knowledge and instructional skill needed for effective teaching, especially at the elementary and middle school level. Direct assessments of teachers' actual mathematical knowledge provide the strongest indication of a relation between teachers' content knowledge and their students' achievement.

Teacher Certification as a Measure of Mathematical Content Knowledge

Across studies that used teacher certification or teacher certification in mathematics as a proxy for teachers' mathematical content knowledge, findings were mixed about the impact of teacher certification on student achievement in mathematics. Research in this area has not provided consistent or convincing evidence that students of teachers who are certified in mathematics gain more than those whose teachers are not.

Content Course Work and Degrees as Measures of Mathematical Content Knowledge

Mathematics course work and field-specific degrees are a second common group of proxies for teachers' mathematical content knowledge; both measures focus on teachers' completion of college-level mathematics study and are often jointly considered within the same analysis. In general, although results are mixed, there appears to be some positive relationship between teacher content course work and degrees, and student math achievement at the high school level. However, the existing research does not show evidence of such a relationship below ninth grade.

Test Scores and Ad Hoc Measures as Measures of Mathematical Content Knowledge

More proximal measures of teachers' mathematical content knowledge included tests of teachers' content knowledge. Such measures allow closer examination of the effect that mathematical knowledge has on student achievement. Although there was variation among the set of studies that used teacher test scores as measures of teacher content knowledge, overall these studies signaled a positive effect of mathematical content knowledge on student achievement. It should be noted that the studies this Task Group found were focused at the elementary level, making comparisons with other findings difficult.

The Mathematical Content and Nature of Teacher Licensure Exams

Although recent research treating teacher licensure as a proxy for teachers' mathematical content knowledge has not consistently or convincingly shown that students of teachers who are licensed in mathematics gain more academically than those whose teachers are not, teacher licensure exams play an important role in determining the math teachers available for employment in schools. Recognizing the importance of teacher licensure exams, the Task Group sought access to these exams together with data on teachers' performance on exam items. Due to issues of confidentiality, however, the Task Group was not able to gather information sufficient to assess the mathematical quality of these exams.

Conclusions and Recommendations Regarding Teachers' Knowledge of Mathematics

Research on the relationship between teachers' mathematical knowledge and students' achievement supports the importance of teachers' content knowledge in student learning. However, because most studies have relied on proxies for teachers' mathematical knowledge (such as teacher certification or courses taken), existing research does not reveal the specific mathematical knowledge and instructional skills needed for effective teaching, especially at the elementary and middle school level.

Direct assessments of teachers' actual mathematical knowledge provide the strongest indication of a relation between teachers' content knowledge and their students' achievement. More precise measures are needed to specify in greater detail the relationship among elementary and middle-school teachers' mathematical knowledge, their instructional skill, and students' learning.

Teachers' Education: Teacher Preparation, Alternative Pathways to Teaching, Induction, and Professional Development

Teaching well requires substantial knowledge and skill. The Task Group sought to understand the impact that teacher education has on teachers' effectiveness. The Task Group considered the empirical evidence on four types of professional training:

- *Preservice teacher preparation*: Initial teacher training, conventionally offered in institutions of higher education;
- *Alternative pathways*: Initial teacher preparation offered outside of conventional teacher education programs*;*
- *Induction programs*: Professional support and additional training within the first years of practice; and
- *Professional development*: Ongoing programmatic professional education of practicing teachers.

The Task Group's review of the available research and the rigor of this research highlights the critical need for more and better studies tracing the relationship between specific approaches to teacher education (i.e., curricula, pedagogy and assessment, instructors, structures, and settings) and teachers' capacity for teaching and their students' learning. However, existing research on aspects of teacher education, including standard teacher preparation programs, alternative pathways into teaching, support programs for new teachers (e.g., mentoring), and professional development, is not of sufficient rigor or quality to permit the Panel to draw conclusions about the features of professional training that have effects on teachers' knowledge, their instructional practice, or their students' achievement.

Preservice Teacher Preparation

The Task Group found very few empirical studies that addressed the question of the impact of preparation programs on student achievement or teachers' mathematical content knowledge. Unfortunately, none were of sufficient rigor or quality to allow us to draw conclusions about the relationship of particular features of teacher preparation programs and their effects.

Alternative Pathways Into Teaching

Determining how different types of pathways into teaching may affect the knowledge and skill teachers' bring to their work is a key policy question in a time when issues of teacher recruitment, retention, and quality are paramount. Overall, evidence is mixed on the effects of teachers' pathways into teaching and their relationship to students' achievement. Because of differences in specifications across studies, drawing specific conclusions about alternative pathways in general would be difficult to do from these studies. However, the evidence suggests that the differences among current pathways are small or not significantly different and that variation within programs is much greater than that across programs.

Induction Programs

National reports calling for higher-quality teaching, higher teacher retention rates, and stronger student achievement identify support of new teachers—induction—as an area for improvement. The Task Group found a dearth of peer-reviewed research on early career support programs for mathematics teachers that looked at outcomes related to student achievement or teacher mathematics knowledge; key outcomes for much of the current induction literature is teacher retention, satisfaction, and beliefs. Given the current expansion of induction programs, it is important to assess the effectiveness of induction programs on teachers' effectiveness. Until induction programs are content-specific or include specific content activities, it may be difficult to determine the effectiveness of induction on the mathematical knowledge of teachers or on students' achievement gains.

Professional Development Programs

Practicing teachers continue their teacher education through in-service professional development. Studies that investigated the impact on teacher content knowledge lacked either comparison groups or direct measures of teacher content knowledge, instead relying on teacher self reports. Across studies that investigated the relationship between teacher professional development programs and students' mathematics achievement, there was a positive effect of professional development on student achievement. However, there was insufficient evidence on the specific features of professional development that impact student achievement to make conclusions about what forms of, or approaches to, professional development are effective.

Conclusions and Recommendations Regarding Teachers' Education

Despite common beliefs about qualities of effective teacher education, there is no strong evidence on the relationship between any specific form of teacher education and either teachers' mathematics content knowledge or their students' mathematics learning. Further, there is even less evidence to identify the specific features of training that might account for any program impacts, providing little insight into the crucial components of teacher education.

High-quality research must be undertaken to create a sound basis for the mathematics preparation of elementary and middle school teachers within preservice teacher education, early-career support, and ongoing professional development programs. Outcomes of different approaches should be evaluated, in part, by using reliable and valid measures of mathematical knowledge that are demonstrably associated with gains in student achievement.

Studies are needed with designs that lead to knowledge about the impact of different approaches to professional development, and that permit comparisons with other potential impacts on teacher capacity and effectiveness (e.g., experience, curriculum, curriculum policy). Such research will depend not only on rigorous designs, but also on valid and reliable measures of the key outcome variables: teachers' mathematical knowledge and skill, instructional quality, and student learning.

Key questions on which robust evidence is needed include:

- Does teacher education (including preservice training of different kinds, professional development, and early career induction programs), have an impact on teachers' capacity for teaching and on their students' achievement?
- What are key features of teacher education (e.g., duration, structure, quantity, content, pedagogy, structure, relationship to practice) that have effects on teacher capacity and on student achievement?
- How do contexts (school, students, teachers, policy) affect the outcomes of professional development?
- How do different amounts of teacher education affect outcomes and effects?

Teacher Incentives

Compensation is often cited as a key factor in improving teacher quality, and programs are focusing on compensation as a recruitment and retention strategy. The Task Group investigated the evidence on different salary schemes that aim to recruit, reward, and retain skillful teachers. The programs utilize a variety of labor market incentives including:

- *Skills-based pay:* Paying more to teachers who have technical skills that are in demand in other sectors of the economy, such as teachers with degrees in mathematics;
- *Location pay:* Compensating teachers for working in conditions they view as unfavorable, such as those associated with high-poverty, low-achieving schools; and
- *Performance-based pay:* Paying more to mathematics teachers who are more productive in raising student achievement.

The Task Group examined research on each of these approaches to teacher compensation.

Skills-Based Pay

College students' decisions to prepare for and enter into teaching depend on how the salary structure for teachers compares with those in competing occupations. The magnitude of the salary differential between the private sector and the teaching profession for those who enter teaching with technical training is large, with a moderate difference on entry and a rapidly increasing gap over the first 10 years of employment. Teachers of mathematics and science are significantly more likely to move from or leave their teaching jobs because of job dissatisfaction than are other teachers, and of those who depart because of job dissatisfaction, the most common reason given is low salary.

Location Pay

Research on the effects of location-based pay, intended to attract or retain skilled teachers in schools that serve under-resourced communities, yielded mixed results. The effectiveness of such salary schemes is affected by the size of differential in pay, the gender and experience of the teacher, whether the bonus is a one-time signing bonus or permanent, as well as other factors.

Performance Pay

There is a substantial variability in extant merit-pay systems. The Task Group identified four different dimensions of variability of "merit" pay: whether salary differentials are tied to schools' performance or that of individual teachers, how significant the differential is, the degree to which the scheme is focused on student performance, and whether the plan seems continuous or is a short-term experiment. Across the studies reviewed, each performance pay approach yielded some positive effects on student achievement.

Recommendations and Cautions Regarding Teacher Incentives

The results from research on teacher incentives generally support the effectiveness of incentives, although the methodological quality of the studies in terms of causal conclusions is mixed. The substantial body of economic research in other fields indicating that salary affects the number of workers entering a field and their job performance is relevant. In the context of the totality of the evidence, the Task Group recommends policy initiatives that put in place and carefully evaluate the effects of:

- Raising base salaries for teachers of mathematics to be more competitive with salaries for similarly trained non-teachers;
- Incentives for teachers of mathematics working in locations that are difficult to staff; and
- Opportunities for teachers of mathematics to increase their base salaries substantially by demonstrable effectiveness in raising student achievement.

Knowing more about how various incentive systems affect teachers would enable the design of more effective and efficient incentives.

Elementary Mathematics Specialist Teachers

There have been many calls for the use of "mathematics specialist teachers" at the K–5 level in recent years, with some arguing that the teaching of mathematics even in the elementary grades calls for specialized knowledge, while most elementary teachers are generalists. The Task Group sought to learn what is known about mathematics specialists at the elementary level. To contribute to thoughtful consideration of the issues involved in restructuring teacher roles around the idea of mathematics specialists, the Task Group reviewed a range of models in current use in the United States and abroad, and sought evidence about their effectiveness.

Models of Elementary Mathematics Specialist Teachers

The Task Group identified three models of mathematics specialist teachers: the lead teacher or mathematics coach, who is a resource person for their coworkers and does not directly instruct students; the specialized teacher, who is responsible for the direct instruction of students; and the pull-out mathematics specialist, who directly instructs individuals or small groups of students within a classroom who have been identified as either failing to meet or exceeding their grade-level standards. Mathematics specialists as lead teachers or mathematics coaches are more common than the other two models.

Evidence for Effectiveness of Elementary Mathematics Specialist Teachers

The Task Group identified very few studies that probed the effectiveness of elementary mathematics specialists. Out of 114 potentially relevant pieces of literature, only one explored the effects of specialized mathematics specialists on student achievement in elementary schools. The study found no difference in the mathematics gain scores of students in an elementary school with a departmentalized structure compared to students in a school with a self-contained class structure.

Costs Associated With Mathematics Specialists

The costs associated with the employment, training, or certification of mathematics specialists were considered. One cost has to do with the funding of the personnel involved and depends on the model used: the specialized-teacher model involves only a redistribution of responsibilities among the existing staff, whereas the lead teacher or mathematics coach or the pull-out program requires the hiring of additional teachers. A second cost is that of the additional training needed for teachers to prepare with the specialization to fill these roles. In addition to tuition costs for participating teachers, there are costs associated with developing and operating programs or courses. Simply taking more college-level mathematics courses would not be sufficient in most cases because regular mathematics courses generally do not focus on the mathematics needed for specialization in elementary and middle school.

Mathematics Specialists Internationally

Full-time elementary mathematics teachers are not widely used in most of the countries that produce high levels of student achievement in mathematics. Only three (China, Singapore, and Sweden) deploy such teachers at the elementary level. Elementary teachers in those countries may enter teaching with a stronger background in mathematics, which may be a factor in the success of those countries with mathematics education.

Conclusions Regarding Elementary Mathematics Specialist Teachers

The Panel recommends that research be conducted on the use of full-time mathematics teachers in elementary schools. These would be teachers with strong knowledge of mathematics who would teach mathematics full-time to several classrooms of students, rather than teaching many subjects to one class, as is typical in most elementary classrooms. This recommendation for research is based on the Panel's findings about the importance of teachers' mathematical knowledge. The use of teachers who have specialized in elementary mathematics teaching could be a practical alternative to increasing all elementary teachers' content knowledge (a problem of huge scale) by focusing the need for expertise on fewer teachers.

I. Introduction

Teachers are crucial to students' opportunities to learn and to their learning of mathematics. Substantial differences in the mathematics achievement of students are attributable to differences in teachers. A meta-analysis of the findings from seven large studies of variation in teacher effects found that 11% of the total variability in student achievement gains in mathematics across one year of classroom instruction was attributable to teachers (Nye, Konstantopoulos, & Hedges, 2004). The authors, noting that all the studies in their review were correlational, conducted a new analysis using data from the Tennessee class-size experiment, also known as the Student-Teacher Achievement Ratio (STAR) project. An analysis of STAR data has a methodological advantage over prior studies of natural variation in that both students and teachers were randomly assigned to classes. Nye et al. found that differences in teachers accounted for 12% to 14% of total variability in students' mathematics achievement gains in each of Grades 1, 2, and 3.

In a similar vein, Gordon, Kane, and Staiger (2006), using data from the Los Angeles Unified School District for teachers in Grades 3 through 5, report that the average student assigned to a teacher in his or her 3rd year of teaching who was in the bottom quartile during his or her first 2 years of teaching lost on average 5 percentile points on a mathematics assessment relative to students with similar baseline scores and demographics. In contrast, the average student assigned to a top-quartile teacher gained 5 percentile points relative to students with similar baseline scores and demographics. Thus, the average difference between being assigned a top-quartile or a bottom-quartile teacher was 10 percentile points.

These are large effects, larger, for example, than those that have been shown to result from significant reductions in class size. Important to note is that these effects are for one year of instruction. Teacher effects are much larger when they combine across years of instruction. Sanders and Rivers (1996) used value-added methods to measure the effectiveness of all math teachers in Grades 3, 4, and 5 in two large metropolitan school districts in Tennessee. The growth in academic achievement by students in each teacher's class relative to all other teachers was used to identify the most effective (top 20%) and the least effective (bottom 20%) teachers. The progress of children assigned to these low- and high-performing teachers was tracked over a 3-year period. Children assigned to three effective teachers in a row scored at the 83rd percentile in mathematics at the end of fifth grade, while children assigned to three ineffective teachers in a row scored at the 29th percentile. Using a different methodology, Rivkin, Hanushek, and Kain (2005) came to a similar conclusion: The cumulative effects of children being taught by highly effective teachers can substantially eliminate differences in student achievement that are due to family background.

Unfortunately, little is known about what accounts for these individual differences in teachers' ability to generate gains in student learning. Investigating the best evidence about instructional practices that affect achievement is the province of the Instructional Practices Task Group. The Teachers and Teacher Education Task Group has sought to learn the impact of a teacher's own knowledge on their students' achievement and how teachers can be best recruited, prepared, supported, and rewarded.

The four questions that structured the work of the Teachers and Teacher Education Task Group are:

1) What is the relationship between the depth and quality of teachers' mathematical knowledge, and students' mathematics achievement?

2) What is known about programs that help teachers develop the necessary mathematical knowledge for teaching? Which of these programs have been shown to affect instructional practice and student achievement?

3) What types of recruitment and retention strategies are used to attract and retain highly effective teachers of mathematics? How effective are the strategies?

4) What evidence exists for the effectiveness of elementary mathematics specialist teachers with respect to student achievement? What models exist for elementary mathematics specialists and their preparation?

The Task Group found an uneven research base to address these questions. Included in this report are the Task Group's summary and recommendations, their observations about the nature of the empirical evidence currently available, and recommendations for the research needed to improve practice and policy with respect to teachers. The Task Group's last question, on elementary mathematics specialist teachers, differs from the others as it relies primarily on anecdotal information regarding the effectiveness of mathematics specialists. Despite the paucity of empirical research, the Task Group believes that there is a useful contribution that the Panel can make to debates about this idea. What the Task Group has produced is a synthesis of the programs and models that exist in the United States, some distinctions among different models, and some commentary on the costs and benefits of those different models.

II. Methodology

The Task Group identified and organized the available scientific evidence into several categories as they reviewed studies related to each of the four research questions addressed by the group. The quality of the evidence varied for each research question, as described in the individual sections of the report. For example, because experimental manipulation of teacher characteristics, such as teacher preparation or teacher content knowledge, is not easily done, the Task Group relied on correlational studies investigating the relationship of these variables to student achievement. Following is a discussion of the categories for studies with quantitative designs that were considered for inclusion because they provided the most rigorous evidence available, followed by a discussion of the procedures for identifying relevant research and synthesizing the results. In their report, the Task Group relied primarily on the highest quality studies available for each question.

A. Categories of Studies

1. High-Quality Studies

The high-quality studies identified for the research questions addressed by the Task Group included those with designs that were randomized controlled trials (RCTs), quasi-experimental studies that included baseline equivalence of treatment and control groups formed other than by random assignment, or correlational studies that met stringent quality standards. The strongest quality quasi-experimental and correlational studies were determined on the basis of: 1) the size of sample, 2) appropriate and adequate statistical controls, 3) use of multiple specifications, or tests for robustness of results, or both 4) use of individual-level versus aggregated data, and 5) the appropriateness and strength of the identification of the outcomes variable. These elements are reported for each study reviewed. The correlational studies in this category used such multivariate methods as regression analysis in the form of either standard ordinary least squares (OLS) or hierarchical linear modeling (HLM).

More specifically, the correlational studies in this category all had these three characteristics:

1) A baseline control for the outcomes measure (e.g., prior student mathematics achievement);

2) Use of multivariate regression analysis; and

3) A strong outcome measure (e.g., standardized test to measure student mathematics achievement).

and at least three of the following four characteristics:

1) Sample size—more than 1,000 observations on student mathematics achievement;

2) Statistical controls—contained a pretest control for student mathematics achievement plus controls for other relevant student and teacher characteristics;

3) Multiple specifications—multiple model specifications or other robustness checks on the results; and

4) Student-level data—the analysis was conducted with student-level data that had not been aggregated beyond the classroom of students instructed by a teacher.

Strong quasi-experimental and correlational studies formed the core of the research available to support conclusions and recommendations of the Task Group.

2. Moderate-Quality Studies

Studies in this category were those that were empirical, but did not meet a sufficient number of the standards established for the highest category. For instance, some of the studies in this category are those with strong research designs but weak outcome measures, such as indirect measures of teacher content knowledge or student achievement. All of the studies included in this category used correlational methods, such as regression analysis in the form of either standard OLS or HLM.

3. Lesser-Quality Studies

Studies in this category met very few of the criteria established for the highest quality studies and/or their measure of a variable of interest was extremely weak. For example, studies where the measure for teacher content knowledge was defined as teacher certification in a field other than mathematics or a score on a mathematics test that the teacher took many years earlier fall into this category.

B. Procedures

1. Literature Search and Study Inclusion

Literature searches were conducted to locate studies on the relationship between selected teacher characteristics and student learning in mathematics. Electronic searches were made in PsycINFO and the Social Sciences Citation Index (SSCI) using search terms identified by the Task Group. Studies were also identified through manual searches of relevant journals and reference lists, and recommendations from experts. Abstracts from these searches were screened for relevance to research questions and appropriate study design. For each study that met the screening criteria, the full study was examined to determine whether it met the inclusion criteria specified below. Citations from those articles and research reviews were also examined to identify additional relevant studies.

Criteria for Inclusion:

- Published between 1975 and 2007;
- Involved Grade K–12 students studying mathematics or involved teachers of mathematics for Grade K–12 students;
- Available in English;
- Published in a peer-reviewed journal, government or national report, book, or book chapter; and
- Used multivariate analysis with statistical controls.

2. Effect Size Calculations

Standardized regression coefficients typically were used as effect size measures because they are often the only data available. The equation used to calculate standardized regression coefficients is equation 2.3 from Bring (1994):

$$B_i = \hat{\beta}_i (s_i / s_y)$$

In this equation, $\hat{\beta}_i$ is the standardized regression coefficient on variable *i*, $\hat{\beta}_i$ is the regression coefficient reported in the paper for variable *i*, s_i is the standard deviation of variable *i*, and s_y is the standard deviation of the dependent variable. Thus, to calculate the standardized regression coefficient, the regression coefficient is multiplied by the standard deviation of the independent variable, and divided by the standard deviation of the dependent variable. The first element required for this calculation, $\hat{\beta}_i$, was available in any paper where a regression coefficient on the variable of interest was reported, which included all of the high-quality studies. However, the two required standard deviations were frequently not reported, and in those cases the standardized regression coefficient could not be calculated.

The standardized regression coefficient is interpreted as the change in the dependent variable as a fraction of the standard deviation that results when the independent variable changes by one standard deviation, holding the other variables constant. It is important to note, however, that although this is the best method available to calculate effect sizes given the data available in these papers, it is somewhat controversial and thus should be interpreted with caution. Essentially, the standard deviations used here are not entirely appropriate, because they are the overall standard deviations, which are much larger than the standard deviations when all independent variables but one are held constant. For more detail and an example of why standardized regression coefficients can be misleading, see Bring (1994).

When possible, the Task Group applied the What Works Clearinghouse (WWC) guidelines to calculate standardized mean differences in mathematics achievement.[1] Using C*omprehensive Meta-Analysis, Version 2* software, Hedges *g* standardized mean differences were calculated for these studies. The standardized mean difference is defined as the difference between the mean score for the treatment group minus the mean score for the comparison group, divided by the pooled standard deviation of that outcome for both the treatment and comparison groups.

In cases where schools, teachers, or classrooms were assigned (either randomly or nonrandomly) into intervention and comparison groups, and the unit of assignment was not the same as the unit of analysis, the effect size and accompanying standard error were adjusted for clustering within schools, teachers, or classrooms. This analysis used WWC guidelines to adjust for clustering, applying the default intra class correlation (ICC) adjustment for achievement guidelines of 0.20 when actual ICC values were unavailable.[2]

[1] See http://ies.ed.gov/ncee/wwc/twp.asp for the guidelines.
[2] See http://ies.ed.gov/ncee/wwc/overview/review.asp?ag=pi for more information on this issue.

III. Findings

A. Teachers' Knowledge of Mathematics

What evidence exists about the relationship between the depth and quality of teachers' knowledge of mathematics and gains in students' achievement?

1. Introduction

It is widely assumed—some would claim common sense—that teachers must know the mathematical content they teach. Yet verifying this assertion and making it more precise for the purposes of teacher education, professional development, and policy have proved challenging for researchers. How much mathematics course work do teachers need to take? How much do they need to know? What exactly do they need to know, and do they need to learn it? On the one hand, a simple answer—and an obvious one—is that teachers must have a strong grasp of the curriculum that they are responsible to teach. Knowing the curriculum well enough also requires knowledge at levels beyond their grade assignment. On the other hand, showing the effect of different kinds or amounts of teacher mathematical knowledge on student achievement has been far from easy or conclusive. Expert opinion can offer insight into the mathematical demands of teaching.

In this section, the Task Group considered empirical evidence that might help to shape teachers' education and policy: How does the level and scope of teachers' own mathematical attainment affect the learning of their students?

This is, of course, hardly a new question. The relationship between teacher characteristics and student achievement outcomes has been examined extensively over the past four decades, reaching as far back as a study by Coleman in the 1960s (Coleman et al., 1966). This education production function research attempted to determine the relationships between inputs, such as school and teacher characteristics, that can be modified (e.g., class size, expenditures on teacher salaries) and desired education outcomes (e.g., increased student achievement). The findings from this body of work have not produced consistent conclusions (e.g., Hanushek, 1986, 1989; Hedges, Laine, & Greenwald, 1994; Krueger, 1999; Monk, 1992), especially regarding matters of the relationship between education expenditures and student performance. Yet, as Goldhaber (2007) noted, a major finding from this literature was that of all school-related factors, teacher quality dominates effects on student achievement. As such, more recent interest in school inputs has turned to examining how the characteristics of teachers and the policies that affect them (e.g., teacher certification, teacher testing) may be related to student achievement. In this section the Task Group turns to examining how teachers' mathematical knowledge is related to student achievement.

Studying the relationship of teachers' mathematical knowledge to student achievement requires a measure of content knowledge. In the studies the Task Group reviewed and that met their standards of rigor, they identified three ways in which teachers' mathematical knowledge has been measured and related to student achievement gains: teacher certification, mathematics course work, and tests of teachers' mathematical knowledge. Other paths to assess teachers' content knowledge include scores on certification examinations, individual interviews and structured tasks, and teachers' skill with mathematics in the context of actual instruction. The Task Group did not locate studies that examined the effects of measures on teachers' performance on the learning of their students, and so these other approaches are not reflected in their initial summary. Consistent with the findings of other researchers who also have examined this body of literature, the Task Group finds that when studies are combined by type, including those that rely on proxies for teacher knowledge, such as certification status, the general results are mixed (Mandeville & Liu, 1997; Hill, Rowan, & Ball, 2005). When studies that use direct measures of teachers' mathematical knowledge are isolated, the effects are more positive.

The Task Group summarizes the evidence available about teacher knowledge effects, comments on the quality and reliability of that evidence, and suggests a set of warranted conclusions and inferences based on the strongest evidence. In addition, for the strongest group of studies, the Task Group provides the magnitude of the findings for illustrative purposes.

2. Teacher Certification as a Measure of Mathematical Content Knowledge

The Task Group begins this synthesis by examining literature that uses teacher certification as a proxy for teachers' mathematical content knowledge. A drawback to certification status is its inexactness as a measure of teachers' content knowledge. This stems from several problems. The first is one of self-selection. If a teacher's certification status is correlated with other characteristics that may also affect student outcomes (e.g., motivation), then any impact of certification on student achievement may not necessarily reflect the effect of greater mathematical content knowledge. The second problem is that, to the extent that certification status does not measure more nuanced elements of teacher mathematical content knowledge, if and where those "undetectable" teacher-to-teacher differences exist and are significant determinants of student outcomes, estimates of the effect of teacher mathematical content knowledge will be less precise, i.e., biased toward finding no impact on student outcomes. Another limitation is that different types of certification status (e.g., standard versus emergency) complicate the understanding of the effect of teacher mathematical content knowledge, primarily because how "true" mathematical content knowledge differs between these two groups of certified teachers is not completely clear (see Darling-Hammond, Berry, & Thoreson, 2001; Fetler, 1999). Finally, the small percentage of uncertified teachers may also raise questions about what is being measured.

a. Overall Findings

Overall, findings about the impact of teacher certification on student achievement in mathematics have been mixed, even among the most rigorous and highest-quality studies. Research in this area has not provided consistent or convincing evidence that students of teachers who are certified in mathematics gain more than those whose teachers are not. This may be in part due to some of the drawbacks mentioned above (Goldhaber & Brewer, 2000; Rowan, Correnti, & Miller, 2002; Fetler, 1999; Mandeville & Liu, 1997).

As can be seen in Table 1, three studies found positive effects of teacher certification on student achievement (Goldhaber & Brewer,[3] 1997b, 2000; King Rice, 2003*)*. Interestingly, each made use of the U.S. Department of Education's National Educational Longitudinal Survey of 1988 (NELS:88), using as the dependent measure gain scores in student achievement, or student achievement with prior grade achievement as a control. The earlier work by Goldhaber and Brewer (1997b) examined differences in student achievement in mathematics in 10th grade with scores in 8th grade as a control, while their 2000 study and King Rice's 2003 research examined 12th-grade differences in student achievement with scores in 10th grade as a control. Each of these studies used scores on the NELS:88 mathematics test as its dependent measure, and each used a number of statistical controls in their regression analyses, including prior mathematics test scores. And though their point estimates of the effect of certification differ, all found positive impacts of teacher certification in mathematics on student achievement.

Conversely, two of the high-quality studies and one of moderate quality found no significant effect of teacher certification as a predictor of student achievement in mathematics. These include studies by Kane, Rockoff, and Staiger (2006); Hill et al. (2005); and Rowan et al. (2002).[4] Each of these studies made use of distinctly different data sets that vary in terms of measures (for both student test scores and sample characteristics) and sample representation. Specifically, Kane et al. used statewide data and state-specific standardized test scores for mathematics in New York City, while the other two studies used data spanning states that were originally collected as part of another study (Hill et al.; Rowan et al.).[5] Across all of these data sources, none of the studies found a significant effect of teacher certification status on students' mathematics achievement. Thus, the strongest evidence currently available about the effect of teacher certification on student achievement in mathematics (i.e., the high- and moderate-quality studies reviewed for this section) is mixed.

[3] A third study (Goldhaber & Brewer, 1997a) uses the same data set and sample of students and teachers as the 1997b study. As the latter study includes additional predictors (i.e., teacher classroom behavior) the Task Group chose to include it as it appears to be a better specified model.

[4] Although all three of these studies represent solid empirical investigation, the Task Group considers the models in the Rowan, Correnti, and Miller (2002) study to be of *moderate* quality due to data limitations that the authors faced in their analysis. Specifically, the authors acknowledged that their results and subsequent claims about the effect of certification reflected estimates based on less than 6% of their sample. The models reported in the Harris and Sass (2007a) study investigated the effect of National Board for Professional Teaching Standards (NBPTS) certification, which is very different than standard certification. Thus, while the Task Group still considers these reports in drawing their conclusion, they represent less powerful evidence to support the argument at hand.

[5] Hill, Rowan and Ball (2005) make use of a study of instructional improvement initiatives in schools. Rowan, Correnti, and Miller (2002) uses the *Prospects: The Congressionally Managed Study of Educational Growth and Opportunity 1991–1994.*

Table 1: Quality Characteristics of Models Examining Impact of Teacher Certification on Student Achievement in Mathematics, by Study and by Overall Quality

			Other Controls							
Authors	**Sample Size**	**Pretest Control[a]**	**Student**	**Teacher**	**Class, School, or District**	**Family**	**Multiple Specifications**	**Student-Level Analysis**	**Identification Measure**	**Grade(s)**
High-Quality Studies										
Goldhaber, & Brewer, 1997b	5,149 students, 2,245 teachers 638 schools 3498 math classes	X	X	X	X	X	X	X	Certified in math	10
Goldhaber, & Brewer, 2000	3,786 students, 2,098 teachers	X	X	X	X	X	X	X	Level of certification[b]	12
King Rice, 2003	3,696 teachers	X		X	X		X		Certified in math	12
Kane, Rockoff, & Staiger 2006	1,462,100 student-year observations	X	X	X	X	X	X	X	Type of certification	Panel data: 3–8
Hill, Rowan, Ball, 2005	2,963 students, 699 teachers	X	X	X	X	X	X	X	Certified	1 and 3
Moderate-Quality Studies										
Rowan, Correnti, & Miller, 2002	Panels of about 4,000 students in more than 300 classrooms and more than 120 schools.[e]	X	X	X	X	X		X	Certified in math[c]	Cohort 1: 1–3; Cohort 2: 3–6
Harris, & Sass, 2007a	1,112,984 student-year observations	X	X	X	X	X	X	X	NBPTS certification[d]	4–10
Lesser-Quality Studies										
Darling-Hammond, 1999	44 NAEP state averages for students, 52,000 public school teachers, 9,500 public schools, 5,6000 school districts								Well-qualified: state certification + equivalent of math major (B.A. or M.A.)	4 and 8
Fetler, 1999	921,437 Total Grade 9: 347,201 Grade 10: 313,303 Grade 11: 260,933			X	X	X			Emergency certification, school-level	9–11
Hawkins, Stancavage, & Dossey, 1998	Unweighted - Grade 4: 6,627; Grade 8: 7,146; Grade 12: 6,904; Weighted - Grade 4: 3,714,998; Grade 8: 3,570,116; Grade 12: 2,830,443								Certified in math, certified in education	4, 8, and 12
Larson, 2000	6,474 students, 185 teachers	X							Certified in math	1st semester algebra students
Mandeville, & Liu, 1997	203 teachers				X				Grade level certification	7

[a] Includes use of pretest as control variable, gain scores, or value-added models.
[b] Teachers responded to a question about their certification in mathematics or science. Based on their responses, they were classified as one of five groups: standard certification (reference group), probationary certification, emergency certification, private school certification, and no certification.
[c] The authors caution that only 6% of the sample had special certification, subject-matter degrees, or both.
[d] National Board for Professional Teaching Standards (NBPTS) certification can be general or subject-specific—this measure includes both types.
[e] The authors conducted 12 analyses, given the design of the parent study *(Prospects: The Congressionally Managed Study of Educational Growth and Opportunity 1991–1994)* the number of students, classrooms, and schools varies over time as the pathways students take diverge.

b. Strength of the Findings

The effect of teacher certification status on students' mathematics achievement remains somewhat ambiguous. Of the 12 studies the Task Group reviewed on this subject, 5 provide the highest-quality evidence due to five features of their design: 1) sample size, 2) appropriate and adequate statistical controls, 3) multiple specifications or tests for robustness of results,[6] 4) micro-level versus aggregated data, and 5) the appropriateness and strength of the identification of teacher content knowledge.[7] These elements are reported in Table 1 for the 12 studies that use teacher certification status as a proxy for mathematical content knowledge.

Seven studies (those in the first two panels of Table 1, the high- and moderate-quality studies) used regression analysis in the form of either standard OLS or HLM. Although the other five studies (in the bottom panel of the table) support the general conclusions of the Task Group's findings, they represent the weakest evidence because they lack important qualities such as adequate controls (e.g., pretest scores for students), highly detailed data sets, or meaningful alternative specifications. The implications of these studies, therefore, must be interpreted with caution.

c. Magnitude of the Findings

Although most of these studies use very different measures of student achievement to assess the effect of teacher certification, it may be beneficial to know something about the magnitude of the effects reported in each piece. Already known are the different measures of dependent variables that can complicate the ability to compare regression coefficients across studies. Additionally, many studies lack sufficient information to transform measures of impacts (e.g., regression coefficients, mean differences) into standardized measures across studies. As a result, this Task Group can do little more than report standardized regression coefficients where data are available, and be cautious in their interpretation and comparison of these effects, regardless of their statistical significance. Table 2 lists the reported impacts of teacher certification on student achievement for the high-quality studies in terms of standardized regression coefficients.

---

[6] Robust results are those that are consistent even when multiple specifications or models are used.
[7] Of the five studies that met the criteria for high quality, only one (King Rice, 2003) does not employ student- and family-level controls in its analysis. This is because King Rice aggregates her student- and family-level controls to the classroom (average) level. Also, the use of a pretest control in conjunction with other classroom- teacher- and school-level controls suggests her analysis satisfies the criteria of appropriate and adequate statistical controls.

Table 2: Reported Impacts of Models Examining the Effect of Teacher Certification on Student Achievement in Mathematics for the High-Quality Studies

Authors	Dependent Measure	Standardized Regression Coefficient[b]	Analytic Technique
High-Quality Studies[a]			
Goldhaber, & Brewer, 1997b	NELS test battery scores (Grade 10)	0.06*	Generalized Least Squares
Goldhaber, & Brewer, 2000	NELS test battery scores (Grade 12)		Ordinary Least Squares
	Probationary certification in subject[c]	0.01	
	Emergency certification in subject[c]	0.00	Ordinary Least Squares
	Private school certification[c]	-0.01	Ordinary Least Squares
	No certification in subject[c]	-0.01*	Ordinary Least Squares
King Rice, 2003	NELS test battery scores (Grade 12)	0.02*	Ordinary Least Squares
Kane, Rockoff, & Staiger, 2006	NYC standardized test scores (value-added Grades 3–8)	0.00	Ordinary Least Squares
Hill, Rowan, & Ball, 2005	Terra Nova scores		Hierarchical Linear Modeling
	- Grade 1	0.00	
	- Grade 3	0.00	

*p < .05
[a] High-quality studies are defined on page 3.
[b] The standardized coefficient was calculated for those studies with sufficient data using the formula provided in Bring (1994) $B_i = \hat{\beta}_i(s_i / s_y)$
[c] The comparison is to standard certification in subject.

As shown in Table 2, where there is a significant effect of teacher certification in mathematics on student achievement, it is quite small. Also, because teacher certification in mathematics is a remote proxy for mathematical content knowledge, this measure does not allow strong inferences about the effect of teachers' knowledge on their students' achievement.

3. Content Course Work and Degrees as Measures of Mathematical Content Knowledge

Mathematics course work and field-specific degrees are a second common proxy for teachers' mathematical content knowledge. Although these are different, they both focus on teachers' completion of college-level mathematics study. Consequently, these measures frequently appear together within the same data set and, thus, are often jointly considered within the same analysis. Therefore, for this section the Task Group discusses and synthesizes the literature and subsequent evidence presented for the relationship between these measures of teacher mathematical content knowledge and student achievement.

Although the amount of course work or the possession of a degree in mathematics are both closer predictors of a teacher's mathematical knowledge than certification status, these are still both proxies for that knowledge and each has unique validity problems. Neither measures the actual command of specific mathematical topics and skills. Neither measures what an individual actually learned, which may vary substantially from person to person. There is similarly no information about the correspondence between particular courses and the school curriculum for which teachers are responsible. Thus, as a measure of the knowledge on which teaching depends, course work or degree attainment may or may not correspond to what teachers use in the course of their work.

Also, as with certification status, there is no guarantee that using the amount of course work completed or type or level of degree circumvents the problems of selection bias previously mentioned. It could be that teachers who engage in generous amounts of mathematics course work or obtain mathematics degrees are particularly motivated to teach mathematics, or possess some other unobservable characteristics unrelated to course work and degree that make them especially effective at teaching mathematics. Finally, these measures do not take into account the passage of time. On the one hand, this may be important if individuals forget lessons from their schooling and change the core of their mathematics instruction over time to satisfy student or school needs or demands (e.g., switching from teaching geometry to teaching calculus).[8] On the other hand, if teachers constantly practice the content-specific skills they need, and those needs do not significantly change over time, then the time lag is decidedly less important.

a. Overall Findings

Much like the research using certification status as a proxy for teachers' mathematical knowledge, the findings in the literature on the impact of content-specific course work and degrees are mixed.[9] Among the seven studies of high quality that examine the impact of a teacher's mathematics course work, or degree, or both, on students' mathematics achievement, several of which look at multiple measures of course work, or degree, or both, four provide estimates of the relationship between degree and student achievement (three are positive; one negative), four provide estimates of the relationship between course work and student achievement (all positive), and two examine the relationship between college credits and student achievement (one positive, one negative). The evidence from this set of studies is somewhat more consistent than the evidence for an effect of teacher certification, thus pointing to a likely positive relationship between teacher content course work and degrees, and student mathematics achievement. Left unexamined are important details, such as how many courses, or which degrees or programs of study make the most difference. Moreover, of the studies that found a positive impact of teacher course work and degrees on student mathematics achievement, many use NELS:88 data and most focus on high school students; thus their scope is limited.

In particular, two reports by Goldhaber and Brewer (1997b, 2000) found positive impacts of course work attainment as noted previously. Both of these studies use the NELS:88 to measure student mathematics achievement, and their samples include students in 10th (1997b) or 12th grades (2000). Findings from both studies indicate that students who have teachers with degrees in mathematics perform significantly higher on the NELS test battery than do students of teachers who are not qualified in math. Rowan, Chiang, and Miller (1997) also use the NELS data for 10th-grade students, finding that teacher's possession of a mathematics degree is associated with higher student achievement. Monk and King (1994), on the other hand, use Michigan State University's Longitudinal Study of American Youth (LSAY), which uses items from the U.S. Department of Education's National Assessment of Educational Progress (NAEP) achievement test in mathematics. That

[8] Though including teacher experience as a control variable (common in the literature) helps to alleviate the inexactness of course work and degrees as measures over time, estimates of the interaction between teacher experience and course work and degrees (not common in the literature) would help to determine actual effects of course work and degrees.
[9] Supported by Wayne and Youngs (2003).

study also focuses on secondary school students but examines teacher course work at the graduate and undergraduate level. They found a positive relationship between number of mathematics courses taken and student achievement. The positive effects of teacher preparation in mathematics thus seem to exist across different data sources, although as noted, these studies all focus on high school grades.

Less clear is how teachers' level of conventional college mathematics study affects student achievement below ninth grade. Rowan et al. (2002) used a survey that followed two cohorts of students, starting in Grades 1 and 3, for 4 years. The authors found a negative impact for an advanced mathematics degree on student achievement, but acknowledge that fewer than 6% of teachers (approximately 43 total) had subject-matter specific degrees. Harris and Sass (2007b) examined a number of different measures for students of all different ages, and although they mostly found no effects, they identified positive impacts for some measures and negative for others. Hill et al. (2005) found that teachers' mathematics course work did not significantly predict gains in student achievement.

Thus, the results across the studies that use teachers' attainment as a proxy for mathematical knowledge are mixed. The strongest suggestion is that the level of teachers' mathematics study may predict student achievement at the high school level. Evidence was not uncovered to support this relationship below ninth grade. It may be that using course work as a proxy for teachers' actual knowledge is a less valid measure at this level than it is at the secondary school level where the content teachers teach is closer to the content they study in college.

b. Strength of the Findings

As described earlier in the section on certification, the Task Group chose to focus on a limited number of studies of the highest quality, based on 1) sample size, 2) appropriate and adequate statistical controls, 3) multiple specifications or tests for robustness of results, 4) micro-level versus aggregated data, and 5) the appropriateness and strength of the identification. These elements are reported in Table 3. Each of the studies used detailed regression analysis, in the form of either standard OLS or HLM.

Studies of lesser quality are also reported in Table 3. While these studies support the general conclusions of the Task Group's findings, they represent weaker evidence because they lack such important qualities as adequate controls (e.g., pretest scores for students), highly detailed data sets, and meaningful alternative specifications. Therefore, the implications of this research must be interpreted with caution.

c. Magnitude of the Findings

Although these studies use differing measures of student achievement and of teacher mathematics course work and degrees, there is value in knowing something about the magnitude of the effects reported in each piece. Table 4 lists the reported impacts of teacher course work and degrees on student achievement for the highest quality studies using standardized regression coefficients to allow comparisons across the studies where data are available. Table 4 shows that for teacher certification, the impact of teacher course work and degrees on student mathematics achievement, when a standardized coefficient can be calculated, is quite small. This finding is consistent across measures and grade levels, although few studies have sufficient data for calculating the magnitude of the effect. Therefore, generalizations are not appropriate here.

Table 3: Quality Characteristics of Models Looking at Impact of Teacher Mathematics Course Work and Degrees on Student Achievement in Mathematics, by Study and by Overall Study Quality

			Other Controls							
Authors	**Sample Size**	**Pretest Control**[a]	**Student**	**Teacher**	**Class, School, or District**	**Family**	**Multiple Specifications**	**Student-Level Analysis**	**Identification Measure**	**Grade(s)**
High-Quality Studies										
Goldhaber, & Brewer, 1997b	5,149 students, 2,245 teachers, 3,498 math classes, 638 schools	X	X	X	X	X	X	X	B.A. and graduate degree in math	10
Goldhaber, & Brewer, 2000	3,786 students, 2,098 math teachers	X	X	X	X	X	X	X	B.A. and graduate degree in math	12
Harris, & Sass, 2007b	Grades 4–5: 514,620 students/ 785,780 observations; Grades 6–8: 542,289 students/ 784,423 observations; Grades 9–10: 426,474 students/ 667,698 observations Teachers: Grades 3–5: 2,134; Grades 6–8: 943; Grades 9–10: 639	X	X	X	X	X	X	X	Course work in math, college credits, math degree	4–10
Hill, Rowan, & Ball, 2005	2,963 students, 699 teachers	X	X	X	X	X	X	X	Course work in math content and math methods	1 & 3
Monk, 1994	2,829 students, 608 math teachers	X		X	X		X	X	College degree in math, college courses in math	10 & 11
Monk, & King, 1994	2,831 students, their teachers	X	X	X	X	X	X	X	College courses in math	Cohort 10–12
Rowan, Chiang, & Miller, 1997	5,381 students	X	X	X	X	X	X	X	College degree in math	10
Moderate-Quality Studies										
Rowan, Correnti, & Miller, 2002	Panels of about 4,000 students in more than 300 classrooms and more than 120 schools	X	X	X	X	X	X	X	College degree in math[b]	Cohort 1: Grades 1–3; Cohort 2: Grades 3–6
Lesser-Quality Studies										
Darling-Hammond, 1999	44 state averages for students, 52,000 public schools, 65,000 teachers								Well-qual: certification + math major	4 & 8
Darling-Hammond, Berry, & Thoreson, 2001	3,786 students, 2,078 teachers	X		X				X	College degree in math (BA or MA)	12
Eisenberg, 1977	807 students, 28 teachers		X					X	College math GPA, courses taken	Junior high Algebra I students

[a] Includes use of pretest as control variable, gain scores, or value-added models. Gain scores are the calculated gains of students across a school year.
[b] The authors also caution that they identified the effect of mathematics college degree from a maximum of 6% of the teachers in their sample.

Table 4: Reported Impacts of Models Examining the Effect of Teacher Math Course Work and Degrees on Student Achievement in Mathematics, by Study and By Overall Quality

		Standardized[a] Regression Coefficients			
Authors	**Dependent Measure**[b]	**College Credits**	**Degree/Major**	**Course Work**	**Analytic Technique**
Goldhaber, & Brewer, 1997b	NELS test battery scores (Grade 10)		0.03*		Ordinary Least Squares
Goldhaber, & Brewer, 2000	NELS test battery scores (Grade 12)		Not available		Ordinary Least Squares
Harris, & Sass, 2007b	FCAT scores				2-stage Ordinary Least Squares
	- Elementary		Not available		
	- Middle		Not available		
	- High		Not available		
Hill, Rowan, & Ball, 2005	Terra Nova scores				Hierarchical Linear Modeling
	- Grade 1	0.02			
	- Grade 3	0.05			
Monk, 1994	Scores on NAEP-based tests				Ordinary Least Squares
	- Grade 10		Not available	Not available	
	- Grade 11		Not available	Not available	
Monk, & King, 1994	Gain scores on NAEP-based tests (Grades 10, 11, and 12)			.09*	Ordinary Least Squares
Rowan, Chiang, & Miller 1997	NELS test battery scores (Grade 10)		Not available		Hierarchical Linear Modeling

*p < .05

[a] The standardized coefficient was calculated for those studies with sufficient data using the formula provided in Bring (1994) $B_i = \hat{\beta}_i(s_i/s_y)$.

[b] Of the studies that consider measures of mathematics course work as their independent variable, only one (Monk, 1994) partitions its measure by the type of course work, in this case graduate and undergraduate course work. Along these measures, the authors find that there is no significant effect (at the 5 % level) of graduate mathematics course work on student achievement, and that there is a significant effect of undergraduate mathematics course work, but only for one (i.e., the juniors) of the two grades tested.

4. Test Scores and Ad Hoc Measures as Measures of Mathematical Content Knowledge

Using test scores and more proximal measures of teacher mathematical content knowledge allows closer examination of the effect that mathematical knowledge has on student achievement. Such measures escape some of the traditional problems of selection bias and inexactness present in other more conservative measures. These types of measures make it possible to probe more directly the causal link between mathematical knowledge and student achievement. One caution with this approach is that specially developed tests or measures, although intuitively appealing or apparently more relevant, have often not been validated or otherwise checked for their psychometric quality.

a. Overall Findings

Research that has used teacher test scores and other ad hoc measures has also produced mixed results. The Task Group's inability to draw solid conclusions from this literature is in part due to the general lack of quality measures of mathematics content knowledge, as well as the absence of an adequate number of high-quality studies using these types of measures. Overall, the Task Group identified five studies that both examined the effect of mathematical content knowledge (as measured by test scores and other instruments) on student achievement and that met their standards for high or moderate quality. Of these, two studies found a positive and significant effect of mathematical content knowledge on student achievement, another found a positive but statistically insignificant effect, and two other studies found ambiguous effects for various measures with their sample. Unlike the studies reviewed earlier on teacher certification and course work that mostly examined student achievement at the high school level, the studies in this group are focused at the elementary level, making comparisons with other findings difficult.

Clotfelter, Ladd, and Vigdor (2007) examined the relationship of teacher test scores to student mathematics achievement in North Carolina. Although the test did not focus solely on mathematics content, they found that higher teacher test scores are a significant predictor of higher student achievement. Hill et al. (2005) used test items specifically designed to measure the mathematical content knowledge used in teaching and controlled for content knowledge in teaching reading. They found that the measure of content knowledge of mathematics is a significant and positive predictor of student success in math.

Harbison and Hanushek (1992) examined the effect of teacher mathematics test scores on fourth-grade tests on student achievement in Brazil; the authors found a positive effect of teacher test score on student achievement: "At fourth grade, a ten-point improvement in the mean teacher's command of her mathematics subject matter … would engender a five-point increase in student achievement; this is equivalent to a 10% improvement over the mean scores of fourth graders" (p. 114). The effects are not significant at the traditional .05 statistical level.

Two other studies used standardized tests as measures of teachers' knowledge of mathematics. These show ambiguous results for determining the impact of teacher content knowledge on achievement. One study (Harris & Sass, 2007b) used teachers' quantitative Scholastic Aptitude Test (SAT) scores taken at the time of college entry, a measure that is substantially more distal than those used in the studies reported above, to assess the impact of teacher content knowledge on math achievement for students taking the Florida Comprehensive Assessment Test (FCAT). Further, they partitioned their sample by grade to differentiate potential impacts at various grade levels. Ultimately they found that there is no significant effect of teachers' previous higher SAT scores on elementary and middle school students, and that there is actually a negative impact on achievement for high school students.

Mullens, Murnane, and Willett (1996) used teacher test scores on the Belize National Selection Exam (BNSE) to analyze their effect on students' understandings of basic and advanced concepts in math. The authors found that while teacher test scores are not significant predictors of achievement for basic concepts in mathematics, they do exert a positive influence on student understanding of advanced concepts.

b. Strength of the Findings

The Task Group's conclusions about the impact of teacher content knowledge on student achievement, as measured by tests or specially designed measures, could only be drawn from a small number (*n* = 3) of the high-quality studies. Still, evidence from the lesser-quality studies is consistent, pointing in a positive direction. For example, Rowan et al. (1997) identified teacher content knowledge using a teacher's response to a single school-relevant mathematics question and found positive effects for that one item. Additionally, two studies (Harris & Sass, 2007b; Mullens et al., 1996) also are considered moderate quality due to their choice of measurement of teacher mathematical content knowledge. It could be argued that the use of quantitative SAT scores (Harris & Sass) may not capture important elements of mathematical content knowledge that are acquired at the collegiate level. By the same token, the use of eighth-grade BNSE scores (Mullens et al.) is even more likely to be a less relevant measure. Nevertheless, this Task Group cannot discount the possibility that these measures may assess mathematical content knowledge relevant for elementary (or even middle) school teaching. Overall, the evidence here (Table 5), based on a small number of studies, does point more strongly in the direction of a relationship between teachers' usable knowledge of mathematics and students' achievement than do the other measures of teacher content knowledge.

c. Magnitude of the Findings

Similar to the other sections, the Task Group reports the magnitude of the findings for each study in this strand using standardized regression coefficients, where data are available (see Table 6 later in this section). Since each one of these studies uses a distinctly different dependent measure, and only two studies have sufficient data for making comparisons, the Task Group cannot make any meaningful relative interpretations of these findings. However, it should be noted that even though the magnitude of the findings where they are available is quite small, the Hill et al. (2005) findings are substantially larger than the relationships noted for other measures of teacher content knowledge.

Hill et al. (2005) also examined the effect of teacher certification in mathematics and mathematics education course work, in addition to their specific content knowledge measure These comparative analyses show that the specific measure of mathematical knowledge for teaching used by Hill et al. (2005) is measuring a type of understanding and skill that is not captured by certification status or course work measures.

Table 5: Quality Characteristics of Models Looking at Impact of Teacher Test Scores And Other Ad Hoc Measures on Student Achievement in Mathematics, by Study and By Overall Study Quality

			Other Controls							
Authors	**Sample Size**	**Pretest Control***	**Student**	**Teacher**	**Class, School, or District**	**Family**	**Multiple Specifications**	**Student-Level Analysis**	**Identification Measure**	**Grade(s)**
High-Quality Studies										
Clotfelter, Ladd, & Vigdor, 2007	Nearly 1 million student-year observations	X	X	X	X	X	X	X	North Carolina teacher test scores	3–5
Harbison, & Hanushek, 1992	Over 2,500 students	X	X	X	X	X			Brazilian math test scores	2–4
Hill, Rowan, & Ball, 2005	2,963 students, 699 teachers	X	X	X	X	X	X	X	CKT-M test measure	1–3
Moderate-Quality Studies										
Harris, & Sass, 2007b	Grades 4–5: 514,620 students/785,780 observations; Grades 6–8: 542,289 students/784,423 observations; Grades 9–10: 426,474 students/ 667,698 observations Teachers: Grades 3–5: 2,134; Grades 6–8: 943; Grades 9–10: 639	X	X	X	X	X	X	X	SAT Quant. Score	4–10
Mullens, Murnane, & Willett, 1996	1,043 students	X	X	X	X	X		X	BNSE Math Score	3
Lesser-Quality Studies										
Rowan, Chiang, Miller, & 1997	5,381 students	X	X	X	X	X	X	X	A single-item teaching-relevant and difficult math test	10
Sheehan, & Marcus, 1978	1,836 students, 119 teachers	X		X	X				WCET scores	1

* Includes use of pretest as control variable, gain scores, or value-added models. Gain scores are the calculated gains of students across a school year.

Table 6: Reported Impacts of Models Examining the Effect of Teacher Test Scores and Other Ad Hoc Measures on Student Achievement in Mathematics, by Study

Authors	Dependent Measure	Standardized Regression Coefficient[a]	Analytic Technique
High-Quality Studies			
Clotfelter, Ladd, & Vigdor, 2007	North Carolina standardized test (Grades 3–5)		Ordinary Least Squares
	- Score Level	Not available	
	- Score Gain	Not available	
Harbison, & Hanushek, 1992	Brazilian math test (from EDURURAL[b])		Ordinary Least Squares
	- Grade 2	Not available	
	- Grade 4	.02	
Hill, Rowan, & Ball, 2005	Terra Nova scores		Hierarchical Linear Modeling
	- Grade 1	.06*	
	- Grade 3	.05*	

*p < .05
[a] The standardized coefficient was calculated for those studies with sufficient data using the formula provided in Bring (1994) $B_i = \hat{\beta}_i(s_i/s_y)$.
[b] EDURURAL is the research project that emerged from an effort to improve educational performance in rural northeast Brazil.

5. The Mathematical Content and Nature of Teacher Licensure Exams

Recent research treating teacher licensure as a proxy for teachers' mathematical content knowledge has not consistently or convincingly shown that students of teachers who are licensed in mathematics gain more academically than those whose teachers are not (Goldhaber & Brewer, 1997a, 1997b, 2000; Hill et al., 2005; Kane et al., 2006; King Rice, 2003; Rowan et al., 2002). However, teacher licensure exams still play an important role in determining the quality and quantity of math teachers available for employment in schools. To assess the quality of teacher licensure exams, it is first necessary to ascertain the mathematical integrity of the exam questions as well as the relevance of these questions to teaching in the classroom of an elementary or middle school. Such precise information turns out to be difficult to obtain.

Most states use a teacher licensure exam that yields a score as a measure of a candidate's achievement in the subject of mathematics. Each state has designed its own unique system of licensure by choosing among different exams and determining cut scores for them.[10] The *Praxis Series* of exams created by Educational Testing Service (ETS) is the most commonly used teacher licensure exam. Overall, one or more of these exams are currently required in 38 states[11] and the District of Columbia, with 7 of these states[12] using

[10] Connecticut, Kansas, and Missouri use an identical set of certification exams and cut scores for the subject of math, as do Oklahoma and Nebraska.
[11] These are Alabama, Alaska, Arkansas, California, Connecticut, Delaware, Georgia, Hawaii, Idaho, Indiana, Kansas, Kentucky, Louisiana, Maine, Maryland, Minnesota, Mississippi, Missouri, Nebraska, Nevada, New Hampshire, New Jersey, North Carolina, North Dakota, Ohio, Oklahoma, Oregon, Pennsylvania, Rhode Island, South Carolina, South Dakota, Tennessee, Utah, Vermont, Virginia, Washington, West Virginia, and Wisconsin.

the *Praxis Series* in conjunction with their own exam. Of the 12 states that do not use *Praxis*, 9[13] use their own licensing exam exclusively, while the other 3 either do not require an exam for licensure or do not use an exam that yields an independent score in the subject of math.[14]

The *Praxis Series* is composed of two separate exams, The *Praxis I* and *II*. The *Praxis I* exams, or Pre-Professional Skills Tests (PPST), are designed to measure basic skills in reading, writing, and mathematics. Most ETS states currently require the *Praxis I* tests for licensure, and often for admission into their teacher education programs. The cut scores currently required for licensure across these states range from 169 to 178 while the average performance range of teacher candidates is from 175 to 183.[15] The *Praxis II* exams in mathematics measure specifically mathematical knowledge and teaching and are not nearly sufficient to warrant the formulation of any conclusions about the licensure of middle school teachers in mathematics.

ETS shared data regarding two *Praxis II* exams in Series 0061 and 0063, one testing mathematical content knowledge, and the other addressing mathematical proofs, models, and problems (both with item performance data). At least one of these two series is used by 33 states as part of the general licensure requirements in mathematics,[16] but 30 of these states[17] also require additional exams, and 24 of these[18] also use a *Praxis II* exam designed specifically for testing mathematical content knowledge at a middle school level.[19] The cut scores in each state for these particular exams supplied by ETS are not known because these exams have been retired by ETS. There is also another unresolved issue. As these exams are for single-subject certification, it would be necessary to know the precise role the exams played in certifying middle school teachers in the 30 states mentioned earlier before any statement can be made about the appropriateness of these exams for the certification of middle school teachers. For example, in the multiple choice exam on content knowledge, 8 of 25 exam items are on high school mathematics (e.g., calculus, trigonometry, conditional

[12] These are Alabama, Alaska, California, Georgia, Indiana, South Carolina, and Washington.
[13] These are Arizona, Colorado, Florida, Illinois, Massachusetts, Michigan, New Mexico, New York, and Texas.
[14] Wyoming requires teacher candidates to demonstrate knowledge of the U.S. Constitution and the Wyoming Constitution either through course work or an exam. Montana does not require any testing for licensure, and Iowa does not use an exam with a separate math score.
[15] The average performance range encompasses the scores earned by the middle 50% of the examinees taking the test (n = 93,805). This statistic provides an indication of the difficulty of the test. The standard error of measurement for this figure (2.4) is a statistic that is often used to describe the reliability of the scores of a group of examinees.
[16] These are Alabama, Alaska, Arkansas, Connecticut, the District of Columbia, Georgia, Hawaii, Idaho, Indiana, Kansas, Kentucky, Louisiana, Maine, Maryland, Minnesota, Mississippi, Missouri, Nevada, New Hampshire, New Jersey, North Dakota, Ohio, Oregon, Pennsylvania, South Carolina, South Dakota, Tennessee, Utah, Vermont, Virginia, Washington, West Virginia, and Wisconsin.
[17] These are Alabama, Alaska, Arkansas, Connecticut, the District of Columbia, Georgia, Hawaii, Idaho, Indiana, Kansas, Kentucky, Louisiana, Maine, Maryland, Minnesota, Missouri, Nevada, New Hampshire, New Jersey, North Dakota, Ohio, Oregon, Pennsylvania, South Carolina, South Dakota, Tennessee, Vermont, Virginia, Washington, and West Virginia.
[18] These are Alabama, Alaska, Connecticut, the District of Columbia, Idaho, Kansas, Kentucky, Louisiana, Maine, Maryland, Minnesota, Missouri, New Hampshire, New Jersey, North Dakota, Ohio, Oregon, Pennsylvania, South Dakota, Tennessee, Vermont, Virginia, Washington, and West Virginia
[19] Here the District of Columbia has been counted as a state and currently uses the *Praxis II* exam designed specifically for testing mathematical content knowledge at a middle school level in addition to an exam testing mathematical content knowledge and one addressing mathematical proofs, models, and problems.

probability, matrices) while only 5 are related to fractions, the central topic of middle school mathematics; of these 5, only 2 (one on percent, the other on rate) directly probe teachers' understanding of fractions. This exam, then, does not seem to assess directly the content for which middle school teachers are responsible.

To determine whether the tests adequately measure teacher content knowledge, precise information about their licensure is needed. At the moment, the collection of such information is greatly hampered by confidentiality issues. The Task Group recommends that there be more openness in this area to facilitate the kind of academic inquiries that are a prerequisite to progress.

a. Implications From the Empirical Evidence

Overall, across the studies reviewed, the signal is that teachers' knowledge of mathematics is a positive factor in students' achievement. However, despite the common-sense nature of this claim, solid evidence about the relationship of teachers' mathematical content knowledge to students' mathematics achievement remains uneven and has been surprisingly difficult to produce. One main reason has been the lack of valid and reliable measures of teachers' mathematical knowledge. The literature has been dominated by the use of proxies for such knowledge, such as certification status and course work. A second reason for the inconsistent findings has been weak study designs. Too few studies set up proper comparisons or use sufficient sample sizes or appropriate analytic methods. Selection bias and failure to isolate variables have further plagued these studies, as have inadequate or imprecise measures of students' mathematics achievement. Finally, no studies identified by the Task Group probed the dynamic that would examine how teachers' mathematical knowledge affects instructional quality, students' opportunities to learn, and their gains over time. However, in the context of a body of literature that is as inexact as it has been, the positive trends identified do support the importance of teachers' knowledge of mathematics as a factor in students' achievement.

To improve the quality of research evidence, sharper measures of teachers' mathematical knowledge in different domains and at different levels are needed, with appropriate psychometric tests for their reliability and validity. Hill et al. (in press) report a series of in-depth validation studies that provide a benchmark for what such tests should involve. Also needed are value-added studies, including experimental studies of interventions designed to develop teachers' mathematical knowledge for teaching and to inspect the effects of interventions on students' mathematics achievement.

As the scientific rigor and validity of research in this area increases, it should be informed by analyses of empirical data on the mathematical demands on teachers in the course of their work. Hypotheses generated through a mathematical perspective on teaching practice can provide researchers with a sharper focus on the question in ways more likely to produce results useful to policymakers, and to improve the relevance and effectiveness of teachers' mathematical training and professional assessment.

6. Recommendations Based on the Mathematical and Logical Analysis of The Demands of Teaching Mathematics

Although the empirical evidence does not yet strongly specify the nature of the mathematical knowledge needed for teaching, the Task Group hypothesizes that teaching mathematics demands knowledge of the subject. Because a direct relationship between conventional mathematical study and teacher effectiveness is not supported by a review of high-quality research, future research should uncover those aspects of teacher knowledge and understanding that are most strongly related to student learning. Policies should be developed and supported that would lead to a more mathematically skilled teacher force. The Task Group recommends investigating different components of content knowledge that may have relationships to instructional effectiveness.

A first component worth investigating is the nature of the requisite competence with the school curriculum that teachers are responsible to teach. This includes the concepts, skills, and strategies that students are to learn, and the prerequisite content, several levels both below and beyond the level at which they teach. Another aspect of this curricular knowledge is the additional need for teachers to know school mathematics at a more advanced level than what is found in school textbooks. This may be seen in the daily tasks that teachers must perform and that appear to entail substantial mathematical judgment, understanding, and skill. For example, answering students' questions may be unexpectedly subtle or complicated, or making up exam problems to focus on central ideas may require more than a minimal knowledge of the mathematics in the grade-level curriculum. It is also worth examining the extent to which teachers may need different types of advanced knowledge—for example, how an understanding of number theory may enrich and strengthen teachers' capacity to teach whole number concepts and operations.

Another important area in need of investigation is what else teachers may need beyond the standard skills and concepts in the school curriculum. Examples of work that teachers do that may require other mathematical knowledge include explaining why a particular topic is worth learning; providing connections within lessons, across lessons, and across grades; and making spontaneous instructional decisions in the classroom. These tasks of teaching seem to require mathematical skill that has not been well established in the research literature.

To help students extend what they know, teachers may also need a deep understanding of the foundational ideas and skills prerequisite to the level of instruction, the mathematics that leads up to the students' mathematical present. Middle school teachers teaching algebra may benefit from a nuanced understanding of operations, including their properties, and their interpretation and representations; upper elementary teachers may be enabled by an understanding of ways of renaming or re-representing mathematical ideas that arise in the early grades (e.g., with the standard subtraction algorithm).

In addition to knowledge of particular content, including concepts and procedures, teaching may also require sensitivity, habits of mind, and attention to particular mathematical principles, such as precision, definitions, reasoning, and coherence. Mathematics uses language in exact ways quite different from its uses in other contexts, including everyday life. Attention to careful use of quantifiers, relations, and logical terms is important. Being clear on the differences between "five apples," "at least five apples," "no more than five apples," and "exactly five apples," or between "if" and "if and only if" is important as young learners begin to express mathematical ideas, or as a teacher attentively uses a curriculum. A second, and closely related, aspect of mathematics is the importance of *definitions*. Terms, ideas, and concepts all require definition in the service of precision and to support reasoning.

The Task Group reviewed how teachers' knowledge of particular definitions, and perhaps equivalent alternatives and their relative advantages, make a difference for teaching quality. Another aspect for consideration may be a sense about the general role and nature of definitions within mathematics. Mathematical development and the solution of problems depend on specific approaches to logical argument and explanation. How does teachers' knowledge of reasoning play a role in teaching, considering both the teachers' own knowledge and their capacity to make mathematical reasoning and explanation accessible to and learnable by students? Being able to do that is rooted in a finely grained understanding of the nature of mathematical reasoning. Also worth investigating is the ability to see connections, and to appreciate and construct coherent links within and across ideas. How is early work with whole numbers connected to later encounters with integers, fractions, and the real number line? How are equivalent fractions related to the regrouping steps of the subtraction algorithm? What are useful geometric or physical models of arithmetic operations, and how can one explain the correspondences?

More work is needed to inspect how mathematical knowledge is needed for and deployed in teaching. Hypotheses about the requisite skill and knowledge can help to advance the question from a blunt investigation of credentials to a more nuanced understanding of the mathematical demands of teaching and the connection of those to students' learning gains. This is a crucial area for further and more precise study.

B. Teachers' Education: Teacher Preparation and Alternative Pathways to Teaching, Professional Development, and Induction

What kinds of programs have been shown to help teachers develop the necessary mathematical knowledge and skills needed for teaching?

a) How can preservice programs effectively increase beginning teachers' mathematical knowledge for teaching?
b) How can in-service programs do so?
c) Do particular designs or curricula make a difference for teachers' instructional skill and their students' achievement?
d) Is there evidence about how different kinds of professional preparation or requirements affect teachers' effectiveness, and how these compare?

Teacher education is regarded as key to building instructional quality and teacher effectiveness. The Task Group uses the term "teacher education" here to refer to four different types of professional training:

- *Preservice teacher preparation*: Initial teacher training, conventionally offered in institutions of higher education;

- *Alternative pathways*: Initial teacher preparation, offered outside of conventional teacher education programs;
- *Induction programs*: Programs of professional support and additional training within the first years of practice; and
- *Professional development*: Ongoing programmatic professional education of practicing teachers.

The Task Group sought to examine the evidence on teacher education in these four forms, asking about the relationship between different forms of teacher education and the learning of teachers and their students.

Many beliefs exist about what constitutes effective professional training for teachers, including what teachers should learn, how it should be structured and taught, and how much training is needed. There are also beliefs about learning from experience, and about what can be learned in formal settings or from practice. Many authoritatively stated positions assert "what we know" about "good" professional development. The Task Group wanted to learn what is known about particular curricula, structures, or approaches to teacher education, and their effects on gains in teachers' mathematical knowledge and skill for teaching, and their demonstrated relationships to students' achievement gains. The Task Group focused on these two outcomes to learn about effective professional training; although other outcomes (e.g., professional satisfaction, retention, teachers' reports of usefulness or relevance of particular programs) may be informative, they fall short of providing links between professional education and actual effects on learning outcomes. To help inform sensible allocation of resources for professional education and to provide direction for continued research in this area, the Task Group focused directly on the current state of knowledge about these effects of professional education. The Task Group's results highlight the critical need for more and better studies tracing the relationship between specific approaches to teacher education (i.e., curricula, pedagogy and assessment, instructors, structures, and settings) and teachers' capacity for teaching and their students' learning.

1. Preservice Teacher Preparation

The Task Group identified and synthesized peer-reviewed research and national reports to answer questions regarding the impact of preparation programs for teachers. Most of the studies found were descriptive—that is, they provided information about programs, described the characteristics of individuals who enrolled in or completed such programs, or simply compared students before and after a program or class without any comparison group. Other studies looked at limited relationships, such as the effect of being taught by a teacher who is certified in mathematics. Such studies were not useful for the question the Task Group sought to answer about features of preservice teacher preparation and their effects on teacher knowledge or student achievement.

Five empirical studies were found that addressed the question related to impacts of preparation programs on student achievement or teachers' mathematical content knowledge. Two of the studies examined effects on student achievement (Levine, 2006; Noell, 2006); the other three examined impacts on teacher mathematics content knowledge (Koehler & Lehrer, 1998; McDevitt, Troyer, Ambrosio, Heikkinen, & Warren, 1995; National Center for Research on Teacher Learning, 1991). These studies employed a variety of analytical techniques, but none was of sufficient rigor or quality to allow the Task Group to draw conclusions about the relationship of particular features of teacher preparation programs and their effects. Only two of the five were peer-reviewed and none controlled for all the relevant factors that might explain variation in impact on either teachers' knowledge or their students' learning.

In Tables 7 and 8, a synthesis is provided of the empirical evidence uncovered providing information on both study characteristics and study findings. Overall, however, the Task Group was unable to draw conclusions from this body of evidence.

Table 7: Quality Characteristics of Models Examining Impact of Teacher Preparation Programs on Student Achievement in Mathematics or Teacher Mathematics Content Knowledge

			Other Controls					
Authors	**Sample Size**	**Pretest Control**	**Student**	**Teacher**	**Class, School, or District**	**Family**	**Matching of Schools**	**Level of Analysis**
Student Achievement in Mathematics								
Levine, 2006	1,611 math teachers, over 30 million observations of students	X		X	X	X	X	Teacher
Noell, 2006	Over 200,000 students in Grades 4–9	X	X		X	X		Student
Teacher Mathematics Content Knowledge								
Koehler, & Lehrer, 1998	10 preservice teachers	X						Preservice teacher
McDevitt et al., 1995	About 150 preservice teachers in the first cohort, 110 in the second cohort							Preservice teacher
National Center for Research on Teacher Learning, 1991	Longitudinal study of over 100 participants in 5 different programs	X						Preservice teacher

Table 8: Reported Impacts From Models Examining the Effect of Teacher Preparation Programs on Student Achievement in Mathematics or Teacher Mathematics Content Knowledge

Authors	Dependent Variable	Independent Variable	Results of Estimated Effects	Math Specific Outcome	Grade(s)	Estimation Technique
Student Achievement in Mathematics						
Levine, 2006	NWEA achievement tests	Teacher trained at NCATE accredited school	Positive & Not Significant	X		ANOVA[a]
	NWEA achievement tests	Teacher trained at doctoral/research university (vs. Masters I)	Positive & Significant	X		ANOVA
Noell, 2006	Louisiana standardized test scores in math	Teacher attended Undergraduate Univ B	Positive & Not Significant	X	4–9	Hierarchical Linear Models (HLM)
	Louisiana standardized test scores in math	Teacher attended Undergraduate Univ E	Negative & Not Significant	X	4–9	Hierarchical Linear Models (HLM)
	Louisiana standardized test scores in math	Teacher attended Undergraduate Univ H	Positive & Not Significant	X	4–9	Hierarchical Linear Models (HLM)
	Louisiana standardized test scores in math	Teacher attended Undergraduate Univ K	Negative & Not Significant	X	4–9	Hierarchical Linear Models (HLM)
	Louisiana standardized test scores in math	Teacher attended Undergraduate Univ L	Positive & Not Significant	X	4–9	Hierarchical Linear Models (HLM)
	Louisiana standardized test scores in math	Teacher attended Undergraduate Univ M	Negative & Not Significant	X	4–9	Hierarchical Linear Models (HLM)
	Louisiana standardized test scores in math	Teacher attended Undergraduate Univ N	Negative & Not Significant	X	4–9	Hierarchical Linear Models (HLM)
	Louisiana standardized test scores in math	Teacher attended Alternative Cert. Univ M	Negative & Not Significant	X	4–9	Hierarchical Linear Models (HLM)
	Louisiana standardized test scores in math	Teacher attended Alternative Cert. Univ P	Positive & Not Significant	X	4–9	Hierarchical Linear Models (HLM)
	Louisiana standardized test scores in math	Teacher attended Undergraduate Univ A	Negative & Not Significant	X	4–9	Hierarchical Linear Models (HLM)
	Louisiana standardized test scores in math	Teacher attended Undergraduate Univ D	Negative & Not Significant	X	4–9	Hierarchical Linear Models (HLM)
	Louisiana standardized test scores in math	Teacher attended Undergraduate Univ F	Negative & Not Significant	X	4–9	Hierarchical Linear Models (HLM)
	Louisiana standardized test scores in math	Teacher attended Undergraduate Univ G	Negative & Not Significant	X	4–9	Hierarchical Linear Models (HLM)
	Louisiana standardized test scores in math	Teacher attended Undergraduate Univ J	Negative & Not Significant	X	4–9	Hierarchical Linear Models (HLM)
	Louisiana standardized test scores in math	Teacher attended Undergraduate Univ I	Negative & Significant	X	4–9	Hierarchical Linear Models (HLM)
	Louisiana standardized test scores in math	Teacher attended Alternative Cert. Univ B	Negative & Significant	X	4–9	Hierarchical Linear Models (HLM)
	Louisiana standardized test scores in math	Teacher attended Alternative Cert. Univ G	Negative & Significant	X	4–9	Hierarchical Linear Models (HLM)
	Louisiana standardized test scores in math	Teacher attended Alternative Cert. Univ L	Negative & Significant	X	4–9	Hierarchical Linear Models (HLM)
Teacher Mathematics Content Knowledge						
Koehler, & Lehrer, 1998	Problem-type sorting task	Learning using a hypermedia tool (vs. text)	Positive & Not Significant	X	N/A[b]	Comparison of Means
	Solution-strategy sorting task	Learning using a hypermedia tool (vs. text)	Positive & Significant	X	N/A	Comparison of Means
McDevitt et al., 1995	Researcher-designed test given at the end of each math class	An experimental program (vs. standard courses)	Positive & Significant	X	N/A	Comparison of Means
National Center for Research on Teacher Learning, 1991	Score on interview tasks and questionnaire	Attended the only program in which developing pre-service teachers' meaningful knowledge was an explicit goal	Positive, Significance not reported	X	N/A	Comparison of Means

[a] Analysis of variance.
[b] N/A means that data were not available.

2. Alternative Pathways Into Teaching

The Task Group identified peer-reviewed research and national reports focused on the impact of alternative preparation programs for teachers. However, most of the studies located were only descriptive. These reports provided information about alternative programs or described the characteristics of individuals who enrolled in or completed such programs.

The Task Group found 10 empirical studies that examined effects of alternative preparation programs on student achievement; all but one used correlational techniques (one was experimental). No empirical studies were found on the impacts of alternative pathways programs on teachers' mathematical content knowledge. Tables 9 and 10 provide a synthesis of the relevant empirical evidence, and information on both study characteristics and study findings. Wherever possible, standardized regression coefficients were calculated. Overall, evidence is mixed on programmatic effects of teachers' pathways into teaching and their relationship to students' achievement. As shown in Table 10, four studies show positive effects for teachers trained in alternative pathways [e.g., Teach for America (TFA)]. Boyd, Grossman, Lankford, Loeb, & Wykoff (2006) found small differences in students' mathematics achievement that could be associated with teacher preparation pathways and these effects were only for first-year teachers working with students in the sixth through eighth grades. Out of 18 different comparisons of different alternative pathways and "college-recommended pathways" investigated, 5 showed significant effects, with 4 of those showing positive effects for an alternative pathway, and 1 showing a negative effect. Overall, according to the authors, variation within pathways tended to be greater than variation across pathways. Decker, Mayer, and Glazerman (2004) found that students in TFA classrooms outperformed their peers, though the size of these effects varied with the characteristics of students (e.g., gender, mobility, and prior achievement status).

Kane, Rockoff, and Staiger (2006) showed significant positive effects for one alternative pathway into teaching (Teach for America, TFA) and significant negative effects for a second alternative pathway (International Program). Results were also inconclusive in the research reported by Darling-Hammond, Holtzman, Gatlin, and Heilig (2005), who found more negative effects for teachers prepared through alternative pathways than for those prepared in traditional certification programs. In Raymond, Fletcher, and Luque (2001), the students of new TFA teachers outperformed those of new traditionally trained teachers at Grades 4 and 5, while there was no difference for teachers overall. At the middle school level, however, the difference was significant for all teachers overall but insignificant for new teachers. In contrast, Miller, McKenna, and McKenna (1998) found no differences between the mathematics achievements of fourth- or fifth-grade students whose teachers were prepared through alternative pathways and those prepared in traditional programs. Additionally, Laczko-Kerr and Berliner (2002) found that students of under-qualified teachers, including TFA teachers, performed less well on mathematics tests than those of comparably experienced certified teachers.

Interpreting the evidence across these studies is not easy. One significant problem is definitional and related to the treatment conditions; "alternative pathways" does not define a clear programmatic type of teacher preparation. Some studies examined certification status while others compared different pathways to certification. Similarly, there is lack of clear specification of the traditional preparation programs to which these alternatives are being compared. These programs—alternative and traditional—are all forms of preservice teacher preparation, and the studies do not probe into key curricular, structural, or other programmatic variables that would permit analysis of the programs and their effects. A second problem rests with massive differences in the design and measures of students' achievement, making it difficult to compare across studies. A third issue is that the programs that tend to be studied over-sampled from one state (New York) and, thus, represent a narrow range geographically and only a tiny fraction of teachers in the system with over-sampling from more elite populations. Drawing conclusions about alternative pathways, in general, would be difficult to do from these studies.

Determining how different types of pathways into teacher preparation may affect teachers' capacity to teach is a key policy question in a time when issues of teacher recruitment, retention, and quality are paramount. Extant evidence suggests that there are no significant differences among current pathways, and, as Boyd et al. (2006) report, variation within programs appears to be greater than that uncovered across programs. Studies that more clearly specified the alternatives to be compared and the outcome measures used, and that used common or comparable designs would help in investigating this important question with more precision and focus.

Table 9: Quality Characteristics of Models Examining Impact of Alternative Pathways On Student Achievement in Mathematics

			Other Controls					
Authors	**Sample Size**	**Pretest Control**	**Student**	**Teacher**	**Class, School, or District**	**Family**	**Matching of Teachers**	**Level of Analysis**
Boyd, Grossman, Loeb, Lankford, & Wyckoff, 2006	1,035,949 Grades 3–8 student-year observations	Yes	X	X	X	X		Student
Fetler, 1999	Total: 921,437 Grade 9: 347,201 Grade 10: 313,303 Grade 11: 260,933 students in schools across CA			X	X	X		School
Goldhaber, & Brewer, 2000	3,786 Grade 12 math students 2,098 math teachers	X	X	X	X	X		Student
Kane, Rockoff, & Staiger, 2006	1,462,100 student-year observations, for Grades 3–8 in NYC schools from 1998–2005	X	X	X	X			Student
Tatto, Nielsen, Cummings, Kularatna, & Dharmadasa, 1993	216 teachers in Sri Lanka		X	X	X			Student
Decker, Mayer, & Glazerman, 2004	1,893 students in across 100 classrooms	X	X	X	X	X		Student
Darling-Hammond, Holtzman, Gatlin, & Heilig, 2005	271,015 students in the Houston ISD, Grades 3 and higher from 1995–96 to 2001–02	X	X	X	X	X		Student
Laczko-Kerr, & Berliner, 2002	232 newly hired teachers across 5 AZ school districts from 1998–2000			X	X		X	Classroom
Raymond, Fletcher, & Luque, 2001	81,814 students in Grades 4 & 5 96,276 students in Grades 6 & 8	X	X	X	X	X		Student
Miller, McKenna & McKenna, 1998	345 students in 18 middle school classrooms in GA			X			X	Student

Table 10: Reported Impacts of Studies Examining the Effect of Teachers From Alternative Paths on Student Achievement in Mathematics

Authors	Dependent Variable	Independent Variable	Standardized Regression Coefficients	Results of Estimated Effects	Math Specific Outcome	Grade(s)	Estimation Technique
Boyd, Grossman, Loeb, Lankford, & Wyckoff, 2006	NYC standardized math exam scores	NYC Teaching Fellow (vs. College Recomm.) Over 1 year	Unknown	Negative & Significant	X	4–5	OLS Regression
	NYC standardized math exam scores	NYC Teaching Fellow (vs. College Recomm.) Over 2 years	Unknown	Positive & Not Significant	X	4–5	OLS Regression
	NYC standardized math exam scores	NYC Teaching Fellow (vs. College Recomm.) Over 3 years	Unknown	Positive & Not Significant	X	4–5	OLS Regression
	NYC standardized math exam scores	TFA Teacher (vs. College Recomm.) Over 1 year	Unknown	Negative & Not Significant	X	4–5	OLS Regression
	NYC standardized math exam scores	TFA Teacher (vs. College Recomm.) Over 2 year	Unknown	Positive & Not Significant	X	4–5	OLS Regression
	NYC standardized math exam scores	TFA Teacher (vs. College Recomm.) Over 3 years	Unknown	Positive & Not Significant	X	4–5	OLS Regression
	NYC standardized math exam scores	Temp. Licensed (vs. College Recomm.) Over 1 year	Unknown	Negative & Not Significant	X	4–5	OLS Regression
	NYC standardized math exam scores	Temp. Licensed (vs. College Recomm.) Over 2 years	Unknown	Positive & Not Significant	X	4–5	OLS Regression
	NYC standardized math exam scores	Temp. Licensed (vs. College Recomm.) Over 3 years	Unknown	Positive & Significant	X	4–5	OLS Regression
	NYC standardized math exam scores	NYC Teaching Fellow (vs. College Recomm.) Over 1 year	Unknown	Negative & Not Significant	X	6–8	OLS Regression
	NYC standardized math exam scores	NYC Teaching Fellow (vs. College Recomm.) Over 2 years	Unknown	Positive & Not Significant	X	6–8	OLS Regression
	NYC standardized math exam scores	NYC Teaching Fellow (vs. College Recomm.) Over 3 years	Unknown	Positive & Significant	X	6–8	OLS Regression
	NYC standardized math exam scores	TFA Teacher (vs. College Recomm.) Over 1 year	Unknown	Positive & Significant	X	6–8	OLS Regression
	NYC standardized math exam scores	TFA Teacher (vs. College Recomm.) Over 2 years	Unknown	Positive & Not Significant	X	6–8	OLS Regression

Continued on p. 5-31

Table 10, continued

Authors	Dependent Variable	Independent Variable	Standardized Regression Coefficients	Results of Estimated Effects	Math Specific Outcome	Grade(s)	Estimation Technique
Boyd, Grossman, Loeb, Lankford, & Wyckoff, 2006	NYC standardized math exam scores	TFA Teacher (vs. College Recomm.) Over 3 years	Unknown	Positive & Not Significant	X	6–8	OLS Regression
	NYC standardized math exam scores	Temp. Licensed (vs. College Recomm.) Over 1 year	Unknown	Negative & Not Significant	X	6–8	OLS Regression
	NYC standardized math exam scores	Temp. Licensed (vs. College Recomm.) Over 2 years	Unknown	Positive & Significant	X	6–8	OLS Regression
	NYC standardized math exam scores	Temp. Licensed (vs. College Recomm.) Over 3 years	Unknown	Positive & Not Significant	X	6–8	OLS Regression
Fetler, 1999	Stanford Achievement Test scores in math	Percent Emergency Certified Teachers	Unknown	Negative & Indeterminate	X	9	OLS Regression
	Stanford Achievement Test scores in math	Percent Emergency Certified Teachers	Unknown	Negative & Indeterminate	X	10	OLS Regression
	Stanford Achievement Test scores in math	Percent Emergency Certified Teachers	Unknown	Negative & Indeterminate	X	11	OLS Regression
Goldhaber, & Brewer, 2000	Standardized test scores in math, for test designed by the ETS (NELS:88)	Probationary Certification (vs. Traditionally Certified)	0.00	Positive & Not Significant	X	12	OLS Regression
	Standardized test scores in math, for test designed by the ETS (NELS:88)	Emergency Certification (vs. Traditionally Certified)	0.00	Positive & Not Significant	X	12	OLS Regression
	Standardized test scores in math, for test designed by the ETS (NELS:88)	Not Certified (vs. Traditionally Certified)	-0.01	Negative & Significant	X	12	OLS Regression

Continued on p. 5-32

Table 10, continued

Authors	Dependent Variable	Independent Variable	Standardized Regression Coefficients	Results of Estimated Effects	Math Specific Outcome	Grade(s)	Estimation Technique
Kane, Rockoff, & Staiger, 2006	NYC standardized math exam scores	NYC Teaching Fellow (vs. Traditionally Certified)	0.00	Positive & Not Significant	X	3–8	OLS Regression
	NYC standardized math exam scores	TFA Teacher (vs. Traditionally Certified)	0.01	Positive & Significant	X	3–8	OLS Regression
	NYC standardized math exam scores	Int’l Program Teacher (vs. Traditionally Certified)	0.00	Negative & Significant	X	3–8	OLS Regression
	NYC standardized math exam scores	Not Certified (vs. Traditionally Certified)	0.00	Zero	X	3–8	OLS Regression
Tatto, Nielsen, Cummings, Kularatna, & Dharmadasa, 1993	Sri-Lankan test of math skills	Untrained Teacher (vs. Teacher College)	Unknown	Positive & Significant	X	4	Comparison of Means
	Sri-Lankan test of math skills	Untrained Teacher (vs. College of Education)	Unknown	Positive & Significant	X	4	Comparison of Means
	Sri-Lankan test of math skills	Untrained Teacher (vs. Distance Education)	Unknown	Positive & Not Significant	X	4	Comparison of Means
Decker, Mayer, & Glazerman, 2004	ITBS math scores	TFA Teacher	Unknown	Positive & Significant	X	1–5	HLM Modeling
Darling-Hammond, Holtzman, Gatlin, & Heilig, 2005	TAAS score in math	TFA Teacher	Unknown	Positive & Significant	X	3–HS	HLM Modeling
	Stanford Achievement Test score in math	TFA Teacher	Unknown	Negative & Significant	X	3–HS	HLM Modeling
	Aprenda score in math	TFA Teacher	Unknown	Negative & Significant	X	3–HS	HLM Modeling
	TAAS score in math	Alternative Certification	Unknown	Negative & Significant	X	3–HS	HLM Modeling
	Stanford Achievement Test score in math	Alternative Certification	Unknown	Negative & Significant	X	3–HS	HLM Modeling
	Aprenda score in math	Alternative Certification	Unknown	Negative & Not Significant	X	3–HS	HLM Modeling
	TAAS score in math	Emer./Temp. Certification	Unknown	Negative & Significant	X	3–HS	HLM Modeling
	Stanford Achievement Test score in math	Emer./Temp. Certification	Unknown	Negative & Not Significant	X	3–HS	HLM Modeling
	Aprenda score in math	Emer./Temp. Certification	Unknown	Negative & Not Significant	X	3–HS	HLM Modeling

Continued on p. 5-33

Table 10, continued

Authors	Dependent Variable	Independent Variable	Standardized Regression Coefficients	Results of Estimated Effects	Math Specific Outcome	Grade(s)	Estimation Technique
Laczko-Kerr, & Berliner, 2002	Average Stanford Achievement Test score in math in 1998	Certified Teacher (vs. Noncertified)	Unknown	Positive & Not Significant	X	2–8	Analysis of variance (ANOVA) and Comparisons of means
	Average Stanford Achievement Test score in math in 1999	Certified Teacher (vs. Noncertified)	Unknown	Positive & Significant	X	2–8	ANOVA and Comparisons of means
	Average Stanford Achievement Test score in math in 1998	Not TFA (vs. TFA teachers)	Unknown	Positive & Not Significant	X	2–8	ANOVA and Comparisons of means
	Average Stanford Achievement Test score in math in 1999	Not TFA (vs. TFA teachers)	Unknown	Positive & Significant	X	2–8	ANOVA and Comparisons of means
Raymond, Fletcher, & Luque, 2001	Average TAAS math score	TFA Teacher (vs. non-TFA teachers, for all teachers)	Unknown	Positive & Not Significant	X	4–5	OLS regression
	Average TAAS math score	TFA Teacher (vs. non-TFA teachers, w/ < 1 yr exp)	Unknown	Positive & Significant	X	4–5	OLS regression
	Average TAAS math score	TFA Teacher (vs. non-TFA teachers, for all teachers)	Unknown	Positive & Significant	X	6–7	OLS regression
	Average TAAS math score	TFA Teacher (vs. non-TFA teachers, w/ < 1 yr exp)	Unknown	Positive & Not Significant	X	6–7	OLS regression
Miller, McKenna, & McKenna, 1998	ITBS reading and math scores	Alternative Certification	Unknown	Positive & Not Significant	X	5–6	Multivariate Analysis of variance (MANOVA) and Comparisons of means

3. Induction Programs

National reports calling for higher-quality teaching, higher teacher retention rates, and stronger student achievement identify support of new teachers—or induction—as an area for improvement (National Commission on Mathematics and Science Teaching for the 21st Century, 2000; National Commission on Teaching and America's Future, 2003; National Council of Teachers of Mathematics, 2002). As the goal of many induction programs is acclimation of new teachers to the school, they often last only the first year. Other induction programs are positioned as the first step on a continuum of professional development for teachers and are multiyear in nature. The assignment of mentors is frequently part, or all, of an induction program.

Feiman-Nemser, Schwille, Carver, and Yusko (1999) discuss three definitions for induction: 1) a unique phase or stage in teacher development, 2) a time for socialization as the teacher transitions from preparation to practice, and 3) a formal program for beginning teachers. In this review, as is common, the term “induction” was taken to mean a formal program for beginning teachers, including mentoring programs. Induction programs are varied in the length of time a person participates, the scope of support provided, and the guidance provided by policies and mandates (Feiman-Nemser et al.). Some states, such as California and Connecticut, have statewide induction programs.

Frequently, at least in the United States, induction is viewed as synonymous with mentoring; however, this is not accurate. Mentoring is a frequent component of induction programs, but it is not the only component. Four goals of induction programs were identified by the National Commission on Teaching and America’s Future (Fulton, Yoon, & Lee, 2005): 1) building and deepening teacher knowledge, 2) integration of new practitioners into a teaching community and school culture that support continuous professional growth of all, 3) support for the constant development of the teaching community in the school, and 4) encouragement of a professional dialogue to articulate the goals, values, and best practices of the community. Induction programs should be systems which are “networks of supports, people, and processes that are all focused on assuring that novices become effective in their work” (p. 4).

The concept of induction is not new in education. Although calls for programs to support new teachers can be dated to the 1960s, initially induction programs were uncommon. Only Florida had a mandated induction program prior to 1980 (Feiman-Nemser et al., 1999). Over the years induction programs, and mandates to support induction programs, increased due to ties between induction and such key issues as school reform, teacher retention, teacher quality, and achievement initiatives.

The Task Group searched for empirical investigations of the effectiveness of teacher induction programs—broadly defined—on teacher mathematics knowledge and student achievement. Careful systematic review of the literature uncovered a dearth of peer-reviewed research on induction. Reviews of induction programs—frequently focusing on retention—were examined (e.g., Feiman-Nemser et al., 1999; Glazerman, Senesky, Seftor, & Johnson, 2006; Ingersoll & Kralik, 2004; Kagan, 1992; Lopez, Lash, Schaffner, Shields, & Wagner, 2004; Totterdell, Bubb, Woodroffe, & Hanrahan, 2004). Overall, they identified more than 100 potentially relevant pieces of literature. However, none focused on the effects of induction for mathematics teachers on student achievement or teacher mathematics knowledge.

Literature examining induction is not scarce. However, rarely does it focus specifically on teachers of mathematics—although some focuses on both mathematics and science teachers (Adams & Krockover, 1997; Davis, Petish, & Smithey, 2006; Garet, Porter, Desimone, Birman, & Yoon, 2001; McGinnis, Parker, & Graeber, 2004). Much of the work is program evaluation of particular initiatives and is not peer-reviewed. For examples, see two programs in California: the New Teacher Center at the University of California Santa Cruz (www.newteachercenter.org), and California’s Beginning Teacher Support and Assessment program (www.btsa.ca.gov). In addition, many of the studies are case studies or qualitative in nature. The empirical evidence on outcomes of the programs is weak as none of the studies reviewed used random assignment, and few used a comparison group of any kind.

The key outcome for much of the extant induction literature is teacher retention. There is also a wealth of literature examining the effects of induction programs on teacher beliefs, satisfaction, and practices. Induction programs continue to expand, some mandated and some not. Given the expansion, it is important to assess the effectiveness of induction programs on outcomes, such as student achievement, and not rely only on evidence about teacher retention, satisfaction, and beliefs. Until induction programs are content-specific or include specific content activities, it may be difficult to determine the effectiveness of induction on the mathematical knowledge of teachers or on students' achievement gains.

4. Professional Development Programs

The final component of teacher education that the Task Group investigated was professional development for practicing teachers. The Task Group searched for peer-reviewed research and national reports that would offer high-quality evidence regarding the impact of professional development programs for teachers. Many of the studies identified were descriptive in that they provided information about the characteristics of the programs, and most of those that were empirical did not include a comparison group, but used a one-group pretest-posttest design. This was the case for every empirical study they identified that examined the effects of teacher professional development programs on teachers' mathematical content knowledge. Moreover, many of those relied on teacher self-reports about their knowledge before and after the professional development rather than on measures of teacher knowledge. As a result, this review includes only studies investigating the relationship between teacher professional development programs and students' mathematics achievement. In other words, the Task Group did not include studies with a pretest-posttest design, and thus no studies related to teacher mathematics content knowledge were included.

The Task Group found eight empirical studies, using a variety of analytical techniques, that examined effects of teacher professional development programs on student achievement. Tables 11 and 12 provide a synthesis of the relevant empirical evidence, and information on both study characteristics and study findings. For all studies except one (Jacob & Lefgren, 2002), the Task Group was able to calculate a standardized effect size using Hedge's *g*. These calculations are shown in Table 12. Accompanying this synthesis is a series of descriptive tables that provide more details on each study included in Tables 11 and 12.

Table 11: Quality Characteristics of Studies Examining Impact of Teacher Professional Development Programs on Student Achievement in Mathematics

			Other Controls					
Authors	**Sample Size**	**Pretest Control**	**Student**	**Teacher**	**Class, School, or District**	**Family**	**Matching of Schools**	**Level of Analysis**
Angrist, & Lavy, 2001	Elementary schools in Jerusalem, 9 intervention schools (7 secular, 2 religious) and 11 comparison schools (6 secular, 5 religious). Approximately 634 secular students; 196 religious students.	X	X		X	X	X	Student
Campbell, 1996	Elementary school teachers (K–Grade 3) in 6 schools in Montgomery County, MD. Treatment group: students in 3 treatment schools; comparison group: student in 3 schools in the same county. *n* = 149 Kindergarten and 292 Grade 1 students.						X	Student
Carpenter et al., 1989	40 Grade 1 teachers in 24 schools (2 private) in Madison, WI.	X						Teacher
Chapin, 1994	Elementary, middle and high school students of 42 teachers in Chelsea, MA, school district versus students in the same schools whose teachers were not involved in the program. Outcomes available for Grades 3 (*n* = 269), 6 (*n* = 244) and 7 (*n* = 210) students.							Student
Jacob, & Lefgren, 2002	246 Elementary schools in Chicago (Grades 3–6), approximately 47,000 students.	X	X		X	X	Regular discontinuity	Student
Karges-Bone, Collins, & Maness, 2002	1 elementary school, 61 Grade 3 students, 48 Grade 4 students.							Student
Saxe, Gearheart, & Nasir, 2001	23 elementary schools in Greater Los Angeles: 9 teachers in Integrated Mathematics Assessment (IMA) program, 8 teachers in collegial support program (SUPP); and 6 in traditional/no additional professional development program (TRAD).	X	X					Classroom
Van Haneghan, Pruet, & Bamberger, 2004	6 elementary schools (4 treatment, 2 comparison) in Mobile, AL. Followed two cohorts over the course of 3 years (approx 200 in K, and approx. 420 in Grade 3).						X	Student

Table 12: Reported Impacts of Studies Examining the Effect of Teacher Professional Development Programs on Student Achievement in Mathematics

Authors	Dependent Variable	Independent Variable	Effect size or Regression Coefficients	Results of Estimated Effects	Math Specific Outcomes	Grade(s)	Estimation Technique
Angrist & Lavy, 2001	1996 math score	Attending secular school with more intensive in-service training	0.431	Positive (p = .097)	X	6	Hedges g based on raw means (adjusted for differences in average pretest scores and adjusted for school-level clustering)
Campbell, 1996	Project-developed student achievement assessment	Project IMPACT vs. comparison	0.39	Positive, not significant	X	K	Hedges g (adjusted for school clustering)
	Project-developed student achievement assessment	Project IMPACT vs. comparison	0.14	Positive, not significant	X	1	Hedges g (adjusted for school clustering)
Carpenter et al., 1989	Computation – ITBS	CGI treatment vs. basic problem-solving workshop	0.41	Positive, not significant	X	1	Hedges g (pretest adjusted means, no adjust for clustering)
	Computation – Number facts	CGI treatment vs. basic problem-solving workshop	0.66	Positive, significant	X	1	Hedges g (pretest adjusted means, no adjust for clustering)
	Problem solving: ITBS	CGI treatment vs. basic problem-solving workshop	0.37	Positive, not significant	X	1	Hedges g (pretest adjusted means, no adjust for clustering)
	Problem solving: simple add/subtract	CGI treatment vs. basic problem-solving workshop	0.43	Positive, not significant	X	1	Hedges g (raw means since adjusted not available, no adjust for clustering)
	Problem solving: complex add/subtract	CGI treatment vs. basic problem-solving workshop	0.42	Positive, not significant	X	1	Hedges g (pretest adjusted means, no adjust for clustering)
	Problem solving: advanced	CGI treatment vs. basic problem-solving workshop	0.11	Positive, not significant	X	1	Hedges g (pretest adjusted means, no adjust for clustering)
	Problem solving: interview	CGI treatment vs. basic problem-solving workshop	0.69	Positive, significant	X	1	Hedges g (pretest adjusted means, no adjust for clustering)
Chapin, 1994	California Achievement Test	Students of CCTDM teachers versus non-project teachers	0.58	Positive, significant	X	3	Hedges g (adjusted for classroom clustering)
	California Achievement Test	Students of CCTDM teachers versus non-project teachers	0.76	Positive, significant	X	6	Hedges g (adjusted for classroom clustering)
	California Achievement Test	Students of CCTDM teachers versus non-project teachers	0.66	Positive, significant	X	7	Hedges g (adjusted for classroom clustering)
Jacob, & Lefgren, 2002	ITBS math score	Probation vs. no	OLS est = -.021 (se = .010)	Negative, not significant	X	3–6	OLS with student and school covariates
Karges-Bone, Collins, & Maness, 2002	BSAP	1997/98 cohort (post-treatment) vs. 1996 cohort	0.70	Positive, significant	X	3	Hedges g (student level)
	MAT	1997/98 cohort (post-treatment) vs. 1996 cohort	0.44	Positive, not significant	X	4	Hedges g (student level)

Continued on p. 5-38

Table 12, continued

Authors	Dependent Variable	Independent Variable	Effect size or Regression Coefficients	Results of Estimated Effects	Math Specific Outcomes	Grade(s)	Estimation Technique
Saxe, Gearheart, & Nasir, 2001	Concepts	IMA vs. TRAD	2.39	Positive, significant	X	4–5	Hedges g (classroom-level analysis)
	Concepts	SUPP vs. TRAD	0.67	Positive, not significant	X	4–5	Hedges g (classroom-level analysis)
	Concepts	IMA vs. SUPP	1.45	Positive, significant	X	4–5	Hedges g (classroom-level analysis)
	Computation	IMA vs. TRAD	-0.53	Negative, not significant	X	4–5	Hedges g (classroom-level analysis)
	Computation	SUPP vs. TRAD	-1.34	Negative, significant	X	4–5	Hedges g (classroom-level analysis)
	Computation	IMA vs. SUPP	0.77	Positive, not significant	X	4–5	Hedges g (classroom-level analysis)
Van Haneghan, Pruet, & Bamberger, 2004	Yr 1, Grade 3: SAT9 Problem Solving NCE	Students of trained MMI teachers vs. control schools	-0.04	Negative, not significant	X	3	Hedges g (student level, adjusted for school-level clustering)
	Yr 1, Grade 3: SAT9 Procedures NCE	Students of trained MMI teachers vs. control schools	0.29	Positive, not significant	X	3	Hedges g (student level, adjusted for school-level clustering)
	Yr 1, Grade 3: SAT9 Total NCE	Students of trained MMI teachers vs. control schools	0.15	Positive, not significant	X	3	Hedges g (student level, adjusted for school-level clustering)
	Yr 1, Grade 3: TIMSS-items	Students of trained MMI teachers vs. control schools	0.35	Positive, not significant	X	3	Hedges g (student level, adjusted for school-level clustering)
	Yr 2, Grade 4: SAT9 Problem Solving NCE	Students in MMI schools versus control schools	0.24	Positive, not significant	X	4	Hedges g (student-level, adjusted for school-level clustering)
	Yr 2, Grade 4: SAT9 Procedures NCE	Students in MMI schools versus control schools	0.24	Positive, not significant	X	4	Hedges g (student level, adjusted for school-level clustering)
	Yr 2, Grade 4: SAT9 Total NCE	Students in MMI schools versus control schools	0.27	Positive, not significant	X	4	Hedges g (student level, adjusted for school-level clustering)
	Yr 3, Grade 2: Fractions	Students in MMI schools versus control schools	0.20	Positive, not significant	X	2	Hedges g (student level, adjusted for school-level clustering)
	Yr 3, Grade 5: Fractions	Students in MMI schools versus control schools	0.70	Positive, not significant	X	5	Hedges g (student level, adjusted for school-level clustering)
	Yr 3, Grade 2: Geometry	Students in MMI schools versus control schools	0.01	Positive, not significant	X	2	Hedges g (student level, adjusted for school-level clustering)
	Yr 3, Grade 2: Mental Math	Students in MMI schools versus control schools	0.22	Positive, not significant	X	2	Hedges g (student level, adjusted for school-level clustering)
	Yr 3, Grade 5: Mental Math	Students in MMI schools versus control schools	0.46	Positive, not significant	X	5	Hedges g (student level, adjusted for school-level clustering)
	Yr 3, Grade 2: Numeration	Students in MMI schools versus control schools	0.11	Positive, not significant	X	2	Hedges g (student level, adjusted for school-level clustering)
	Yr 3, Grade 5: Numeration	Students in MMI schools versus control schools	0.36	Positive, not significant	X	5	Hedges g (student level, adjusted for school-level clustering)

Continued on p. 5-39

Table 12, continued

Authors	Dependent Variable	Independent Variable	Effect size or Regression Coefficients	Results of Estimated Effects	Math Specific Outcomes	Grade(s)	Estimation Technique
Van Haneghan, Pruet, & Bamberger, 2004	Yr 3, Grade 2: Story Problems	Students in MMI schools versus control schools	0.27	Positive, not significant	X	2	Hedges g (student level, adjusted for school-level clustering)
	Yr 3, Grade 5: Story Problems	Students in MMI schools versus control schools	0.47	Positive, not significant	X	5	Hedges g (student level, adjusted for school-level clustering)
	Yr 3, Grade 5: SAT9 Problem Solving NCE	Students in MMI schools versus control schools	0.57	Positive, not significant	X	5	Hedges g (student level, adjusted for school-level clustering)
	Yr 3, Grade 5: SAT9 Procedures NCE	Students in MMI schools versus control schools	0.67	Positive, not significant (p = .097)	X	5	Hedges g (student level, adjusted for school-level clustering)
	Yr 3, Grade 5: SAT9 Total NCE	Students in MMI schools versus control schools	0.66	Positive, not significant	X	5	Hedges g (student level, adjusted for school-level clustering)
	Yr 3, Grade 5: TIMSS items	Students in MMI schools versus control schools	0.71	Positive, not significant (p = .078)	X	5	Hedges g (student level, adjusted for school-level clustering)

Across these eight studies that investigated the relationships of professional development to students' achievement, few significant effects were identified. The study that yielded the most consistently positive effects was Chapin (1994). In this study, the achievement of 723 third-, sixth-, and seventh-grade students of 42 teachers who participated in a professional development program was compared with students in the same schools whose teachers were not involved in the program. The study design does not permit analysis of particular features of the professional development that might account for differences in teacher performance. Two other studies that showed positive effects on student achievement were Carpenter, Fennema, Peterson, Chiang, and Loaf (1989) and Saxe, Gearhart, and Nasir (2001). Carpenter et al. studied a professional development program in which teachers were provided with knowledge of students' number fact concepts and reasoning, and showed that students in experimental classes exceeded students in control classes in number fact knowledge and problem solving. Saxe et al. compared three different professional development programs: an extensive subject-matter-focused professional development program whose goal was to improve teacher understanding of content, and of student thinking and learning processes; a program focused on building collegial support; and a traditional in-service program.[20]

The researchers found significant differences for pupils' conceptual and computational outcomes with the subject-matter-focused program leading to greater effects on students' conceptual learning and the traditional program yielding greater impact on students' computational skills. Angrist and Lavy (2001) studied a professional development program that involved weekly meetings between trainers and teachers to review teaching methods and plans for the following week, based on a "humanistic mathematics" philosophy of teaching. Treatment schools received an average of 10.5 more hours per week in training than comparison schools. Students in the treatment classes where the intervention was

[20] "Subject-matter-focused" denotes a program in which mathematics is central to teachers learning opportunities, and "traditional" ones are more focused on pedagogical classroom processes.

implemented performed significantly better than students in control schools. Still, none of these studies offers clear signals about the features of professional development that affect teachers' capacity to teach or students' achievement gains.

Only one study (Jacob & Lefgren, 2002) directly investigated effects of the amount of professional development, a factor frequently thought to be crucial for effective professional development. This study reported no effects from modestly increased amounts of professional development on students' learning in schools that were placed on probation for poor achievement. Little can be concluded from this study as the professional development was not specifically focused on mathematics, and no detailed information was provided about the nature of this professional development other than that it varied widely and was administered by a variety of organizations.

Overall, the Task Group was not able to draw conclusions about the features of professional development that have an impact on students' achievement because of the paucity of studies that investigated this link. For the studies the Task Group did identify that probed this connection, specificity is lacking regarding the features of the professional development programs where effects were found.

Professional development is often regarded as one of the key policy levers for improving instruction and student achievement with currently practicing teachers. To probe this assumption, the Task Group sought to identify and review the available evidence about effects of professional development, and features that make a difference for either student outcomes or their teachers' capacity to teach. The Task Group uncovered no studies, however, of sufficient quality where the designs and measures permitted them to ask and answer questions about teachers' learning. Most studies used a simple pre- and posttest design with no comparison group or used self-report data on teachers' learning. To ascertain the impact of professional development on students' achievement, the Task Group did identify a small number of studies, but overall, these did not support any specific claims about the nature of professional development that affects teachers' effectiveness.

5. Conclusions

The Task Group reviewed research on teacher education, including preservice teacher preparation, alternative pathways into teaching, induction programs, and professional development for practicing teachers. Despite the many beliefs about effective teacher education in any of these forms, the Task Group did not find strong evidence for the relationships between teacher education, and either teachers' capacity to teach or their students' learning. Even for the few studies that did produce significant effects, so little was unpacked about the features of the training that might account for a program's impact that the Task Group was left without much greater insight into the crucial components of teacher education.

Note that most research in this domain lacks rigorous designs and measures and is descriptive more often than not. Without comparison groups, or designs that permit analysis of program effects, it is difficult to draw conclusions about how teacher education works or about what the key features of effective professional training are.

6. Recommendations

Studies are needed that use designs that lead to knowledge about the impact of different approaches to professional development and permit comparisons with other potential impacts on teacher capacity and their effectiveness (e.g., experience, curriculum, curriculum policy). Such research will depend not only on rigorous designs but also on valid and reliable measures of the key outcome variables: teachers' mathematical knowledge and skill, instructional quality, and student learning. Self-reported data cannot continue to be the main source of information about professional development outcomes.

Key questions on which robust evidence is needed include the following:

- Does teacher education (e.g., preservice training of different kinds, professional development, early career induction programs) have an impact on teachers' capacity to teach and on students' achievement?
- What are key features of teacher education (e.g., duration, structure, quantity, content, pedagogy, structure, relationship to practice) that have effects on teachers' capacity to teach and on students' achievement?
- How do contexts (e.g., school, students, teachers, policy) affect the outcomes of professional development?
- How do different amounts of teacher education affect outcomes and effects?

Given the vast investment made in teacher education and the call for more of it, knowledge about its effects is vitally needed. Efforts to build measures and to implement better research designs should be supported.

C. Teacher Incentives

What types of recruitment and retention strategies are used to attract and retain highly effective teachers of mathematics? How well do they work?

As the Task Group has previously noted, substantial differences in the mathematics achievement of students are attributable to differences in teachers. The Task Group has focused on the role of teachers' mathematical knowledge in predicting student achievement, with teachers' knowledge measured by proxies, such as certification, college course work, and scores on tests. Teachers' college course taking in mathematics is associated with student gains in high school. Teacher's mathematical knowledge as measured by tests is also linked to student achievement. Other characteristics of teachers including years of teaching experience, general cognitive ability, and selectivity of the institution awarding the teacher's baccalaureate degree are associated with teacher effectiveness (Greenwald, Hedges, & Laine, 1996). It is important to note, however, that the largest reported positive effect of any of these characteristics is small relative to the magnitude of the natural variation in teacher effectiveness.

Thus the Task Group knows that there are large differences in the on-the-job performance of teachers of mathematics as measured by student gains. Some portion of the differences in teacher effectiveness can be predicted by such known characteristics of teachers as their college course work and their scores on tests. But prior on-the-job performance of teachers is by far the strongest predictor of their future on-the-job performance.

1. Utilizing Labor Market Incentives for Good Teaching

One possible mechanism for recognizing and leveraging differences in teaching ability is salary. According to the National Center for Education Statistics' Schools and Staffing Survey (Gruber, Wiley, Broughman, Strizek, & Burian-Fitzgerald, 2002), 70% of public school teachers in the United States in 2000 worked under a "uniform salary schedule" in which teachers with the same number of years of employment and the same level of postsecondary education received the same pay. In private industry and higher education, in contrast, pay is typically contingent on performance and area of specialization, as well as years of experience and level of education. In universities, for example, economists typically receive higher salaries than historians, reflecting the greater demand for economists outside the university sector. Within academia, economists who publish more influential work and bring in more external funding are paid more than less productive economists. Parallels in K–12 education would take the form of paying more to teachers who have technical skills that are in demand in other sectors of the economy, such as teachers with degrees in mathematics (*skills-based pay*), and paying more to mathematics teachers who are more productive in raising student achievement (*performance-based pay*). Another type of incentive intends to compensate teachers for working in conditions they view as unfavorable, such as those associated with high-poverty, low-achieving schools (*location pay*).

a. Skills-Based Pay

Skills-based teacher incentives are based on two premises: Certain types of preparation and training are necessary to teach certain subjects, and individuals with that preparation and training are less likely to enter into or remain in teaching if their levels of compensation are substantially below market rates. Evidence presented previously is consistent with the first premise in demonstrating a relationship between certain types of teacher preparation and student outcomes. With respect to the second premise, a large and consistent body of economic research indicates that college students' decisions to prepare for and enter into teaching depend on how the salary structure for teachers compares with those in competing occupations (Dolton & van der Kaauw, 1995; Bacolod, 2007; Goldhaber, DeArmond, Liu, & Player, 2007).

The magnitude of the salary differential between the private sector and the teaching profession for those who enter teaching with technical training is large. Four years after graduation, the gap in annual salary between teachers and non-teachers who have training in math and science is $13,469. Ten years out of college, the annual salary gap is $27,890 (Goldhaber et al., 2007). The salary differential for mathematically trained teachers versus non-teachers is likely to account at least in part for the significant teacher shortage in mathematics. In 2003–04, 74.1% of public high schools reported having teaching vacancies in mathematics, and 32.4% of those schools indicated that it was very difficult, or they were not able to fill those vacancies. Both of these levels were higher than for any other field in which high schools reported vacancies (Strizek, Pittsonberger, Riordan, Lyter, & Orlofsky, 2006). Differential salaries may also be responsible in part for the higher attrition rate from teaching of teachers with training in mathematics and science; Ingersoll (2000) reports that math and science teachers are significantly more likely to move from or leave their teaching jobs because of job dissatisfaction than are other teachers (40% of math and science, and 29% of all teachers). Of those who depart because of job dissatisfaction, the most common reason given is low salaries (57% of respondents).

b. Location Pay

"Location" pay is premised on the well-documented tendency of the most qualified teachers to select or migrate towards schools that serve the most economically advantaged children (Lankford, Loeb, & Wyckoff, 2002; Hanushek, Kain, & Rivkin 2004; Reed, Rueben, & Barbour 2006). Could this problem be remedied by paying teachers more who serve in high-needs schools? Research on the effects of location pay provides mixed results that are likely affected by the size of differential pay, the gender and experience of the teacher, and whether the bonus is a one-time signing bonus or permanent, among other factors (Hanushek et al.; Loeb & Page, 2000). Hanushek et al., using longitudinal data from Texas and taking advantage of naturally occurring variation in teacher salaries across districts, found that women were much less responsive to salary differences than men in determining whether to transition out of a high-minority school. They estimate that it would require an 8.8% salary premium for nonminority males with 3–5 years of teaching experience to keep them from moving from large urban to suburban districts, but a 42.6% differential to retain nonminority females. On the assumption that location pay could not be targeted to male teachers, they concluded that offering teachers pay differentials to take jobs in low-performing schools is not a cost-effective means of improving achievement. In contrast, Clotfelter, Glennie, Ladd, and Vigdor (2006) found that a moderately-sized addition to salary ($1,800) was effective in encouraging mid-career and more senior math and science teachers to stay in high-needs districts in North Carolina.

One important difference in the two studies is that Hanushek et al. (2004) estimated the size of the incentive that would neutralize teacher movement out of high-poverty urban schools, whereas Clotfelter et al. (2006) estimated the effect of the particular $1,800 bonus used in North Carolina, which was a 12% reduction in turnover rates. It may require much lower levels of location pay to reduce the outflow of experienced teachers from high-needs schools than it would take to eliminate it.

c. Performance Pay

Both skills-based pay and location pay as currently conceptualized and implemented provide incentives based on characteristics of teachers, such as college course work and experience, which are relatively weak predictors of student achievement. Thus they may be relatively inefficient mechanisms for enhancing the supply of effective teachers, where "effective" is defined as a teacher's above-average ability to increase the measured academic achievement of students. If, as previously documented, the strongest predictor of a teacher's effectiveness is the teacher's history of effectiveness, and if teachers' performance is affected by salary, then pay-for-performance might generate greater yields than similar investments in other incentives.

What is the evidence that pay-for-performance, or merit pay, has positive effects on teaching quality in mathematics? Before addressing that question it is important to note and briefly describe the substantial variability in the design of extant merit-pay systems.

One major categorical distinction is between incentive schemes focused on individuals versus those focused on schools. In the latter, teachers within schools receive a bonus if the entire school in which they teach makes progress on measures of student learning. In the former, individual teachers receive salary increments based on the gains of their own students over the course of the school year. School-based incentive plans typically have political advantages over individual plans, and some have argued that they enhance cooperation among teachers compared to individual plans. But others have argued that school-based plans, depending as they do on cooperation and continuity of effort by the teaching staff, are poorly designed for schools in which teacher mobility is high, a striking characteristic of many central city, high-poverty schools.

A second important dimension of variation among pay-for-performance plans is the level of compensation bonus that is available for higher performing teachers. Existing research on pay-for-performance involves bonuses that range from a couple of hundred dollars (Lavy, 2002) up to 40% or more of base salary (Glewwe, Ilias, & Kremer, 2003).

Another distinction involves the degree to which the merit system is focused on student outcomes. In some systems, teacher bonuses are totally dependent on student gains on standardized assessments of learning (e.g., Winters, Ritter, Barnett, & Greene, 2006). At the other end of the dimension are systems in which input and output measures are mixed together in complex arrays. An example is Mexico’s Carrera Magisterial, which rewards teachers with salary bonuses based on a number of criteria, such as seniority, educational attainment, professional development, teacher performance, and student achievement (Santibáñez et al., 2007).

Continuity is yet another dimension on which pay-for-performance systems differ. The majority of existing pay-for-performance systems is put forward as trials of the concept. A system that seems temporary may motivate teachers differently than one around which longer-term plans can be made.

The Task Group identified 14 quantitative studies on teacher merit pay (see Table 13). Of these, 13 showed positive effects on student outcomes (see Table 14). These studies varied methodologically from randomized control trials (Muralidaran, & Sundararaman, 2006; Glewwe et al., 2003) to causally weaker correlational studies (e.g., Figlio & Kenny, 2007) and quasi-experiments (e.g., Ladd, 1999). The design of the performance pay plans varied across the studies on all the design features identified in this section. Given the variability in program design and evaluation methodology, it is striking that each of the studies found some positive effects on student achievement.

Table 13: Quality Characteristics of Studies Examining Impact of Teacher Pay for Performance on Student Achievement in Mathematics

Authors	Sample Size	Pretest Control	Other Controls: Student	Other Controls: Teacher	Other Controls: Class, School, or District	Other Controls: Family	Matching of Schools	Level of Analysis	Incentive Scheme	Math Specific Test
Atkinson et al., 2004	182 secondary school teachers in the U.K. across 18 schools (with 23,000 students)	X	X	X				Teacher	Teacher-based: greater than 9% increase in salary	X
Cooper, & Cohn, 1997	541 classes, comprised of 532 teachers and 13,646 students in 18 school districts	X		X	X	X		Class/Teacher	2 plans—teacher-based or teacher- and school-based: each with similar pay incentives from $2,000–$3,000	X
Dee, & Keys, 2004	24,000 observations pooled from over 11, 000 students in 79 participating schools over 4 years	X	X	X	X	X		Student	Teacher-based: $2,000–$7,000 raise	X
Eberts, Hollenbeck, & Stone, 2002	2 high schools in MI (collectively comprised of over 17,000 student/year observations)	X					X	School	Teacher-based: student retention bonus 12–12.5% of base pay, performance bonus 5% of base pay, additional 10% bonus for both	
Figlio, & Kenny, 2006	4,515 students	X	X		X	X		Student	Varied within sample	
Glewwe et al., 2003	Over 50,000 students from 100 primary schools	X	X		X			Student	School-based: 21–43% of base salary	X
Ladd, 1999[a]	1,118 school-year observations in 5 cities				X			School	School-based: $1,000 to teacher and principals, and $2,000 to schools	X
Lavy, 2002	Over 22,000 observations across 190 schools		X		X	X	X	Student	2 school-based plans: amount varies by plan	X
Lavy, 2004	Over 120,000 student-year observations across 350 schools		X		X	X	X	Student	Teacher-based: $1,750–$7,500+	X
Muralidharan, & Sundararaman, 2006	Over 68,000 students from 300 schools	X			X	X		Student	School-based, teacher-based: Unrestricted. 500% gain in ave. student test score – 5%)	X

Continued on p. 5-46

Table 13, continued

Authors	Sample Size	Pretest Control	Other Controls: Student	Other Controls: Teacher	Other Controls: Class, School, or District	Other Controls: Family	Matching of Schools	Level of Analysis	Incentive Scheme	Math Specific Test
Richards, & Sheu, 1992	994 schools	X						School	School-based: amount varies, during study was $30–$50 per pupil or approximately $10,000 per school	
Santibáñez et al., 2007	Over 850,000 classroom-year observations in Mexico			X	X	X		Class/Teacher	(MX$)[b] 1,599 to (MX$) 12,462. Representing 27–215% of base salary	
Slotnik, Smith, Glass, & Helms, 2004	Over 100,000 student-year observations across 16 schools in CO		X		X	X	X	Student	$500–$1,500	X
Winters, Ritter, Barnett, & Greene, 2006	608 student-year observations across 5 schools in AR	X	X		X	X	X	Student	$1,800–$11,200	X

[a] Another piece, a book chapter by Clotfelter and Ladd (1996), has been cited by such authors as Podgursky and Springer (2007). However this piece has been omitted from Tables 13 and 14 because the analysis in that chapter later became integrated as part of a published article by Ladd (1999), and presents the same analysis as the Ladd (1999) piece included here.
[b] MX$ means Mexican dollars.

The methodologically strongest studies have been conducted in developing countries. For example, Muralidaran and Sundararaman (2006) reported results from a randomized trial of individual- and school-level performance-based incentives implemented across a representative sample of government-run rural primary schools in the Indian state of Andhra Pradesh. The program provided bonus payments to teachers based on the average improvement of their students' test scores in independently administered learning assessments (with a mean bonus of 3% of annual pay). The effect size for students in incentive schools was .19 standard deviation units for mathematics. The students scored significantly higher on "conceptual" as well as "mechanical" components of the test suggesting that the gains in test scores represented an actual increase in learning outcomes. Incentive schools also performed better on subjects for which there were no incentives. There was no significant difference in the effectiveness of group versus individual teacher incentives. Incentive schools performed significantly better than other randomly chosen schools that received additional paraprofessional teachers and cash block grants that were equivalent in costs to the teacher incentives.

No experiments on performance-based pay have been reported in the United States, although one large randomized trial is underway in Nashville, TN by the National Center on Performance Incentives. A recent correlational study by Figlio and Kenny (2006) examined locally generated, individual-based merit-pay programs in the United States by combining data from the NELS with the authors' own survey on the use of incentives. The performance plans varied from school to school. The authors found that merit-pay plans had positive impacts on student achievement and appeared to be effective when other types of interventions, such as more frequent teacher evaluation, were not.

2. Recommendations

The results from research on teacher incentives generally support the effectiveness of incentives, although the methodological quality of the studies in terms of causal conclusions is mixed. The substantial body of economic research in other fields indicating that salary affects the number of workers entering a field and their job performance is relevant. In the context of the totality of the evidence, and acknowledging the substantial number of unknowns, the Task Group recommends policy initiatives that put in place and carefully evaluate the effects of the following:

- Raising base salaries for teachers of mathematics to be more competitive with salaries for similarly trained non-teachers;
- Incentives for teachers of mathematics working in locations that are difficult to staff; and
- Opportunities for teachers of mathematics to increase their base salaries substantially by demonstrable effectiveness in raising student achievement.

3. Cautions

The lack of results from randomized trials of performance-pay systems in the United States and the difficulty of estimating a cost-benefit ratio for particular types of bonuses means that much work remains to be done before the nation will know enough to put particular pay-for-performance systems in place and predict their outcomes confidently. Currently, the effects of pay-based incentives on teachers are largely unknown. Knowing more about how various incentive systems affect teachers would enable the design of more effective and efficient incentives. Beyond the uncertainties about the effects of particular incentive systems, there is substantial evidence that teachers' decisions to remain in teaching and to continue teaching in particular schools are affected by work conditions in addition to salary. This includes the proximity of their residence to the school, their support from school administrators, their teaching assignment, and the characteristics of their students (Marvel, Lyter, Peltola, Strizek, & Morton, 2006; Hanushek et al., 2004). It is important to note that increasing the pay of mathematics teachers necessarily involves redirecting resources from other purposes, including investments that might have greater effects on student outcomes. Informed policy decisions need to take into consideration the relative returns of alternative investments in improving student achievement. In light of the substantial number of unknowns, policy initiatives involving teacher pay should be carefully evaluated as they are put in place.

Table 14: Reported Impacts of Models Examining the Effect of Teacher Pay for Performance on Student Achievement in Mathematics

Authors	Dependent Measure	Independent Variable	Standardized Regression Coefficients	Results of Estimated Effects	Math Specific Outcome	Grade(s)	Estimation Technique
Atkinson et al., 2004	Gains in GSCE mean math test score	Teacher's eligible for incentives	Data not available	Positive, Significance unknown	X	8–10	OLS regression
	Gains in (KS3 - GSCE) math test scores	Teacher's eligible for incentives	Data not available	Positive, Significance unknown	X	8–10	OLS regression
Cooper, Cohn, 1997	Gains in median math test score	Participation in Teacher Bonus Model	0.22	Positive & Significant	X	HS & Elem.	OLS regression
	Gains in median math test score	Participation in Campus/ Individual Incentive Model	0.13	Positive & Significant	X	HS & Elem.	OLS regression
	Gains in median math test score	Participation in Teacher Bonus Model	0.20	Positive & Significant	X	HS & Elem.	Frontier regression
	Gains in median math test score	Participation in Campus/ Individual Incentive Model	0.13	Positive & Significant	X	HS & Elem.	Frontier regression
Dee, & Keys, 2004	Stanford Achievement Test score in math	Participation in TN Career Ladder System	Data not available	Positive & Significant	X	K–3	OLS regression
Eberts, Hollenbeck, & Stone, 2002	Course completion	Implementation of performance pay incentives in MI	Data not available	Positive & Significant		HS	Difference-in difference in mean outcomes
	Student GPA	Implementation of performance pay incentives in MI	Data not available	Negative & Not significant		HS	Difference-in difference in mean outcomes
	Course-passing rates conditional on course completion	Implementation of performance pay incentives in MI	Data not available	Negative & Significant		HS	Difference-in difference in mean outcomes
Figlio, & Kenny, 2006	Sum of 12th-grade NELS:88 scores across subjects	Number of Teacher incentives offered	Data not available	Positive & Significant		12	OLS regression
	Sum of 12th-grade NELS:88 scores across subjects	The existence of large incentives for teachers	Data not available	Positive & Significant		12	OLS regression
	Sum of 12th-grade NELS:88 scores across subjects	The existence of medium incentives for teachers	Data not available	Positive & Significant		12	OLS regression
	Sum of 12th-grade NELS:88 scores across subjects	The existence of small incentives for teachers	Data not available	Positive & Significant		12	OLS regression

Continued on p. 5-49

Table 14, continued

Authors	Dependent Measure	Independent Variable	Standardized Regression Coefficients	Results of Estimated Effects	Math Specific Outcome	Grade(s)	Estimation Technique
Glewwe et al., 2003	Change in district exam scores in math	First year as incentive school	0.04	Positive & Not significant	X	4–8	GLS regression in a random effects model
	Change in district exam scores in math	Second year as incentive school	0.08	Positive & Significant	X	4–8	GLS regression in a random effects model
	Change in district exam scores in math	First year after two-year incentive program	-0.05	Negative & Not significant	X	4–8	GLS regression in a random effects model
	Change in KCPE exam scores in math	First year as incentive school	0.06	Positive & Not significant	X	4–8	GLS regression in a random effects model
	Change in KCPE exam scores in math	Second year as incentive school	0.07	Positive & Not significant	X	4–8	GLS regression in a random effects model
	Change in KCPE exam scores in math	First year after two-year incentive program	0.04	Positive & Not significant	X	4–8	GLS regression in a random effects model
Ladd, 1999[a]	Pass rates in math on the TX Assessment of Academic Skills (TAAS)	First year of incentive reform	Data not available	Positive & Significant	X	7	OLS regression
	Pass rates in math on the TX Assessment of Academic Skills (TAAS)	Second year of incentive reform	Data not available	Positive & Significant	X	7	OLS regression
	Pass rates in math on the TX Assessment of Academic Skills (TAAS)	Third year of incentive reform	Data not available	Positive & Significant	X	7	OLS regression
	Pass rates in math on the TX Assessment of Academic Skills (TAAS)	Fourth year of incentive reform	Data not available	Positive & Significant	X	7	OLS regression
Lavy, 2002	Ave. matriculation test scores in math, in secular schools	Two years after adopting incentive program	Data not available	Positive & Significant	X	12	OLS regression using regression discontinuity
	Ave. matriculation test scores in math, in religious schools	Two years after adopting incentive program	Data not available	Positive & Significant	X	12	OLS regression using regression discontinuity
	Ave. matriculation test scores in math, in secular schools	Two years after adopting incentive program	Data not available	Positive & Significant	X	12	OLS regression using RDD & PS matching
	Ave. matriculation test scores in math, in religious schools	Two years after adopting incentive program	Data not available	Positive & Significant	X	12	OLS regression using RDD & PS matching

Continued on p. 5-50

Table 14, continued

Authors	Dependent Measure	Independent Variable	Standardized Regression Coefficients	Results of Estimated Effects	Math Specific Outcome	Grade	Estimation Technique
Lavy, 2004	Awarded credits in math	Adoption of incentive program	Data not available	Positive & Significant	X	12	Regular OLS regression
	Passing rates in math on matriculation exams	Adoption of incentive program	Data not available	Positive & Significant	X	12	Regular OLS regression
	Awarded credits in math	Adoption of incentive program	Data not available	Positive & Significant	X	12	OLS regression using PS matching
	Passing rates in math on matriculation exams	Adoption of incentive program	Data not available	Positive & Significant	X	12	OLS regression using PS matching
	Awarded credits in math	Adoption of incentive program	Data not available	Positive & Significant	X	12	OLS regression using RDD
Muralidharan, & Sundararama, 2006	Difference in avg. math test designed by "educational initiatives"[b]	School assignment to individual incentives	Data not available	Positive & Significant	X	1–5	OLS regression
	Difference in avg. math test designed by "educational initiatives"	School assignment to group incentives	Data not available	Positive & Significant	X	1–5	OLS regression
Richards, & Sheu, 1992	School mean gain score on Basic Skills Assessment Program (BSAP)	Rank or "band" within established SC incentive structure	Data not available	Positive, Significance unknown		1–11	Comparison of means
	School mean gain score on Comprehensive Test of Basic Skills (CTBS)	Rank or "band" within established SC incentive structure	Data not available	Positive, Significance unknown		1–11	Comparison of means
Santibáñez et al., 2007	Carrera Magisterial test score (across many subjects)	IP Score, i.e. incentive scheme ranking metric	0.16	Positive & Significant		Primary	OLS regression using RDD
	Carrera Magisterial test score (across many subjects)	Closest to cutoff for receiving incentives	0.00	Negative & Not significant		Primary	OLS regression using RDD
	Carrera Magisterial test score (across many subjects)	IP Score, i.e. incentive scheme ranking metric	0.07	Positive & Significant		Secondary	OLS regression using RDD
	Carrera Magisterial test score (across many subjects)	Closest to cutoff for receiving incentives	0.01	Positive & Not significant		Secondary	OLS regression using RDD

Continued on p. 5-51

Table 14, continued

Authors	Dependent Measure	Independent Variable	Standardized Regression Coefficients	Results of Estimated Effects	Math Specific Outcome	Grade	Estimation Technique
Slotnik, Smith, Glass, & Helms, 2004	Iowa Test of Basic Skills (ITBS)	Marginal effect of being a pilot school (PFP plan adopted)	Data not available	Negative & Significant	X	Elem.	2 stage HLM model
	Colorado Student Assessment Program (CSAP)	Marginal effect of being a pilot school (PFP plan adopted)	Data not available	Negative & Significant	X	Elem.	2 stage HLM model
	Iowa Test of Basic Skills (ITBS)	Marginal effect of being a pilot school (PFP plan adopted)	Data not available	Positive & Not significant	X	Middle	2 stage HLM model
	Colorado Student Assessment Program (CSAP)	Marginal effect of being a pilot school (PFP plan adopted)	Data not available	Positive & Significant	X	Middle	2 stage HLM model
	Iowa Test of Basic Skills (ITBS)	Marginal effect of being a pilot school (PFP plan adopted)	Data not available	Positive & Significant	X	HS	2 stage HLM model
	Colorado Student Assessment Program (CSAP).	Marginal effect of being a pilot school (PFP plan adopted)	Data not available	Negative & Not significant	X	HS	2 stage HLM model
Winters, Ritter, Barnett, & Greene, 2006	NCE in math from SAT minus NCE in math from ITBS	Participation in incentive program	Data not available	Positive & Significant	X	K–5	OLS regression, Difference-in-difference

[a] Another piece, a book chapter by Clotfelter and Ladd (1996), has been cited by such authors as Podgursky and Springer (2007). However this piece has been omitted from Tables 13 and 14 because the analysis in that chapter later became integrated as part of a published article by Ladd (1999), and presents the same analysis as the Ladd (1999) piece included here.

[b] Educational Initiatives refers to incentives given or provided to educators based on achievement or attainment.

D. Elementary Mathematics Specialist Teachers

a) *What models exist for elementary math specialist teachers and their preparation? What is known about the qualifications and responsibilities of mathematics specialist teachers?*
b) *What evidence exists for the effectiveness of elementary math specialist teachers with respect to student achievement?*

There have been many calls for the use of math specialists at the Grade K–5 level in recent years (National Research Council, 2001; Maryland State Department of Education, 2001; National Council of Teachers of Mathematics, 2000; Horowitz, Fuchs, & Clarke, 2006; American Mathematical Society, 2001; Fennell, 2006). The pressing need of math specialists comes from two sources. On the one hand, there is now a general awareness that many elementary teachers lack adequate knowledge of mathematics for teaching. Moreover, evidence exists for substantial variability in teachers' knowledge of mathematics for teaching, and evidence exists that teachers' grasp of such knowledge is directly and very strongly related to the mathematical quality of their classroom instruction (Learning Mathematics for Teaching, 2006). On the other hand, with more than 2 million current elementary teachers, the scale problem of raising the mathematical knowledge of such a large number of teachers becomes intractable. The hope of training a small cadre of mathematically knowledgeable teachers and letting them teach the elementary mathematics classes leads some to consider the use of math specialists an important and scalable route to improving the quality of the mathematics instruction that students receive. The use of math specialists at the Grade K–5 level to reduce the number of teachers who must know mathematics well for teaching therefore seems like a sensible strategy.

However, despite multiple recommendations to use math specialist teachers, the meaning of the term "math specialist" varies. While educators and administrators who carry the title of "math specialist" can be found in most states, they frequently have different sets of responsibilities and qualifications. No national survey of their numbers, responsibilities, or qualifications has been conducted. Recent surveys of math specialists in Iowa[21] and Maryland[22] have shown that the use of math specialists is widespread, and in the case of Maryland this is true across multiple models of the position. Not long ago the Virginia legislature mandated the use of math specialists across the state; the state currently has one of the most developed and researched math specialist programs. Similar legislation may soon be proposed in the state of Maryland (Wray, 2007). Washington has also adopted legislation funding the use of math specialists to act as teacher coaches across the state (Thompson, 2007). The use of different models of math specialists is thus by no means uncommon, but as the following discussion shows, their presence in the education hierarchy has been poorly defined and their effectiveness has been insufficiently studied.

To contribute to thoughtful consideration of the issues involved in restructuring teacher roles around the idea of mathematics specialists, the Task Group reviewed a range of models in current use in the United States and abroad, and sought evidence about their effectiveness. However, there is a paucity of rigorous empirical research to answer the question posed in this section.

1. What Models Exist for Elementary Math Specialist Teachers and Their Preparation?

Math specialists can be found working at every level of U.S. public school systems. They hold positions that oversee all or groups of districts within a state, a single district, a single school within a district, classrooms within a school, and even particular students within a classroom. Some math specialists even take on several of these duties all at once (W. Haver, personal communication, April 1, 2007). In middle schools, math specialists are

[21] A survey of superintendents in Iowa shows that 63.2% of elementary schools in the state have some type of departmental model that treats math as a separate subject requiring a math specialist working under what likely is the specialized-teacher model. Not all elementary grades worked under this model. However, only 2.7% of Grades 1 and 2 were departmentalized across all elementary schools. This compares to 9% of all Grades 3, 46.8% of Grades 4, 89.2% of Grades 5, and 66.7% of Grades 6 that were departmentalized across all elementary schools (Kemis, Heiting, Spitzli, & Lang, 2003).

[22] "In a recent survey of Maryland's local mathematics supervisors, data indicated that Maryland public schools (Grades PreK–12) currently employ approximately 445 school-based (non-teaching) mathematics teacher-leadership specialists at 439 schools, 158 school-based mathematics intervention specialists (who work primarily with students), and 134 district-level mathematics curriculum/instruction specialists, statewide" (Ruehl & Wray, 2006).

employed most often specifically to teach mathematics, while in elementary schools they may teach their students about multiple subjects, not just mathematics (Fennell, 2006). Overall, the location of employment of math specialists within the structure of a state's school system partially determines both the certification needed to qualify for the position and the responsibilities of the position.

Different sets of qualifications are currently required of mathematics specialists. Minimally, a math specialist is someone who has demonstrated expertise, usually through experience as a former math teacher, testing, or both.[23] Math specialists are frequently certified through course work leading to a master's degree (W. Haver, personal communication, April, 2007). For example, particular colleges and universities in the states of Virginia[24] and Michigan[25] have graduate-level certification programs for mathematics specialists. Math specialists have also been known to gain their certification through professional development course work not leading to a higher degree from any accredited institution of higher or postsecondary education (C. Chapman, personal communication, May 1, 2007).

At present, there are at least three models for the use of math specialists: the lead-teacher or math coach model, the specialized-teacher model, and the pull-out model (Fennell, 2006). Math specialists as lead teachers are more common than those working under the other two models. In practice, however, math specialists frequently take on responsibilities that cut across all three models.

[23] For example, the governor of the state of Massachusetts recently recommended laws requiring math specialists to be elementary school teachers who have spent at least 80% of their weekly teaching time teaching math and who also have passed the elementary math section of the Massachusetts Tests for Educator Licensure (Romney, 2007).

[24] There are currently six colleges and universities in the state that offer a master's degree program that endorses mathematics specialist in accordance with states' licensure regulations for school personnel (Virginia Legislature, 2005).

[25] "In conjunction with the Horace H. Rackham School of Graduate Studies, the School of Education of the University of Michigan-Dearborn offers a Master of Arts degree in Education. This is a degree that is designed for educators who desire to fulfill all requirements for a University of Michigan master's degree, including residency, at UM-D. The program is also designed for teachers who wish to strengthen their competencies, expand their professional outlook and gain greater knowledge and understanding of their subject specialization. Through this program teachers may apply for an endorsement/certificate in Early Childhood and Early Childhood Special Education Inclusion UM-D Certificate, English as a Second Language, Middle Level Education, Middle School Mathematics Endorsement and Middle Grades Mathematics Leadership, Reading Specialist K–12, other endorsements for which the School of Education is approved, a renewal of the Provisional Certificate, or obtain a Professional Education Certificate. Eligibility for regular admission into the program includes completion of a bachelor's degree, a 3.0 (B) undergraduate grade point average or better, and a teaching certificate. Individuals whose grade point is less than 3.0 may be considered for probationary admission status and may be required to submit evidence of potential for success in a graduate program." (Retrieved on 8/13/2007 from http://www.umich.edu/~bhlumrec/acad_unit/rackham/degree_req/ www.rackham.umich.edu/Programs/other.camp/Dearborn/educ-dbn.html.)

a. The Lead Teacher or Math Coach Model

Math specialists of this type act as resource persons for their coworkers and do not directly instruct students. They work at the state, district, and school levels, providing leadership and information to teachers and staff and often coordinating mathematics programs within a school, a district, or across districts (V. Inge, personal communication, May 1, 2007). Frequently, these math specialists are elementary teachers who have been released from all or some of their responsibilities of classroom instruction (Reys & Fennell, 2003; Rowan & Campbell, 1995; Ohanian, 1998). They facilitate teachers' use of instructional strategies; align curricular frameworks with state, district, or local standards; distribute, interpret, and apply the findings of research on the teaching and learning of math to teachers; and organize professional development opportunities for teachers (M. Madden, personal communication, April 1, 2007). One of the most important responsibilities of math specialists of this type is their charge to foster a self-reflective culture of learning among teachers (Moon, 2002; P. Hess, personal communication, May 1, 2007). As Campbell (2007; Campbell & White, 1997) points out, theirs is a very challenging task as,

> [math specialists of this type] are called upon to navigate not only the complexity of teaching and student learning as it emerges in the classrooms of multiple teachers, but to do so while provoking the development of those teachers by advocating for their change, nurturing their performance, advancing their thinking, increasing their mathematical understanding, and saluting their attempts (Campbell & White, 1997, p. 328).

Some math specialists are also empowered to officially assess the performance of math teachers in the classroom, but this role is less common as it requires particular organizational structures within schools or districts to support this form of professional evaluation.[26] Math specialists of this type have been funded in part with Title I dollars.[27] In practice, the responsibilities of math specialists who work as lead teachers are quite fluid and are primarily determined by the context of their employment. They are referred to by many other names, including mentors, coaches, and resource, lead, and peer teachers (J. Lott, personal communication, May 1, 2007). As with their responsibilities, the meaning of these terms also depends on context (Miller, Moon, & Elko, 2000).

[26] For instance, the Charlottesville City Schools District in Charlottesville, VA, has hired elementary math specialists to work in individual schools who supports "the professional growth of elementary teachers by strengthening classroom teachers' understanding of math content … will co-teach lessons, develop curriculum and lessons, and create appropriate assessments, as needed. The math specialist will spend a minimum of 75% to 80% of his/her time in classrooms directly with students. He or she may also assist administrative and instructional staff in interpreting data and designing approaches to improve student achievement and instruction. Qualified applicants must hold a valid Virginia teaching certificate and be enrolled in a university-based elementary math specialist program which will result in a Virginia endorsement as a K–8 math specialist. Applicants must have teaching experience and a master's degree is preferred" (Job posting found April 2007, at: http://www.ccs.k12.va.us/uploads/Mathematics%20Specialist%20JD.pdf).

[27] For instance, the town of Carlisle, MA, has used a Title 1 grant to partially fund a math specialist position in their elementary schools since 2003 (The Carlisle School Administration, 2005).

b. The Specialized-Teacher Model

Math specialists of this type are responsible for the direct instruction of students. They work at the school and district levels, but most frequently take responsibilities in one school. In the upper grade levels (particularly for Grades 6–8) these math specialists frequently have the responsibility for the instruction of a single grade level (Fennell, 2006). They work with other teachers at their grade level to divide up the subjects being taught (e.g., math, social studies, science), and frequently all teachers retain some responsibility for reading and language arts instruction in a “homeroom” environment (Reys & Fennell, 2003).

c. The Pull-Out Model

This is a variation of the specialized-teacher model. In this model, math specialists directly instruct individuals or small groups of students within a classroom who have been identified as either failing to meet or exceeding the standards attached to their grade level (V. Mills, personal communication, May 1, 2007). Therefore, this type of math specialists does not address the problem of the deficiency of mathematics instruction in the generic elementary classroom. They have been funded in part with Title I dollars, and in some cases Title II class reduction dollars (V. Inge, personal communication, May 1, 2007).

2. What Evidence Exists for the Effectiveness of Elementary Mathematics Specialist Teachers With Respect to Student Achievement?

The search for empirical investigations about the effectiveness of elementary mathematics specialists turns up little research on the subject. In all, the searches identified 114 potentially relevant pieces of literature, but only one (McGrath & Rust, 2002) was an investigation that explored the effects of specialized mathematics teachers on student achievement in elementary schools. These authors found no difference in the mathematics gain scores of students in an elementary school with a departmentalized structure compared to students in a school with a self-contained structure.

In the series of ethnographic studies compiled as part of the Recognizing and Recording Reform in Mathematics Education Project (R^3M), the presence of mathematics specialists or leaders were cited as being critical elements of reform (Ferrini-Mundy & Johnson, 1996). But the authors caution that the presence of the math specialists in these studies is not sufficient to make the argument for mathematics specialists.

Anecdotal reports addressing the effectiveness of math specialists are more common. These include descriptions of the use of math specialists working under the lead-teacher model (Campbell, 1996; West & Straub, 2003; Miller, Moon, & Elko, 2000) and the specialized-teacher model (Sconiers, 1991). Additional work presents anecdotal evidence showing that math specialists working under the lead-teacher model have a positive effect on students’ academic achievement, or on their teachers’ beliefs about teaching and learning, or on both (Brosnan, 2007; Campbell, 2007; Inge, 2002; Ruehl & Wray, 2006; Virginia Mathematics and Science Coalition, 2002; Virginia Legislature, 2006; Wray, 2007). In addition, Wu (2007) argues from a curricular perspective that the use of math specialists in the specialized-teacher model is a necessity for adequate mathematics instruction in elementary school.

Given the paucity of evidence that general teacher certification has a positive effect on student achievement, it may seem counterintuitive to think that the use of elementary mathematics specialists would have positive effects. It is likely, however, that if the use of elementary math specialists is to have a positive effect, it will be because the training of specialists develops, in a more focused way, the specialized mathematical knowledge for teaching shown to have effects on student achievement. This suggests that policies and programs for elementary math specialist need to be developed in tandem with research that attempts to uncover those aspects of teacher knowledge and understanding most strongly related to student learning.

a. Costs Associated With Mathematics Specialists

The costs associated with the employment, training, or certification of math specialists can be substantial.

Funding of employment. One widely held belief is that any commitment made to employ a math specialist must be maintained over time, usually for at least five years (Reys & Fennell, 2003). The costs quickly escalate when officials try to do so in multiple locations across a district or state.[28] These financial costs can change dramatically depending upon the type of math specialists employed. The lead-teacher model often requires a substantial commitment in resources because teachers may have to be reassigned to take on leadership roles. On the other hand, the specialized-teacher model may not require the hiring of new personnel but rather a redistribution of teaching tasks among existing employees and therefore could be less costly than the lead-teacher model (Reys & Fennell, 2003). The pull-out model may be more expensive than the specialized-teacher model because it requires additional teachers who target their instruction only at specific and small groups of students. Additionally, the use of the pull-out model may not rule out the need for employing multiple math specialists within one school.

Required training. Math specialists who work under the lead-teacher model may need instruction in math content, teaching pedagogy, education policy, organizational leadership, adult education, and professional development practices in education. While math specialists who work under the specialized-teacher model may not need training in organizational leadership, or adult education, or both, they do require the special training in mathematics necessary to directly instruct students with special needs (W. Haver, personal communication, April 1, 2007; V. Mills, personal communication, May 1, 2007).

[28] For example, the town of Carlisle, MA, hired one part-time math specialist at the cost of $25,000 a year in 2005 (The Carlisle School Administration, 2005). If math specialists are placed in every school in any given district, the cost quickly escalates. For example, in Delaware, a budget of $2.7 million for the math specialist program would provide a specialist for every school containing seventh and eighth grades. The governor of Alabama proposed the Math, Science and Technology Program—which provides professional development, equipment, and at-school support by math and science specialists to improve math and science instruction—that would allow the program to expand to more than 600 schools and serve more than 300,000 students at a cost $33 million. While it is impossible to extrapolate any real cost estimates from these figures, they do give a sense of the scale.

Training programs vary greatly in quality and length, and although a greater emphasis on the study of mathematics is needed to train current elementary teachers to be math specialists, simply taking more college-level mathematics courses would not be sufficient. In most cases, college-level mathematics courses are generally not well designed for the benefit of Grades K–6 teachers (Battista, 1994; Reys & Fennell, 2003). Many training programs do not require the completion of a master's degree, but rely on professional development opportunities, often lasting several days or weeks at least and frequently including some college-level course work.[29] The most rigorous training programs for math specialists require the completion of a master's degree, such as that found in the state of Virginia.[30]

Implementation. There are many costs associated with the adoption of any type of math specialist position. The scarce supply of qualified candidates available to work as math specialists plays a large role in driving up the costs associated with their use (Reys & Fennell, 2003). Additionally, other common barriers to the use of math specialists in elementary schools have been cited by researchers and practitioners.

Across most types of math specialists, there is currently little general consensus as to the appropriate certification required for employment and few common job descriptions. This can be considered a cost of implementation because an agreement on the necessary training and responsibilities of a math specialist may aid in the creation of necessary research and public policy (W. Haver, personal communication, April 1, 2007; L. Pitt. personal communication, May 1, 2007).

Differences between urban, suburban, and rural schools are so far-reaching and substantial that no single definition of the qualifications and responsibilities of math specialists can be easily adopted that would be valid across the board (L. Pitt, personal communication, May 1, 2007). For example, some rural areas use on-site teacher leaders as a means of offering leadership to small but spread-out populations of teachers and some urban districts are positioning several math specialists within one school (Campbell, 2007).

[29] The Alliance for Improvement of Mathematics Skills PreK–16 (AIMS PK-16) is a partnership of nine independent school districts in south Texas and two Hispanic-serving institutions of higher education, Del Mar (community) College and Texas A&M University-Kingsville. This partnership includes course work or organized collaboration with peers or experts in the subject of mathematics but does not lead to a degree from any accredited institution of higher or postsecondary education. (Retrieved on 8/13/2007 from http://www.delmar.edu/aims/)

[30] Virginia is home to six colleges and universities that offer a master's degree program for mathematics specialists. For example, the Mathematics Education Leadership Program at George Mason University offers a 33-credit master of education leadership degree with a concentration in math specialist leader (Grades K–8). The concentration is a unique 3-year program for persons who desire part-time study to become specialists in the teaching and leadership of school mathematics. Students in the program study mathematics content, teaching, curriculum, and leadership.

b. Mathematics Specialists Internationally

Eleven nations (Singapore, Belgium (Flemish), Sweden, Japan, China, the Netherlands, Latvia, Lithuania, United Kingdom, Hungary, and the Russian Federation) scored above the United States in the fourth-grade Trends in Mathematics and Science Study (TIMSS).[31] Only Singapore, Sweden, and China use math specialists in the specialized-teacher model in Grades 1 through 6 and the other seven do not; no data could be obtained from the Russian Federation. In summary, the utilization of math specialists in countries scoring higher than the United States on standardized tests, such as the TIMSS, is not widespread.

3. Conclusions

The lack of data precludes any definitive recommendations on the use of mathematics specialists. It may be noted, nevertheless, that the specialized-teacher model of math specialists is one that comes closest to the original intent of using math specialists: It is the least expensive of the three models, and it is the one among the three that seems the most realistic in solving the scale problem of overcoming the content-knowledge deficiency among elementary teachers. But the absence of data to support its potential effectiveness presents a problem in formulating policies for its widespread adoption. It might be a surprise that only 3 of the 11 nations that outperform the United States in the fourth-grade TIMSS use mathematics specialists. This is, however, difficult to interpret because in contexts where teachers' mathematical preparation is strong, the need for a subset of teachers selected to be specialists may be reduced. Still, there are compelling reasons to encourage research to examine the effectiveness of the specialized model. In addition, the need for this kind of math specialist may be sufficiently compelling so that one may wish to proceed, with caution, to create a corps of such specialists. In terms of content knowledge, the criteria that should be used for the certification of these specialists remain unknown.

The lead-teacher model is the most expensive of the three. It is also limited by the expectation that it is possible to produce a large number of teachers who not only possess superior mathematical knowledge to mentor teachers but also who are superior in pedagogical knowledge and organizational skills. Unless there is substantial evidence that this model of specialists is effective, any pursuit of this model may be premature at this point. To the extent that the pull-out model is not designed to meet the needs of the generic classroom, this model is not pertinent to the present considerations.

[31] Major sections of the research synthesis reported here were prepared by Institute for Defense Analyses Science Technology Policy Institute**.**

BIBLIOGRAPHY

Adams, P.E., & Krockover, G.H. (1997). Concerns and perceptions of beginning secondary science and mathematics teachers. *Science Teacher Education, 81*, 29–50.

American Mathematical Society (2001). *The mathematical education of teachers*. Washington, DC: Author.

Angrist, J.D., & Lavy, V. (2001). Does teacher training affect pupil learning? Evidence from matched comparisons in Jerusalem public schools**.** *Journal of Labor Economics, 19*(2), 343–369.

Atkinson, A., Burgess, S., Croxson, B., Gregg, P., Propper, C., Slater, H., & Wilson, D. (2004). *Evaluating the impact of performance-related pay for teachers in England.* Bristol, UK: Centre for Market and Public Organization, University of Bristol.

Bacolod, M. (2007). Who teaches and where they choose to teach: College graduates of the 1990s. *Educational Evaluation and Policy Analysis*. *29*(3), 155–168.

Battista, M. (1994). Teacher beliefs and the reform movement in mathematics education. *Phi Delta Kappan, 75*(6), 462–470.

Boyd, D., Grossman, P., Lankford, H., Loeb, S., & Wyckoff, J. (2006). How changes in entry requirements alter the teacher workforce and affect student achievement. *Education Finance and Policy, 1*(2) 176–216.

Bring, J. (1994). How to standardize regression coefficients. *The American Statistician*, 48 (3): 209–213.

Brosnan, P. (2007). *Mathematics coaching influence on teaching and learning middle school mathematics*. Unpublished presentation of preliminary findings, The Ohio State University, Mathematics Coaching Program, Columbus, OH.

Campbell, P.F. (1996). Empowering children and teachers in the elementary mathematics classrooms of urban schools. *Urban Education, 30*(4), 449–475.

Campbell, P.F. (2007). *The development, activity, pedagogical practices and impact of mathematics coaches in elementary and middle schools*. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago.

Campbell, P.F., & White, D.Y. (1997). Project IMPACT: Influencing and supporting teacher change in predominantly minority schools. In E. Fennema & B. S. Nelson (Eds.), *Mathematics teachers in transition* (pp. 309–355). Mahwah, NJ: Erlbaum.

Carlisle School Administration, The. (2005). Carlisle public schools potential new initiatives, submitted to the Carlisle finance committee. Retrieved on April 1, 2007 from http://www.carlisle.mec.edu/district/business/CSABudget.pdf.

Carpenter, T.P., Fennema, E., Peterson, P., Chiang, C., & Loef, M. (1989). Using knowledge of children's mathematics thinking in classroom teaching: an experimental study. *American Educational Research Journal, 26*(4), 499–531.

Chapin, S. (1994). Implementing reform in school mathematics. *Journal of Education, 176*(1), 67–76.

Clotfelter, C.T., Glennie, E., Ladd, H.F., & Vigdor, J. (2006). *Would higher salaries keep teachers in high-poverty schools? Evidence from a policy intervention in North Carolina* (NBER, Working Paper No. 12285). Cambridge, MA: National Bureau of Economic Research.

Clotfelter, C.T., & Ladd, H. (1996). Recognizing and rewarding success in public schools. In H. F. Ladd (Ed.), *Holding schools accountable: Performance-based reform in education.* Washington, DC: The Brookings Institute.

Clotfelter, C.T., Ladd, H.F., & Vigdor, J.L. (2007). *How and why do teacher credentials matter for student achievement* (NBER, Working Paper No. 12828). Cambridge, MA: National Bureau of Economic Research.

Coleman, J.S., Campbell, E.Q., Hobson, C.J. McPartland, J., Mood, A.M., Weinfeld, F.D., et al. (1966). *Equality of educational opportunity*. Washington, DC: U.S. Department of Education.

Conant, J.B. (1963). *The education of American teachers*. New York: McGraw-Hill.

Cooper, S.T., & Cohn, E. (1997). Estimation of a frontier production function for the South Carolina educational process. *Economics of Education Review, 16*(3), 313–327.

Darling-Hammond, L. (1999). *Teacher quality and student achievement: A review of state policy evidence* (Document R-99-1). Seattle: Center for the Study of Teaching and Policy.

Darling-Hammond, L., Berry, B., & Thoreson, A. (2001). Does teacher certification matter? Evaluating the evidence. *Educational Evaluation and Policy Analysis, 23,* 57–77.

Darling-Hammond, L., Holtzman, D.J., Gatlin, S.J., & Heilig, J.V. (2005). Does teacher preparation matter? Evidence about teacher certification, Teach for America, and teacher effectiveness. *Education Policy Analysis Archives, 13*(42). Retrieved on February 13, 2008 from http://epaa.asu.edu/epaa/v13n42/.

Davis, E.A., Petish, D., & Smithey, J. (2006). Challenges new science teachers face. *Review of Educational Research, 76*(4), 607–651.

Decker, P.T., Mayer, D.P., & Glazerman, S. (2004). *The effects of Teach for America on students: Findings from a national evaluation.* Princeton, NJ: Mathematica Policy Research, Inc.

Dee, T.S., & Keys, B. (2004). Does merit pay reward good teachers? Evidence from a randomized experiment. *Journal of Policy Analysis and Management, 23*(3), 471–88.

Dolton, P., & van der Klaauw, W. (1995). Leaving teaching in the UK: A duration analysis. *The Economic Journal, 105***,** 431–444.

Eberts, R., Hollenbeck, K., & Stone, J. (2002). Teacher performance incentives and student outcomes. *Journal of Human Resources, 37*(4), 913–927.

Eisenberg, T.A. (1977). Begle revisited: Teacher knowledge and student achievement in algebra. *Journal for Research in Mathematics Education, 8*(3), 216–222.

Feiman-Nemser, S., Schwille, S., Carver, C., & Yusko, B. (1999). *A conceptual review of literature on new teacher induction*. East Lansing, MI: National Partnership for Excellence and Accountability in Teaching, Michigan State University.

Fennell, F. (2006). We need elementary school mathematics specialists now. *NCTM News Bulletin*. *43*(4).

Ferrini-Mundy, J., & Johnson, L. (1996). Chapter 7: Highlights and implications. In J. Ferrini-Mundy & T. Schram (Eds.), *Journal for Research in Mathematics Education* (Monograph 8) (pp. 111-128, 129–133).

Fetler, M. (1999). High school staff characteristics and mathematics test results. *Education Policy Analysis Archives*, *7*(9). Retrieved February 13, 2008, from http://epaa.asu.edu/epaa/v7n9.html.

Figlio, D.N., & Kenny, L.W. (2006). *Individual teacher incentives and student performance* (NBER, Working Paper No. 12627). Cambridge, MA: National Bureau of Economic Research.

Figlio, D.N., & Kenny, L.W. (2007). Individual teacher incentives and student performance. *Journal of Public Economics, 91*(5–6), 901–914.

Fulton, K., Yoon, I., & Lee, C. (2005). *Induction into learning communities*. Washington, DC: National Commission on Teaching and America's Future.

Garet, M.S., Porter, A.C., Desimone, L., Birman, B.F., & Yoon, K.S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal, 38*(4), 915–945.

Gearhart, M., Saxe, G.B., Seltzer, M., Schlackman, J., Ching, C.C., Nasir, N., et al. (1999). Opportunities to learn fractions in elementary mathematics classrooms. *Journal for Research in Mathematics Education*, *30*(3), 286–315.

Glazerman, S., Senesky, S., Seftor, N., & Johnson, A. (2006). *Design of an impact evaluation of teacher induction programs: Final report*. Princeton, NJ: Mathematica Policy Research, Inc.

Glewwe, P., Ilias, N., & Kremer, M. (2003). *Teacher incentives* (NBER, Working Paper No. 9671). Cambridge, MA: National Bureau of Economic Research.

Goldhaber, D. (2007). *Everyone's doing it, but what does teacher testing tell us about teacher effectiveness?* (Center for Analysis of Longitudinal Data in Education Research, working paper 9). Washington, DC: Urban Institute.

Goldhaber, D.D., & Brewer, D.J. (1997a). Evaluating the effect of teacher degree level on educational performance. *Developments in School Finance*, 197–210.

Goldhaber, D.D., & Brewer, D.J. (1997b). Why don't schools and teachers seem to matter? Assessing the impact of unobservables on educational productivity. *The Journal of Human Resources*, *XXXII*(3), 505–523.

Goldhaber, D.D., & Brewer, D.J. (2000). Does teacher certification matter? High school certification status and student achievement. *Educational Evaluation and Policy Analysis, 22*, 129–146.

Goldhaber, D., DeArmond, M., Lui, A., & Player, D. (2007). *Returns to skill and teacher wage premiums: What can we learn by comparing the teacher and private sector labor markets?* (working paper 8). Seattle: School Finance Redesign Project.

Gordon, R., Kane, T.J., & Staiger, D.O. (2006). *Identifying effective teachers using performance on the job*. Washington, DC: The Brookings Institution.

Greenwald, R., Hedges, L., & Laine, R.D. (1996). The effect of school resources on student achievement. *Review of Educational Research*, *66*(3), 361–396.

Gruber, K.J., Wiley, S.D., Broughman, S.P., Strizek, G.A., & Burian-Fitzgerald, M. (2002). School and staffing survey, 1999-2000: An overview of the data for public, private, public charter, and bureau of India affairs elementary and secondary schools. Washington, DC: U.S. Department of Education. Retrieved on November 1, 2007 from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2002313.

Hanushek, E. (1986). The economics of schooling—production and efficiency in public schools. *Journal of Economic Literature*, *24*(3), 1141–1178.

Hanushek, E. (1989). The impact of differential expenditures on school performance. *Educational Researcher*, *18*(4), 45–51, 62.

Hanushek, E., Kain, J.F., & Rivkin, S.J. (2004). Why public schools lose teachers. *The Journal of Human Resources*, *39*(2), 326–354.

Harbison, R.W., & Hanushek, E.A. (1992). *Educational performance of the poor: Lessons from rural northeast Brazil* (pp. 81–177). Washington, DC: World Bank.

Harris, D.N., & Sass, T.R. (2007a). *The effects of NBPTS-certified teachers on student achievement.* (National Center for Analysis of Longitudinal Data in Education Research, working paper #4). Washington, DC: Urban Institute.

Harris, D.N., Sass, T.R. (2007b). *Teacher training, teacher quality and student achievement.* (National Center for Analysis of Longitudinal Data in Education Research: working paper #3). Washington, DC: Urban Institute.

Hawkins, E.F., Stancavage, F.B., & Dossey, J.A. (1998). *School Policies and Practices Affecting Instruction in Mathematics: Findings From the National Assessment of Educational Progress.* Washington, DC: National Center for Educational Statistics.

Hedges, L.V., Laine, R., & Greenwald, R. (1994). An Exchange: Part I: Does money matter? A meta-analysis of studies of the effects of differential school input on student outcomes. *Educational Researcher*, *23*(3), 5–14.

Hill, H.C., Ball, D.L., & Schilling, S.G. (in press). Unpacking "pedagogical content knowledge": Conceptualizing and measuring teachers' topic-specific knowledge of students. *Journal for Research in Mathematics Education.*

Hill, H.C., Rowan, B., & Ball, D.L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal, 42*, 371–406.

Horowitz, S.H., Fuchs, L., & Clarke, B. (2006*). [Written comments to the national mathematics advisory panel.] 1-3*. Retrieved on September 1, 2007 from http://www.ed.gov/about/bdscomm/list/mathpanel/3rd-meeting/presentations/wendorf.james.public.pdf.

Inge, V. (2002, September 20). A Virginia forum: Moving from teacher leaders to mathematics teacher specialist: Notes from the Virginia mathematics & science coalition forum.

Ingersoll, R.M., (2000). Turnover among mathematics and science teachers in the U.S. Retrieved on October 1, 2007 from http://repository.upenn.edu/gse_pubs/96.

Ingersoll, R.M., & Kralik, J.M. (2004). *The impact of mentoring on teacher retention: What the research says*. Denver: Education Commission of the States.

Jacob, B.A., & Lefgren, L. (2002). The impact of teacher training on student achievement: Quasi-experimental evidence from school reform efforts in Chicago. *The Journal of Human Resources, 39*(1), 50–79.

Kagan, D.M. (1992). Professional growth among pre-service and beginning teachers. *Review of Educational Research, 62*(2), 129–169.

Kane, T.J., Rockoff, J.E., & Staiger, D.O. (2006). What does certification tell us about teacher effectiveness? Evidence from New York City. Retrieved on October 1, 2007 from http://www.dartmouth.edu/~dstaiger/Papers/nyc%20fellows%20march%202006.pdf.

Karges-Bone, L., Collins, F.J., & Maness, E. (2002). Last place is a fine place to begin. *Journal of Staff Development, 23*(2), 24–48.

Kemis, M., Heiting, T., Spitzli, J., & Lang, E. (2003). 2002 Iowa mathematics and science needs assessment: A study of Iowa teachers, superintendents, AEA mathematics and science coordinators, and higher education faculty. Research Institute for Studies in Education, Iowa State University: Iowa Department of Education.

King Rice, J. (2003). The incidence and impact of teacher professional development: Implications for education productivity. In M.L. Plecki, & D.H. Monk (Eds.), *School finance and teacher quality: Exploring the connections* (pp. 111–136). Larchmont, NY: Eye on Education, Inc.

Koehler, M.J., & Lehrer, R. (1998). Designing a hypermedia tool for learning about children's mathematical cognition. *Journal of Educational Computing Research, 18*(2), 123–145.

Krueger, A.B., (1999). Experimental estimates of education production functions. *Quarterly Journal of Economics. 114,* 497–532.

Laczko-Kerr, I., & Berliner, D.C. (2002). The effectiveness of "Teach for America" and other under-certified teachers on student academic achievement: A case of harmful public policy. *Education Policy Analysis Archives, 10*(37). Retrieved July 25, 2007 from http://epaa.asu.edu/epaa/v10n37/.

Ladd, H. (1999). The Dallas school accountability and incentive program: An evaluation of its impacts on student outcomes. *Economics of Education Review, 18*(1), 1–16.

Lankford, H., Loeb, S., & Wyckoff, J. (2002). Teacher sorting and the plight of urban schools: A descriptive analysis. *Educational Evaluation and Policy Analysis, 24*(1), 37–62.

Larson, J.C. (2000). The role of teacher background and preparation in students' algebra success. Rockville, MD: Office of Shared Accountability, Montgomery County Public Schools.

Lavy, V. (2002). Evaluating the effect of teachers' group performance incentives on pupil achievement. *Journal of Political Economy 110*(6), 1286–1316.

Lavy, V. (2004). *Performance pay and teachers' effort, productivity and grading ethics*. Working paper 10622. Cambridge: National Bureau for Economic Research.

Learning Mathematics for Teaching Project (2006). *Technical materials on the number and operations content knowledge measure*. Ann Arbor, MI: Author.

Levine, A. (2006). *Educating school teachers*. Washington, DC: The Education Schools Project.

Loeb, S., & Page, M. (2000). Examining the link between teacher wages and student outcomes: The importance of alternative labor market opportunities and non-pecuniary variation. *Review of Economics and Statistics, 82*(3), 393–408.

Lopez, A., Lash, A., Schaffner, M., Shields, P., & Wagner, M. (2004). *Report of research on the impact of beginning teacher induction on teacher quality and retention*. Menlo Park, CA: SRI International.

Mandeville, G.K., & Liu, Q.D. (1997). The effect of teacher Certification and task level on mathematics achievement. *Teaching and Teacher Education, 13,* 397–407.

Marvel, J., Lyter, D.M., Peltola, P., Strizek, G.A., & Morton, B.A. (2006). Teacher attrition and mobility: Results from the 2004–05 teacher follow-up survey (NCES 2007-307). Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Maryland State Department of Education (2001). *Keys to Math Success: A Report from the Maryland Mathematics Commission*. Baltimore: Author.

McDevitt, T.M., Troyer, R., Ambrosio, A.L., Heikkinen, H.W., & Warren, E. (1995). Evaluating prospective elementary teachers' understanding of science and mathematics in a model preservice program. *Journal of Research in Science Teaching, 32*(7), 749–775.

McGinnis, J.R., Parker, C., & Graeber, A.O. (2004). A cultural perspective of the induction of five reform-minded beginning mathematics and science teachers. *Journal of Research in Science Technology, 41,* 720–747.

McGrath, C., & Rust, J. (2002). Academic achievement and between-class transition time for self-contained and developmental upper-elementary classes. *Journal of Instructional Psychology, 29*(1), 40–43.

Miller, J.W., McKenna, M.C., & McKenna, B.A. (1998). A comparison of alternatively and traditionally prepared teachers. *Journal of Teacher Education, 49*(3), 165–177.

Miller B., Moon, J., & Elko, S. (2000). *Teacher leadership in mathematics and science: Casebook and facilitators guide*. Portsmouth, NH: Heinemann.

Monk, D.H. (1992). Education productivity research: An update and assessment of its role in education finance reform. *Educational Evaluation and Policy Analysis, 14*(4), 307–332.

Monk, D.H. (1994). Subject area preparation of secondary mathematics and science teachers and student achievement. *Economics of Education Review, 13,* 125–145.

Monk, D.H., & King, J.A. (1994). Multilevel teacher resource effects on pupil performance in secondary mathematics and science: The case of teacher subject-matter preparation. In R.G. Ehrenberg (Ed.), *Choices and consequences: Contemporary policy issues in education* (pp. 29–58). Ithaca, NY: ILR Press.

Moon, J. (2002). Guest editorial: Why research has a role with teachers. *American Journal of Undergraduate Research, 1*(3).

Mullens, J.E., Murnane, R.J., & Willett, J. (1996). The contribution of training and subject matter knowledge to teaching effectiveness: A multilevel analysis of longitudinal evidence from Belize. *Comparative Education Review*, *40*, 139–157.

Muralidaran, K., & Sundararaman, V. (2006). *Teacher incentives in developing countries: Experimental evidence from India*. Boston: Department of Economics, Harvard University.

National Center for Research on Teacher Learning (1991). *Findings from the teacher education and learning to teach study: Final report.* East Lansing, MI: The National Center for Research on Teacher Learning.

National Commission on Mathematics and Science Teaching for the 21st Century. (2000). *Before it's too late: A report from the National Commission on Mathematics and Science Teaching for the 21st Century*. Washington, DC: U.S. Department of Education.

National Commission on Teaching and America's Future. (2003). *No dream denied: A pledge to America's children*. Washington, DC: Author.

National Council of Teachers of Mathematics (2000). P*rinciples and standards for school mathematics*. Reston, VA: Author.

National Council of Teachers of Mathematics (2002). *Induction and mentoring of new teachers*. Retrieved July 31, 2007, from www.nctm.org/about/content.aspx?id=6362.

National Research Council, Center for Education, Division of Behavioral and Social Sciences and Education (2001). *Adding it up: Helping children learn mathematics*. J. Kilpatrick, J. Swafford, & B. Findell (Eds.). Mathematics learning study committee. Washington DC: National Academy Press.

Noell, G.H. (2006). *Annual report of value added assessment of teacher preparation*. Baton Rouge, LA: Louisiana State University.

Nye, B., Konstantopoulos, S., & Hedges, L.V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis, 26*(3), 237–257

Ohanian, S. (1998). *Tiger in the children's garden: The Exxon education foundation K–3 mathematics specialist*. Dallas: Exxon Educational Foundation.

Podgursky, M., & Springer, M. (in press). Teacher performance pay. A review. *Journal of Policy Analysis and Management.*

Raymond, M., Fletcher, S., & Luque, J. (2001). *Teach for America: An evaluation of teacher differences and student outcomes in Houston, Texas*. Stanford, CA: CREDO.

Reed, D., Rueben, K.S., & Barbour, E. (2006). *Retention of new teachers in California*. San Francisco: Public Policy Institute of California.

Reys, B.J., & Fennell, F. (2003). *Who should lead mathematics instruction at the elementary school level? A case for mathematics specialists*. Reston, VA: National Council of Teachers of Mathematics.

Richards, C.E., & Sheu, T.M. (1992). The South Carolina school incentive reward program: A policy analysis. *Economics of Education Review, 11,* 1, 71–86.

Rivkin, S.G., Hanushek, E.A., & Kain, J.F. (2005). Teachers, schools, and academic achievement. *Econometrica*, *73*(2), 417–458.

Romney, M. (2007). *The governor's budget recommendation, house 2, fiscal year 2007: Education reform #10, section 50, chapter 74B.* Retrieved on August 13, 2007 from http://www.mass.gov/bb/fy2007h1/2007outsec/.

Rowan T.E., & Campbell, P.F. (1995). *School-based mathematics specialists providing on site support for instructional reform in urban mathematics classrooms. The project impact model*. Arlington, VA: National Science Foundation. Retrieved on November 1, 2007 from http://eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/14/8a/49.pdf.

Rowan, B., Chiang, F., & Miller, R.J. (1997). Using research on employees' performance to study the effects of teachers on students' achievement. *Sociology of Education*, *70,* 256–284.

Rowan, B., Correnti, R., & Miller, R.J. (2002). What large-scale survey research tells us about teacher effects on student achievement: Insights from the *prospects* study of elementary schools. *Teachers College Record*, *104,* 1525–1567.

Ruehl, S., & Wray, J. (2006). PreK–8 and 9–12 Maryland LEA mathematics specialist survey. Ellicott City, MD: Howard County Public School System.

Sanders, W., & Rivers, J. (1996) *Cumulative and residual effects of teachers on future student academic achievement.* Knoxville, TN: University of Tennessee Value-Added Research and Assessment Center.

Santibáñez, L., Martínez, J.F., Datar, A., McEwan, P.J., Setodji, C.M., & Basurto-Dávila, R. (2007). *Breaking ground: Analysis of the assessment system and impact of Mexico's teacher incentive program "carrera magisteria."* Santa Monica, CA: RAND Education.

Saxe, G.B., Gearhart, M., & Nasir, N.S. (2001). Enhancing students' understanding of mathematics: A study of three contrasting approaches to professional support. *Journal of Mathematics Teacher Education, 4*(1), 55–79.

Sconiers, S. (1991). The UCSMP elementary mathematics specialist project. In I. Wirszup, & R. Streit (Eds.), Developments in school mathematics education around the world. *Teaching Children Mathematics, Vol. 2,* 58, 253–260.

Sheehan, D.S., Marcus, M. (1978). Teacher performance on the national teachers examination and student mathematics and vocabulary achievement. *The Journal of Educational Research*, *71*(3), 134–136.

Slotnik, W., Smith, M.D., Glass, R.J., Helms, B.J., & Ingwerson, D.W. (2004). *Catalyst for change: Pay for performance in Denver. Final report*. Denver: Community Training and Assistance Center.

Strizek, G.A., Pittsonberger, J.L., Riordan, K.E., Lyter, D.M., & Orlofsky, G.F. (2006). Characteristics of schools, districts, teachers, principals, and school libraries in the United States: 2003–04 schools and staffing survey (NCES 2006-313 Revised). Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Tatto, M.T., Nielsen, H.D., Cummings, W., Kularatha, N.G., & Dharmadasa, K.H. (1993). Comparing the effectiveness and costs of different approaches for educating primary school teachers in Sri Lanka. *Teaching and Teacher Education, 9*(1), 41–64.

Thompson, L. (2007). Math coaches help teachers help kids. *The Seattle Times*. Retrieved on February 12, 2008, from http://seattletimes.nwsource.com/html/education/2003713566_mathcoaches19m.html

Totterdell, M., Bubb, S., Woodroffe, L., & Hanrahan, K. (2004). *The impact of newly qualified teachers (NQT) induction programmes on the enhancement of teacher expertise, professional development, job satisfaction or retention rates: A systematic review of research literature on induction.* London, England: EPPI-Centre.

Van Haneghan, J.P., Pruet, S.A., & Bamberger. (2004). Mathematics reform in a minority community: Student outcomes. *Journal of Education for Students Placed at Risk*, *9*(2), 189–211.

Virginia General Assembly (2007). Summary of the amended 2006-2008 biennial budget adopted by the 2007 general assembly affecting the direct aid to public education and department of education central office budgets. Retrieved on May 10, 2007 from http://www.doe.virginia.gov/VDOE/suptsmemos/2007/reg001a.pdf.

Virginia Legislature (2005). Senate joint resolution no. 428, offered January 19, 2005. Requesting the board of education include the mathematics specialist endorsement in the licensure regulations for school personnel report. Retrieved on September 1, 2007 from http://leg1.state.va.us/cgi-bin/legp504.exe?051+ful+SJ428.

Virginia Legislature (2006). House joint resolution no. 258, offered February 2, 2006. Commending Virginia school boards that employ mathematics specialists in order to increase student achievement by increasing the quality of mathematics instruction. Retrieved September 1, 2007 from http://leg1.state.va.us/cgi-bin/legp504.exe?061+ful+HJ258.

Virginia Legislature (2007). House bill number 2093, offered January 10, 2007. A bill to amend and reenact §§ 22.1-253.13:1 through 22.1-253.13:7 and 22.1-274 of the code of Virginia, relating to the standards of quality.

Virginia Mathematics and Science Coalition, Special Issue: The role of the elementary mathematics specialists—overview. *Journal of Mathematics and Science*. (2002).

Wayne, A.J., & Youngs, P. (2003). Teacher characteristics and student achievement gains: A review. *Review of Educational Research*. *73*(1), 89–122.

West, L., & Staub, F.C. (2003). *Content-focused coaching*. Portsmouth, NH: Heinemann.

Winters, M., Ritter, G., Barnett, J., & Greene, J. (2006). *An evaluation of teacher performance pay in Arkansas.* Fayettville, AR: University of Arkansas, Department of Education Reform.

Wray, J. (2007). *PreK–8 and 9–12 mathematics leadership specialist certificate programs*. Unpublished Manuscript.

Wu, H. (2007). What should be taught in the elementary curriculum. Retrieved on May 1, 2007 from http://math.berkeley.edu/~wu/.

Chapter 6: Report of the Task Group on Instructional Practices

Russell Gersten, Co-Chair
Joan Ferrini-Mundy, Ex Officio, Co-Chair
Camilla Benbow
Douglas H. Clements
Tom Loveless
Vern Williams
Irma Arispe, Ex Officio
Marian Banfield, U.S. Department of Education Staff

CONTENTS

Figures

Tables

Abbreviations

AERA	American Educational Research Association
AGO	Adaptive Instruction and Cooperative Learning/Adaptief Groeps-Onderwijs
ANCOVA	Analysis of Covariance
APA	American Psychological Association
ATB	Active Training with Basals
ATCD	Active Teaching with Empirically Validated Curriculum Design
BASIC	Beginner’s All-purpose Symbolic Instruction Code
CAI	Computer-Assisted Instruction
CAT	California Achievement Test
CBI	Computer-Based Instruction
CBL	Computer-Based Laboratories
CGI	Cognitively Guided Instruction
CMI	Computer-managed Instruction
CP	Contextualized Problem
CRA	Concrete-representational-abstract
CTBS	California Test of Basic Skills
CTGV	Cognition and Technology Group at Vanderbilt
EAI	Enhanced Anchored Instruction
EDC	Education Development Center
ERIC	Education Resources Information Center
ETS	Educational Testing Service
FA	Formative Assessment
FIAC	Flanders Interaction Analysis Categories
FT	Project Follow Through
GSI	General Strategy Instruction
ICC	Intra-Class Correlation
IDEA	Individuals with Disabilities Act
ILS	Integrated Learning System
IP	Instructional Practices
ITBS	Iowa Test of Basic Skills
LA	Low Achieving
LD	Learning Disability
MANS	Math Applied to Novel Situations
MASTER	Mathematics Strategy Training for Educational Remediation
MAT	Metropolitan Achievement Test
NAEP	National Assessment of Educational Progress
NALT	Northwest Evaluation Association
NCLB	No Child Left Behind
NCTM	National Council of Teachers of Mathematics
NCME	National Council on Measurement in Education
NICHD	National Institute of Child Health and Human Development
NMP	National Math Panel
NRC	National Research Council

OECD	Organisation for Economic Co-operation and Development
OSEP	Office of Special Education Programs
PA	Performance Assessment
PALS	Peer-Assisted Learning
PDA	Personal Digital Assistant
PIAT	Peabody Individual Achievement Test
PISA	Programme for International Student Assessment
PMI	Peer Mediated Instruction
QED	Quasi-experimental Design
RA	Representative Abstract
RCT	Randomized Control Trials
RME	Realistic Mathematics Education
RPT	Reciprocal Peer Tutoring
SAT	Scholastic Aptitude Test
SAT-M	Scholastic Aptitude Test-Math
SAT-V	Scholastic Aptitude Test-Verbal
SBI	Schema-Broadening Instruction
SBTI	Schema-Based Transfer Instruction
SES	Socioeconomic Status
SESAT	Stanford Early School Achievement Test
SSCI	Social Sciences Citation Index
STAD	Student Teams Achievement Division
STAR	Standardized Testing and Reporting
TAI	Team Assisted Individualization
TEEM	Tucson Early Education Model
USMES	Unified Science and Mathematics for Elementary Schools
WP	Word Problems
WPS	Word Problem Solving
WRAT	Wide Range Achievement Test
WWC	What Works Clearinghouse

Executive Summary

Introduction

Mathematics teaching is an extraordinarily complex activity involving interactions among teachers, students, and the mathematics to be learned in real classrooms (Cohen, Raudenbush, & Ball, 2003). It involves making choices about material and tools to use, planning ways to group and interact with students of differing backgrounds and with differing interests and motivation. It is within this set of areas that some of today's most pressing and debated questions about mathematics instruction are situated.

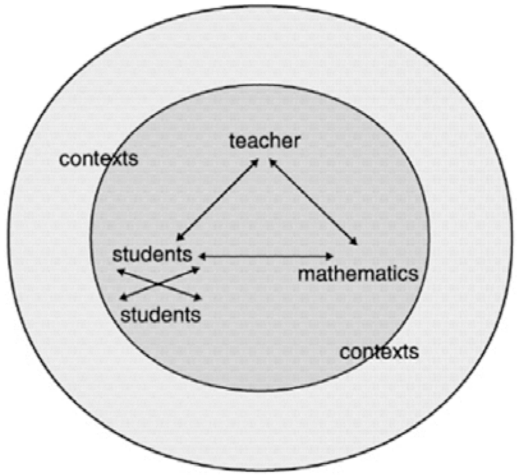
The Instructional Practices (IP) Task Group needed to consider the challenges that this complexity creates while determining what might be learned from research studies on the teaching of mathematics. Not all of the questions that teachers, policymakers, and the public wish to have answered are easily studied or lend themselves to experimental and quasi-experimental research, types of research from which generalizations to practice or for policy can be made. Moreover, many important questions that could be studied using these methods, unfortunately, have not been addressed in these ways. This limits what can validly be said about possible effective practices for the teaching of mathematics. The Task Group's undertaking was to marshal the scientific evidence to make policy recommendations and, thus, only experimental and quasi-experimental studies could be examined.

This situation is hardly unique to mathematics education, or educational research in general. It is—and has been—true in the development of scientific research in any field from engineering to economics to clinical psychology to public health. The accumulation of findings is slow at first, with the expensive experimental designs employed only after a certain amount of knowledge has emerged. Research on teaching and learning is a relatively young field.

With these caveats in mind, the overarching question the Task Group approached is: *What instructional practices enable students to learn mathematics most successfully?* Fortunately, while the knowledge base is not uniformly deep, there has been some progress at assembling evidence about questions of causal impact that has implications for practice and for policy within specific areas of mathematics instructional practice.

Therefore, within this general question, the Task Group identified six questions for investigation, addressing topics that were deemed important by the field, often including issues that have been hotly debated. Questions were identified within all three of the types of interactions comprising teaching as indicated in Figure 1; the Task Group recognizes that most of the questions here engage all three types of interactions specified in the figure, but have classified them according to the types of interactions that seem most salient.

Figure ES-1: Instructional Triangle



Source: *Adding It Up*, National Research Council, 2001, p. 314.

The Task Group realizes that by no means is the list of questions discussed below a comprehensive list of questions about each of these three types of interactions; indeed, it only begins to scratch the surface about what might be learned to inform mathematics teaching practice through research. The Task Group was aware that there are many widely used instructional practices that might have been examined here but that were not included because of limitations of time, resources, and available research. Nonetheless, it is a list of specific issues that will allow the Task Group to draw some conclusions from a small set of rigorous research studies, thereby setting the foundation for a far more expansive program of rigorous research that would fill the gaps in the research on these issues and also take up the many other issues that practitioners face in improving mathematics teaching and learning.

The methodology used in the Instructional Practices Task Group research review process, including an account of how the topics were selected, and the criteria for standards of evidence, are discussed in the full report introduction and in Appendix A.

Interactions Between Teachers and Students

Most contemporary perspectives on instruction argue that finding the best form for those interactions is a complex problem that is dependent on teachers' backgrounds, students' characteristics, school culture, the mathematical topics being addressed, and the instructional materials being used. One advantage of rigorous experimental research is that, over time, the professional community can discern which practices tend to be effective across a broad array of teacher and learning characteristics and a broad array of mathematical topics. One major goal of the Task Group's effort is to critically review the research literature for the small body of rigorous experimental studies and to discern patterns of findings that suggest specific means for improving instructional practice.

It is agreed that there is no single, ideal form in which students and teachers should consistently interact. Nonetheless, there are certain “positions” taken by various organizations and individuals arguing in favor, or in opposition to, such practices as direct instruction, cognitive-strategy instruction, student-centered approaches, cooperative learning, discovery learning, guided inquiry, situated cognition approaches, collaborative learning, and lecture-recitation.

A less polarizing issue, but one that is of great importance to classroom teachers of mathematics, is the challenge of how to best interact with low-achieving students and specifically with students having learning disabilities. A major challenge of mathematics teaching for teachers is to find the combination of instructional approaches and materials that will best meet the needs of the diversity of students in their classrooms.

Research was examined that addresses two basic questions about the forms of teacher and student interactions.

How Effective Is Teacher-Directed Instruction in Mathematics in Comparison to Student-Centered Approaches, Including Cooperative and Collaborative Groups, in Promoting Student Learning?

A controversial issue in the field of mathematics teaching and learning is whether classroom instruction should be more teacher directed or student centered. These terms have come to incorporate a wide array of meanings, with teacher directed ranging from highly scripted direct instruction approaches to interactive lecture styles, and with student centered ranging from students having primary responsibility for their own mathematics learning to highly structured cooperative groups. Schools and districts must make choices about curricular materials or instructional approaches that often seem more aligned with one instructional orientation than another. This leaves teachers wondering about when to organize their instruction one way or the other, whether certain topics are taught more effectively with one approach or another, and whether certain students benefit more from one approach than the other. The review was limited to studies that directly compared these two positions. The studies in the review compare an instructional regime in which teachers do more teaching (and therefore students less) with one in which students do more teaching and teachers less.

Only eight studies were found that met the Task Group’s standards for quality that were consistent with this definition. The studies presented a mixed and inconclusive picture of the relative impact of these two forms of instruction. High-quality research does not support the contention that instruction should be either entirely “child centered” or “teacher directed.” Research indicates that some forms of particular instructional practices can have a positive impact under specified conditions. All-encompassing recommendations that instruction should be entirely “child centered” or “teacher directed” are not supported by research. The limited research base of rigorous research does not support the exclusive use of either approach.

One of the major shifts in education over the past 25–30 years has been advocacy for the increased use of cooperative learning groups and peer-to-peer learning (e.g., structured activities for students working in pairs) in the teaching and learning of mathematics.

Cooperative learning is used for multiple purposes: for tutoring and remediation, as an occasional substitute for independent seatwork, for intricate extension or collaborative groups has been advocated in various mathematics education reports, policies, and state curricular frameworks and instructional guidelines.

Research has been conducted on a variety of cooperative learning approaches. One such approach, Team Assisted Individualization (TAI) has been shown to significantly improve students' computation skills. This instructional approach involves heterogeneous groups of students helping each other, individualized problems based on student performance on a diagnostic test, and rewards based on both group and individual performance. Effects on conceptual understanding and problem solving were not significant. There is evidence suggesting that working in dyads with a clear structure also improves computation skills in the elementary grades. However, additional research is needed.

What Instructional Strategies for Teaching Mathematics to Students With Learning Disabilities and to Low-Achieving Students Show the Most Promise?

A major challenge of mathematics teaching for teachers is to find the combination of instructional approaches and materials that will best meet the needs of the diversity of students in their classrooms. The Task Group chose to examine research that specifically looks at issues addressing students who bring a range of diversity to mathematics classrooms—those students with learning disabilities and those students who struggle with learning mathematics but who do not have a mathematics learning disability.

Obviously this topic has been of high interest for special educators, but increasingly, surveys of teachers have indicated that, as increasing numbers of students with learning disabilities (LD) receive their mathematics instruction in their regular classroom, strategies for teaching these students have become a high priority for all educators. Fortunately, there is an appreciable body of research on this topic that meets the standards for rigorous scientific research established by this Task Group.

A review of 26 high-quality studies, mostly using randomized control designs, was conducted. These studies provide a great deal of guidance concerning some defining features of effective instructional approaches for students with LD as well as low-achieving (LA) students.

Explicit systematic instruction typically entails teachers explaining and demonstrating specific strategies, and allowing students many opportunities to ask and answer questions and to think aloud about the decisions they make while solving problems. It also entails careful sequencing of problems by the teacher or through instructional materials to highlight critical features. More recent forms of explicit systematic instruction have been developed with applications for these students. These developments reflect the infusion of research findings from cognitive psychology, with particular emphasis on automaticity and enhanced problem representation.

This analysis of the body of research indicated that explicit methods of instruction are consistently and significantly effective with students with learning disabilities in the performance of computations, solving word problems, and solving problems that require the application of mathematics to novel situations.

Only a small number of studies were located that investigated the use of visual representations or student "think alouds." Therefore, no inferences about their effectiveness can be drawn. The research suggests that they are most useful when they are integrated with explicit instruction.

Based on this admittedly small body of research, the Task Group concludes that students with learning disabilities and other students with learning problems should receive some time on a regular basis with some explicit systematic instruction. There is no reason to believe that this type of instruction should comprise all the mathematics instruction these students receive. However, it does seem essential for building proficiency in both computation and the translation of word problems into appropriate mathematical equations and solutions. Some of this time should be dedicated to ensuring that students possess the foundational skills and conceptual knowledge necessary for understanding the mathematics they are learning at their grade level.

Interactions Between Students and the Mathematics They Are Learning

In discussions about effective mathematics instruction, there are multiple questions about the ways the curriculum, instructional materials, and resources for mathematics learning influence student performance in mathematics. The Task Group chose to focus the research review on three controversial areas of this domain: a curricular issue concerning how the mathematics is presented; an issue about the impact of tools as a means of interacting with the mathematics; and a curricular organization issue about the pace and nature of the mathematics for gifted students.

Do 'Real-World' Problem Approaches to Mathematics Teaching, and Efforts to Ensure that Students Can Solve 'Real-World' Problems, Lead to Better Mathematics Performance Than Other Approaches?

The importance of addressing this topic as an especially controversial "hot button" issue in the field was stressed, both by Task Group members, as well as by members of the public in testimony to the Panel. Many textbooks begin each unit with "real-world" problems and consider this a potentially motivating approach. Some instructional materials use "real-world" contexts as a means of introducing mathematical ideas. State and national standards typically include as goals students' ability to apply mathematics to situations that occur in a child's life, or that might occur in future jobs. Consequently, high-stakes assessments such as the National Assessment of Educational Progress (NAEP) and many state tests include "real-world" problems. There are strong perspectives both in support of, and in opposition to, the use of "real-world" problems as a means for students to interact with the mathematics they are to learn. For these reasons, a serious examination of the research on this topic seemed warranted.

The research review focused on two key issues. The first was the extent to which problems that authors call "real-world" problems do, in fact, pique students' interest and engage them more fully in exploration of mathematical concepts, with a goal of learning mathematics. A related issue is the extent to which use of "real-world" problems in instruction increases students' ability to transfer the mathematical knowledge they possess to novel situations. Unfortunately, there is no agreed upon definition of "real-world" problems; the terminology is used in very different ways by researchers, teachers, mathematicians, and mathematics educators. And, the matter that what is a "real-world" problem to one student may not be a "real-world" problem to another is an issue. Conducting research in this area is complex; fidelity of the teachers' implementation of the instructional materials or instructional strategy is difficult to assess. Although not addressed in the studies we examined, teachers' knowledge and capacity to use such problems effectively varies greatly. Given these caveats, the Task Group addressed the question of whether using "real-world" contexts to introduce and teach mathematical topics and procedures is preferable to more typical instructional approaches.

The body of high-quality studies for this topic is small. Five studies addressed the question of whether the use of "real-world" problems as the instructional approach led to improved performance on outcome measures of ability to solve "real-world" problems, as well as on more traditional assessments. Four of these were similar enough to combine in a meta-analysis. The meta-analysis revealed that if mathematical ideas are taught using "real-world" contexts, then students' performance on assessments involving similar problems is improved. However, performance on assessments of other aspects of mathematics learning, such as computation, simple word problems, and equation solving, is not improved.

For certain populations (upper elementary and middle grade students and remedial ninth-graders) and for specific domains of mathematics (fraction computation, basic equation solving, and function representation), instruction that features the use of "real-world" contexts can have a positive impact on certain types of problem solving. Additional research is needed to explore the use of "real-world" problems in other mathematical domains, at other grade levels, and with varied definitions of "real-world" problems.

What Is the Relative Impact on Mathematics Learning When Students Use Technology Compared to Instruction That Does Not Use Technology?

There are several types of educational technology that provide opportunities for students to interact with mathematics. The review includes focus on computer software, calculators, and graphing calculators.

Among the many categories of technology, calculators, including graphing calculators, have generated the greatest amount of debate. Some have championed their use in developing problem-solving abilities, by allowing students to perform far more, and more complex, arithmetic operations than would have been possible without technology. Others believe that calculators may reinforce independent skill mastery, or even that they should, along with mental arithmetic, replace some of the paper-and-pencil calculations that dominate elementary school mathematics. On the other hand, some have bemoaned their misuse. One concern is that calculators may have an insidious effect on paper-and-pencil arithmetic and algebraic skills. Some are concerned that reliance on calculators can preclude the development of proficiency with standard calculation algorithms and thus deprive students of an understanding and appreciation of the mathematics that underlies the standard algorithms, as well as ability to quickly retrieve basic arithmetic facts.

A review of 11 studies that met the Task Group's rigorous criteria (only one study was less than 20 years old) found limited to no impact of calculators on calculation skills, problem solving, or conceptual development over periods of up to one year. Unfortunately, these studies cannot be used to judge the advantages or disadvantages of multiyear calculator use beginning in the early years because such long-term use has not been adequately investigated. The Task Group cautions that to the degree that calculators impede the development of automaticity, fluency in computation will be adversely affected.

The Task Group found that computer-assisted instruction (CAI) drill and practice, if of high quality, can improve students' performance compared to conventional instruction. Drill and practice programs **can be** useful tools in developing students' automaticity, or fast, accurate, and effortless performance on computation, freeing working memory so that attention can be directed to the more complicated aspects of complex tasks.

Research has demonstrated that tutorials (CAI programs, often combined with drill and practice) that are well designed and implemented can have a positive impact on mathematics performance. CAI tutorials have been used effectively to introduce and teach new subject-matter content. However, these studies also suggest several important caveats. Care must be taken that there is evidence that the software to be used has been shown to increase learning in the specific domain and with students who are similar to those who are under consideration. Educators should critically inspect individual software packages and studies that evaluate them critically. Furthermore, support conditions to use the software effectively (sufficient hardware and software; technical support; adequate professional development, planning, and curriculum integration), should be in place, especially in large-scale implementations, to achieve optimal results.

Research indicates that computer programming improves students' performance compared to conventional instruction on both mathematics achievement in general and on problem solving. However, computer programming by students can be employed in a wide variety of situations using distinct pedagogies, not all of which may be effective (e.g., integration into the mathematics curriculum may be required for substantial effects). Therefore, the findings are limited to the careful, targeted application of computer programming for learning used in the studies reviewed.

What Instructional Arrangements for Engaging with Mathematics Are Most Promising for Mathematically Gifted Students?

Zimmer, Christina, Hamilton, and Weber Prine (2006) noted that, in a recent survey of teachers implementing the No Child Left Behind Act (NCLB), over half the teachers surveyed felt that implementation of the law resulted in improved learning opportunities for low-performing students, but that teachers and administrators at all levels of schooling worried about high-achieving students receiving adequate instructional challenge in all curricular areas. This review of the research literature explored the immediate and delayed impacts of gifted education approaches aimed at accelerating students' mathematics instruction (e.g., by covering two, or even four years of high school mathematics in 15 months) and those that attempt to provide enrichment or extension activities for mathematically precocious students. This question is addressed in the category of student-mathematics interactions because it is very much about the pace and structure for engaging gifted students with mathematics content.

The Task Group's review of the literature about the kind of mathematics instruction would be most effective for gifted students focused on the impact of programs involving acceleration, enrichment, and the use of homogeneous grouping. The extensive literature searches we conducted yielded few studies that met the Task Group's methodologically rigorous criteria for inclusion. Thus for this topic—and this topic only—we relaxed these criteria in order to fulfill our charge of evaluating the "best available scientific evidence." One randomized control trial study and seven quasi-experimental studies were located. All but one of these studies have limitations.

Despite the flaws in any one study, the set of studies suggests there is value to differentiating the mathematics curriculum for students who are gifted in mathematics and possess sufficient motivation, especially when acceleration is a component (i.e., pace and level of instruction are adjusted). A small number of studies suggest that individualized instruction, where the pace of learning is increased and often managed via computer instruction, produces gains in learning.

Gifted students who are accelerated by other means not only gained time and reached educational milestones earlier (e.g., college entrance) but appear to achieve at levels at least comparable to those of their equally able same-age peers on a variety of indicators even though they were younger when demonstrating their performance on the various achievement benchmarks. One study suggests that gifted students also appear to become more strongly engaged in science, technology, engineering, or mathematical areas of study.

Some support also was found for supplemental enrichment programs. Of the two programs analyzed, one explicitly utilized acceleration as a program component and the other did not. This supports the view in the field of gifted education that acceleration and enrichment combined should be the intervention of choice. We believe it is important for school policies to support appropriately challenging work in mathematics for gifted and talented students.

Interactions Between Teachers and Mathematics

Teachers engage with the mathematical content that they teach in various aspects of teaching practice: in planning and designing lessons, in interpreting and responding to student questions, and in the work of assessing their students' mathematical knowledge. Fortunately, formative assessment is an area of great contemporary interest and is also an area with a rich set of rigorous experimental field studies.

What Is the Impact of Use of Formative Assessment in Mathematics Teaching?

Educators at all levels realize the importance of assessing their students' progress during the year. Formative assessment—the ongoing monitoring of student learning to inform instruction—is generally considered a hallmark of effective instruction in any discipline. Interest in formative assessment has dramatically increased since No Child Left Behind required states to establish accountability systems. Teachers' interpretation and use of the data available to them from instructionally embedded, in-class assessments in the context of teaching, along with high-stakes assessments are critical for improving outcomes for all students. However, many different systems have been established and touted for use as formative assessments. These range from the end-of-unit and mastery tests that accompany major commercial textbook series, to more contingent and informal probes of students' understandings to be used while they solve problems, to weekly tests that sample from the year's instructional objectives in mathematics. The Task Group examined rigorous experimental studies of the impact of teachers' use of formative assessment on students' growth in mathematics proficiency. The Task Group's review of the high-quality studies of this topic produced several conclusions.

Teachers' regular use of formative assessment is marginally significant in improving their students' learning. This is especially true if teachers have additional guidance on using the assessment to design and individualize instruction.

Although the research base is smaller, and less consistent than that on the general effectiveness of formative assessment, the research suggests that several specific tools and strategies can help teachers use formative assessment information more effectively. The first promising strategy is providing formative assessment information to teachers (via technology) on content and concepts that require additional work with the whole class. The second promising strategy involves using technology to specify activities needed by individual students. Both of these aids can be implemented via tutoring, computer-assisted instruction, or help provided by a professional (teacher, mathematics specialist, trained paraprofessional).

The Task Group cautions that only one type of formative assessment has been studied with rigorous experimentation. These are assessments that include random sampling of items that address state standards. These assessments tend to take between 2 and 8 minutes to administer and thus are practical for regular use.

The regular use of formative assessment particularly for students in the elementary grades is recommended. These assessments need to provide information not only on their content validity but also on their reliability and their criterion-related validity (i.e., correlation of these measures with other measures of mathematics proficiency). For struggling students, frequent (e.g., weekly or biweekly) use of these assessments appears optimal, so that instruction can be adapted based on student progress.

Research is needed regarding the content and criterion-related validity and reliability of other types of formative assessments (such as unit mastery tests included with many published mathematics programs, performance assessments, and dynamic assessments involving “think alouds”). This research should include studies of consequential validity (i.e., the impact they have on helping teachers improve the effectiveness of their instruction).

Use of formative assessments in mathematics can lead to increased precision in how instructional time is used in class and can assist teachers in identifying specific instructional needs. Formative measures provide guidance as to the specific topics needed for assistance. Formative assessment should be an integral component of instructional practice in mathematics.

Conclusion

Mathematics instruction is a complex professional practice. The educational research community has made important forays into several of the most controversial and pressing questions about the effectiveness and impact of various types of instructional practice, and in particular have conducted some studies that examine the effects of various interpretations and implementations of practices that have been advocated in the “reform” documents in mathematics education over the past two decades.

The question asked by the Task Group is: *What can be learned from a review of the best available evidence in six important aspects of practice?* These practices included: the use of “real-world” problems in mathematics teaching, the use of technology, the enrichment and acceleration of instruction for mathematically precocious students, the use of cooperative groups and peer instruction, the use of direct instruction with learning disabled students, and the use of formative assessment.

For none of the areas examined did the Task Group find sufficiently strong and comprehensive bodies of research to support all-inclusive policy recommendations of any of the practices addressed. Nor did the Task Group find sufficient evidence to support policy recommendations favoring the status quo in mathematics teaching.

Across all of the areas, the Task Group found that **several instructional practices in mathematics teaching show some promise, in comparison to typical practice, for affecting student learning.** In each case the “promising” practice is clearly specified, somewhat prescriptive, and involves a mix, or combination, of particular distinct practices. Thus, for example, it cannot be said that cooperative learning is a practice whose effectiveness is supported by research—but the Team Assisted Individualization (TAI) approach, with particular students in a particular area of mathematics does appear to be effective. Although formative assessment to inform instruction is useful, it is enhanced when teachers use assessment tools with known validity and reliability. For students performing in the lower third of grade level expectations, explicit instruction using clear models of proficient performance, many opportunities to verbalize their problem-solving strategies, and adequate practice and review should be a part of the mathematics program. It is not surprising that what the Task Group found about effective instructional practice is far more subtle and nuanced than direct answers to the starkly stated questions investigated.

The Task Group found some rather robust findings, but these findings must be accompanied by a caveat. When a practice is demonstrated by high-quality experimental research to have some promise, it is critical to be clear about the promise "for what aspects of mathematics proficiency." Different practices and approaches impact different kinds of outcomes, ranging from computational performance, to "real-world" problem solving, to identifying extraneous problem information, to long-term participation and interest in studying mathematics.

Because researchers and practitioners use different definitions to describe their interventions, it is conceptually problematic to place too much stock in generalizing that a broad category of practice (e.g., using technology or using "real-world" problems) has impact because a set of studies working on the same particular component of this category has impact, which was the case in some of the Task Group's reviews.

The Task Group's process included asking mathematicians and mathematics education reviewers to examine the mathematical content of the research studies—to look at the assessments and interventions, to the extent possible, based on the published reports. They expressed important concerns, including the possibility that an outcome measure item purported to measure computation might not do so because it really measures ability to use the context, for instance. They expressed concern that some topics were underdeveloped (i.e., failed to help students access the underlying mathematics in the topic covered), or that items were mislabeled (e.g., as "problem solving") when the mathematics expert might classify them otherwise. However, they also did note that several of the studies reviewed seemed to help students increase their knowledge of mathematics and how to apply that knowledge to novel situations in a way that is valid from a mathematical perspective.

Seeing how few robust findings emanated from a review of the rigorous research on the topics addressed, it is clear that most practitioners would like more guidance for several areas of instruction. Yet even the inconclusive and limited findings can provide a real service to the profession. If an administrator, a developer or a parent comments, "Research says that lessons must start with 'real-world' problems," or "Students will really learn mathematics only if they are taught using direct instruction," consumers and professionals now know that research is inconclusive on these topics.

This is a necessary step in the evolution of educational research into a more mature science. The paucity of findings and the paucity of high-quality experimental research in the field led the Task Group to realize, early on in the process, that few definitive answers to the research questions posed would be found.

What Would the Instructional Practices Task Group Say to the Practitioner?

There is no one ideal approach to teaching mathematics; the students, the mathematical goals, the teacher's background and strengths, and the instructional context, all matter. The findings here do suggest that it is especially important to:

- monitor what students understand and are able to do mathematically;
- design instruction that responds to students' strengths and weaknesses based on research when it is available; and
- employ instructional approaches and tools that are best suited to the mathematical goals, recognizing that a deliberate and conscious mix of strategies will be needed.

Also, it is important for teachers, school administrators, and the public to understand the importance of helping to formulate research questions and being willing to participate in the types of experimental and quasi-experimental studies that are described here.

What Would the Instructional Practices Task Group Say to the Researcher?

More research that can identify causal claims is needed to guide both policy and practice. Building the mathematics education research portfolio to include this work will involve:

- Formulation of research questions that are of interest to practitioners and policy-makers;
- Collaborations among mathematicians, mathematics education researchers, methodologists, and psychometricians; and
- Motivation to design and undertake rigorous studies.

The work of this Task Group has substantiated understanding of the complexity and challenge of effective mathematics instruction. It is now up to practitioners, policymakers, mathematicians, and mathematics education researchers to take up the challenges of clarifying the definitions of mathematics instructional practices, debunking myths about mathematics instruction, and formulating the types of research studies that can answer the pressing questions that need to be addressed.

In conclusion, instructional practice should be informed by high-quality research, when available, and by the best professional judgment and experience of accomplished classroom teachers.

I. Introduction

A. Instructional Practices

Mathematics teaching is an extraordinarily complex activity involving interactions among teachers, students, and the mathematics to be learned in real classrooms (Cohen, Raudenbush, & Ball, 2003). It involves making choices about material and tools to use, planning ways to group and interact with students of differing backgrounds and with differing interests and motivation. It is within this set of areas that some of today's most pressing and debated questions about mathematics instruction are situated.

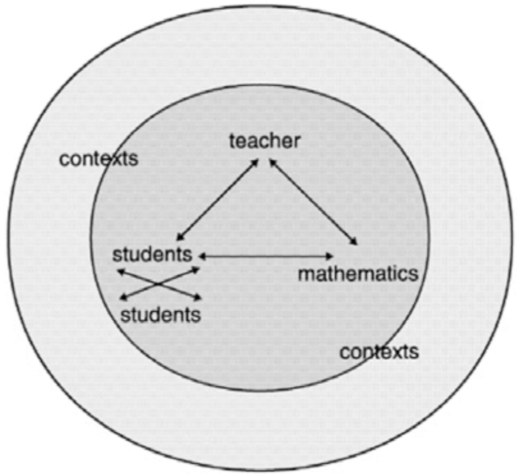
The Instructional Practices Task Group needed to consider the challenges that this complexity creates while determining what might be learned from research studies on the teaching of mathematics. Not all of the questions that teachers, policymakers, and the public wish to have answered are easily studied or lend themselves to experimental and quasi-experimental research, types of research from which generalizations to practice or for policy can be made. Moreover, many important questions that could be studied using these methods, unfortunately, have not been addressed in these ways. This limits what can validly be said about possible effective practices for the teaching of mathematics. The Task Group's undertaking was to marshal the scientific evidence to make policy recommendations and, thus, only experimental and quasi-experimental studies were examined.

This situation is hardly unique to mathematics education or educational research in general. It is—and has been—true in the development of scientific research in any field from engineering to economics to clinical psychology to public health. The accumulation of findings is slow at first, with the expensive experimental designs employed only after a certain amount of knowledge has emerged. Research on teaching and learning is a relatively young field.

With these caveats in mind, the overarching question the Task Group approached is: *What instructional practices enable students to learn mathematics most successfully?* Fortunately, while the knowledge base is not uniformly deep, there has been some progress at assembling evidence about questions of causal impact that has implications for practice and for policy within specific areas of mathematics instructional practice.

Therefore, within this general question, the Task Group identified six questions for investigation, addressing topics that were deemed important by the field often including issues that have been hotly debated. Questions were identified within all three of the types of interactions comprising teaching as indicated in Figure 1; the Task Group recognizes that most of its questions here engage all three types of interactions specified in the figure, but have classified them according to the types of interactions that seem most salient.

Figure 1: Instructional Triangle



Source: *Adding It Up,* National Research Council, 2001, p. 314.

The Task Group realizes that by no means is the list of questions discussed below a comprehensive list of questions about each of these three types of interactions; indeed, it only begins to scratch the surface about what might be learned to inform mathematics teaching practice through research. The Task Group was aware that there are many widely used instructional practices that might have been examined here but that were not included because of limitations of time, resources, and available research. Nonetheless, it is a list of specific issues that will allow the Task Group to draw some conclusions from a small set of rigorous research studies, thereby setting the foundation for a far more expansive program of rigorous research that would fill the gaps in the research on these issues and also take up the many other issues that practitioners face in improving mathematics teaching and learning.

1. Notes About Methodology and Reporting

The methodology used in the Instructional Practices Task Group research review process, including an account of how the topics were selected, and the criteria for standards of evidence, are included in Appendix A. For ease in reading this report key points are summarized here. The studies used in the meta-analyses and syntheses that follow were designated as either Category 1 or 2. Category 1 studies are experimental and quasi-experimental studies that meet or meet with reservations the What Works Clearinghouse (WWC) standards. Studies in this category provide evidence of causal claims and include randomized control trials (RCTs) and strong quasi-experimental studies. Some exceptions to the WWC criteria were allowed; these are described in Appendix A. Category 2 consisted of weak group comparison studies (e.g., failed RCTs and weak nonequivalent comparison designs; other flaws discussed in Appendix A). Category 2 studies are always open to multiple interpretations with regard to causal inferences; however, they are not necessarily weak studies for other purposes such as description. If there were no acceptable experimental studies, sections of the report may include brief discussion of Category 2 studies. If there is a pattern of findings across the studies this may also be mentioned. Panelists were free to use any type of research (descriptive, correlational, qualitative) to set the context for the meta-analyses.

For all studies that met the criteria for inclusion, the What Works Clearinghouse guidelines were used to calculate standardized mean differences in mathematics achievement. Hedges' g standardized mean differences were calculated for each of the studies. In cases in which schools, teachers, or classrooms were assigned (either randomly or nonrandomly) into intervention and comparison groups and the unit of assignment was not the same as the unit of analysis, the effect size and accompanying standard error were adjusted for clustering within schools, teachers, or classrooms. When judged appropriate, effect sizes were pooled across studies meta-analytically using random effects models. Specifically, weighted mean effect sizes were computed using inverse variance weights to reflect the statistical precision of the respective studies stemming from both the subject-level and study-level sampling error.

Multiple contrasts: For each study that included at least three conditions, effect sizes were calculated for all relevant contrasts, provided that they were orthogonal. When pooling the effects using meta-analytic techniques, only independent effect sizes per study were included, i.e., those not based on the same participant samples.

Multiple outcomes: For studies that reported effects on more than one mathematics achievement outcome, either one outcome was chosen, or the results from multiple outcomes were averaged, with decisions made by the authors on a case-by-case basis. Assessments that were overly aligned with an intervention were either not used or noted when used.

Multiple independent samples within a study: In cases in which impacts on independent samples within a study were reported, all independent effect sizes were included separately in the pooled analysis.

Throughout this report, effect sizes are reported as statistically significant only when $p < .01$. Effect sizes where $p < .10$ are described as "bordering on significance". This report conforms with the National Math Advisory Panel (Panel) *Guidelines for Standards of Evidence* in using the following terminology: strong evidence, moderately strong evidence, suggestive evidence, inconsistent evidence, and weak evidence.

2. Interactions Between Teachers and Students

Most contemporary perspectives on instruction argue that finding the best form for those interactions is a complex problem that is dependent on teachers' backgrounds, students' characteristics, school culture, the mathematical topics being addressed, and the instructional materials being used. One advantage of rigorous experimental research is that, over time, the professional community can discern which practices tend to be effective across a broad array of teacher and learning characteristics and a broad array of mathematical topics. One major goal of the Task Group's effort was to critically review the research literature for the small body of rigorous experimental studies and to discern patterns of findings that suggest specific means for improving instructional practice.

It is agreed that there is no single, ideal form in which students and teachers should consistently interact. Nonetheless, there are certain "positions" taken by various organizations and individuals arguing in favor of, or in opposition to, such practices as direct instruction, cognitive-strategy instruction, student-centered approaches, cooperative learning, discovery learning, guided inquiry, situated cognition approaches, collaborative learning, and lecture-recitation.

A less polarizing issue, but one that is of great importance to classroom teachers of mathematics, is the challenge of how to best interact with low-achieving students and specifically with students having learning disabilities. A major challenge of mathematics teaching for teachers is to find the combination of instructional approaches and materials that will best meet the needs of the diversity of students in their classrooms. Research was examined that addresses two basic questions about the forms of teacher and student interactions.

a. How Effective Is Teacher-Directed Instruction in Mathematics in Comparison to Student-Centered Approaches, Including Cooperative and Collaborative Groups, in Promoting Student Learning?

A controversial issue in the field of mathematics teaching and learning is whether classroom instruction should be more teacher-directed or student-centered. These terms have come to incorporate a wide array of meanings, with teacher-directed ranging from highly scripted direct instruction approaches to interactive lecture styles, and with student-centered ranging from students having primary responsibility for their own mathematics learning to highly structured cooperative groups. Schools and districts must make choices about curricular materials or instructional approaches that often seem more aligned with one instructional orientation than another. This leaves teachers wondering about when to organize their instruction one way or the other, whether certain topics are taught more effectively with one approach or another, and whether certain students benefit more from one approach than the other. The review was limited to studies that directly compared these two positions. The studies in the review compare an instructional regime in which teachers do more teaching (and therefore students less) with one in which students do more teaching and teachers less.

One of the major shifts in education over the past 25–30 years has been advocacy for the increased use of cooperative learning groups and peer-to-peer learning (e.g., structured activities for students working in pairs) in the teaching and learning of mathematics. Cooperative learning is used for multiple purposes: for tutoring and remediation, as an occasional substitute for independent seatwork, for intricate extension activities, for initial brainstorming and for numerous other purposes. Use of cooperative or collaborative groups has been advocated in various mathematics education reports, policies, and state curricular frameworks and instructional guidelines.

Provided in a subsequent section of the report is a synthesis of the research that met Task Group criteria on the topic of teacher-directed vs. student-centered learning. The section includes a review of studies that compare general versions of teacher-directed and student-centered mathematics instruction in accordance with the Task Group’s definition. There are only a limited number of sufficiently rigorous research studies making this comparison, within this definition. There is also a review of studies that examine various forms of cooperative and collaborative groups, including such specific approaches as Team Assisted Instruction and Peer Assisted Learning, as well as the use of cooperative groups with technology, and other approaches.

b. What Instructional Strategies for Teaching Mathematics to Students with Learning Disabilities and to Low-Achieving Students Show the Most Promise?

A major challenge of mathematics teaching for teachers is to find the combination of instructional approaches and materials that will best meet the needs of the diversity of students in their classrooms. The Task Group chose to examine research that specifically looks at issues addressing students who bring a range of diversity to mathematics classrooms—those students with learning disabilities (LD) and those students who struggle with learning mathematics but who do not have a mathematics learning disability.

Obviously this topic has been of high interest for special educators, but increasingly, surveys of teachers have indicated that, as increasing numbers of students with LD receive their mathematics instruction in their regular classroom, strategies for teaching these students has become a high priority for all educators. Fortunately, there is an appreciable body of research on this topic that meets the standards for rigorous scientific research established by this Task Group.

3. Interactions Between Students and the Mathematics They Are Learning

In discussions about effective mathematics instruction, there are multiple questions about the ways the curriculum, instructional materials, and resources for mathematics learning influence student performance in mathematics. The Task Group chose to focus the research review on three controversial areas of this domain: a curricular issue concerning how the mathematics is presented; an issue about the impact of tools as a means of interacting with the mathematics; and a curricular organization issue about the pace and nature of the mathematics for gifted students.

a. Do 'Real-World' Problem Approaches to Mathematics Teaching and Efforts to Ensure That Students Can Solve 'Real-World' Problems Lead to Better Mathematics Performance Than Other Approaches?

The importance of addressing this topic as an especially controversial "hot button" issue in the field was stressed, in particular, by Task Group members, as well as by members of the public testifying before the Panel. Many textbooks begin each unit with "real-world" problems and consider this a potentially motivating approach. Some instructional materials use "real-world" problems as a means of introducing mathematical ideas. State and national standards typically include as goals students' ability to apply mathematics to situations that occur in a child's life or that might occur in future jobs. Consequently high-stakes assessments such as the National Assessment of Educational Progress (NAEP) and many state tests include "real-world" problems. There are strong perspectives both in support of, and in opposition to, the use of "real-world" problems as a means for students to interact with the mathematics they are to learn. For these reasons, a serious examination of the research on this topic seemed warranted.

The research review focused on two key issues. The first was the extent to which problems that authors call “real-world” problems do, in fact, pique students’ interest and engage them more fully in exploring mathematical concepts with a goal of learning mathematics. A related issue is the extent to which “real-world” problems increase students’ ability to transfer the mathematical knowledge they possess to novel situations. Unfortunately, there is no agreed upon definition of “real-world” problems; the terminology is used in very different ways by researchers, teachers, mathematicians, and mathematics educators.

b. What Is the Relative Impact on Mathematics Learning When Students Use Technology Compared to Instruction that Does Not Use Technology?

There are several types of educational technology that provide opportunities for students to interact with mathematics. The review includes focus on computer software and calculators, including graphing calculators.

Among the many categories of technology, calculators, including graphing calculators, have generated the greatest amount of debate. Some have championed their use in developing problem-solving ability by allowing students to perform far more, and more complex, arithmetic operations than would have been possible without technology. Others believe that calculators may reinforce independent skill mastery, or even that they should, along with mental arithmetic, replace some of the paper-and-pencil calculations that dominate elementary school mathematics. On the other hand, some have bemoaned their misuse. One concern is that calculators may have an insidious effect on paper-and-pencil arithmetic and algebraic skills. Some are concerned that reliance on calculators can preclude the development of proficiency with standard calculation algorithms and thus deprive students of an understanding and appreciation of the mathematics that underlies the standard algorithms, as well as ability to quickly retrieve basic arithmetic facts.

c. What Instructional Arrangements for Engaging with Mathematics Are Most Promising for Mathematically Gifted Students?

Zimmer, Christina, Hamilton, and Weber Prine (2006) noted that, in a recent survey of teachers implementing the No Child Left Behind Act, more than half the teachers surveyed felt that implementation of the law resulted in improved learning opportunities for low-performing students but that teachers and administrators at all levels of schooling worried about high-achieving students receiving adequate instructional challenge in all curricular areas. This review of the research literature explored the immediate and delayed impacts of gifted education approaches aimed at accelerating students’ mathematics instruction (e.g., by covering 2, or even 4 years of high school mathematics in 15 months) and those that attempt to provide enrichment or extension activities for mathematically precocious students. This question is addressed in the category of student-mathematics interactions because it is very much about the pace and structure for engaging gifted students with mathematics content.

4. Interactions Between Teachers and Mathematics

Teachers engage with the mathematical content that they teach in various aspects of teaching practice: in planning and designing lessons, in interpreting and responding to student questions, and in the work of assessing their students' mathematical knowledge. Fortunately, formative assessment is an area of great contemporary interest and is also an area with a rich set of rigorous experimental field studies.

a. What Is the Impact of Use of Formative Assessment in Mathematics Teaching?

Educators at all levels realize the importance of assessing their students' progress during the year (i.e., formative assessment). Interest in formative assessment has dramatically increased since No Child Left Behind required states to establish accountability systems. Teachers' interpretation and use of the data available to them from instructionally embedded, in-class assessment in the context of teaching, along with high-stakes assessments are critical for improving outcomes for all students. However, many different systems have been established and touted for use as formative assessments. These range from the end-of-unit and mastery tests that accompany major commercial textbook series, to more contingent and informal probes of students' understandings to be used while they solve problems, to weekly tests that sample from the year's instructional objectives in mathematics. The Task Group examined rigorous experimental studies of the impact of teachers' use of formative assessment on students' growth in mathematics proficiency.

BIBLIOGRAPHY

Anderson, J.R., Reder, L.M., & Simon, H.A. (1996). Situated learning and education. *Educational Researcher, 25*(4), 5–11.

Cohen, D.K., Raudenbush, S.W., & Ball, D.L. (2003). Resources, instruction, and research. *Educational Evaluation and Policy Analysis, 25(2),* 119–142.

National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (NIH Publication No. 00-4769). Washington, DC: Author.

National Research Council. (2001). Adding it up: Helping children learn mathematics. In J. Kilpatrick, J. Swafford, & B. Findell (Eds.), *Mathematics learning study committee, center for education, division of behavioral and social sciences and education*. Washington, DC: National Academy Press.

National Research Council. (2002). Scientific research in education. In R.J. Shavelson and l. Towne (Eds.), *Committee on scientific principles for educational research*. Washington, DC, Center for Education: National Academy Press.

National Research Council. (2004). *On evaluating curricular effectiveness: Judging the quality of K–12 mathematics evaluations.* Committee for a Review of the Evaluation Data on the Effectiveness of NSF-Supported and Commercially Generated Mathematics Curriculum Materials. Washington, DC: The National Academies Press.

Woodward, J., & Montague, M. (2002). Meeting the challenge of mathematics reform for students with learning disabilities. *Journal of Special Education, 36*(2), 89.

Zimmer, R., Christina, R., Hamilton, L. S., & Weber Prine, D. (2006). *Evaluation of two out-of-school programs in Pittsburgh public schools: No Child Left Behind's supplemental educational services and state of Pennsylvania's educational assistance program*. (No. WR-451-PPS). Santa Monica, CA: RAND.

II. Teacher-Directed and Student-Centered Instruction in Mathematics

The Task Group, in its initial activity of formulating key flashpoint questions about mathematics instruction, identified the question, "Is teacher-directed instruction more effective than student-centered instruction?" as one needing particular attention because of pressures faced by teachers being urged to use one or the other of these styles exclusively.

The terms "teacher-directed instruction" and "student-centered instruction" are sometimes used as labels to stake out starkly contrasting views in discussions about mathematics teaching. In their purest forms, these labels convey images of instruction that are in some sense polar opposites. For some, these terms convey differences of perspective about whether the goals of teachers or the needs of students should have primacy in determining what specific mathematics teaching interventions will be used in the mathematics classroom. Some have interpreted "student-centered" instruction to mean that students, rather than teachers, control the direction and content of the mathematical discussion, or that students are expected to somehow learn all mathematics on their own, by teaching one another. "Teacher directed" instruction has been interpreted in similarly extreme ways, to mean that teachers are not responsive to, or aware of, students' learning issues, and instead dispense mathematics instruction in a way that is disconnected from the learners. The distinction has been summarized by some with the ubiquitous "sage on the stage rather than a guide on the side" maxim. The idea of the "guide on the side" is often associated with intentions of mathematics education reforms in the past two decades concerning the role of the teacher.

The National Research Council (NRC) report *Adding It Up* acknowledges the challenge of such labels in discussing teaching: "Much debate centers on forms and approaches to teaching: 'direct instruction' versus 'inquiry,' 'teacher-centered' versus 'student-centered,' 'traditional' versus 'reform.' These labels make rhetorical distinctions that often miss the point regarding the quality of instruction. The quality of instruction is a function of teachers' knowledge and use of mathematical content, teachers' attention to and handling of students, and students' engagement in and use of mathematical tasks." (2001, p. 315). This section undertakes a circumscribed treatment of what have perhaps become positions that have hardened into ideologies that rarely offer pragmatic guidance to teachers on how they should teach.

A. Literature Review

The caricatures of teacher-directed and student-centered instruction that have sometimes emerged in debates on this subject are not validated in the versions of teacher-directed and student-centered instruction that were examined in the studies reviewed. Indeed, teacher-directed instruction involves assessment and careful attention to student progress—students were very much involved in the versions of teacher-directed instruction described in these studies. And, teachers have a key role in the versions of student-centered instruction described here as well—they choose tasks, direct discussion, and work toward mathematical goals. The Task Group found no examples of studies in which students were teaching themselves or each other without any teacher guidance; nor did the Task Group find studies in which teachers conveyed mathematics content directly to students without any attention to their understanding or response. The fact that these terms, in practice, are neither clearly nor uniformly defined, nor are they true opposites, complicates the challenge of providing a review and synthesis of the literature.

The literature review presented below will not end the debate over student-centered and teacher-centered instruction. Instead, it offers a summary of what is currently known about effective instructional practices in mathematics as they relate to teacher-directed or student-centered approaches, drawing on an exhaustive search based on terms that have been used in the literature to describe both teacher-directed and student-centered instructional approaches. Using the search terms provided in Appendix B, only studies of how instruction influences mathematics achievement were included. Studies of how instructional approaches affect students' motivation, social skills, attitudes toward mathematics, or other noncognitive outcomes were not reviewed. The search found 40 randomized experiments or quasi-experiments that were determined to have a rigorous design and to be relevant to the topic.

Blanket statements endorsing a philosophy of mathematics education will not be found. Even when examining high-quality studies, considering context is crucial to properly interpreting results. In other words, some approaches may be shown to be effective, but confidence in their effectiveness is only warranted under specified conditions. Factors such as the age of students, the mathematical content that is taught, the duration of the instructional program, the preparation of the teachers, and the outcomes that are sought must be taken into account.

Consequently, this literature review comes with a warning. Educators should be leery of sweeping claims that "best practices" in mathematics instruction are known and supported by research. Most efforts to promote any single all-encompassing style of instruction, to the exclusion of any others, are based on beliefs, not science, and much of the research cited to promulgate those beliefs does not meet minimal standards of quality. A body of high-quality research simply does not exist to answer such broad questions as whether teacher-directed or student-centered instruction should be dominant in teaching mathematics.

1. What Is Meant by Teacher-Directed Instructional Strategies?

Interpretations of teacher-directed instructional strategies gleaned from the literature vary widely. Common to most is the notion that the teacher has complete control of the instruction. Perhaps the best-known instantiation of teacher-directed instructional strategies as conceptualized in the late 1960s and 1970s was in the context of Project Follow Through, and was called Direct Instruction (Gersten & Carnine, 1984). Project Follow Through, a part of President Lyndon Johnson's War on Poverty in 1967, has been reported to be the largest and most expensive federally funded experiment in education ever conducted (Becker, 1977; Gersten et al., 1984). There were 17 distinct instructional models represented in the Follow Through evaluation (Stebbins, St. Pierre, Proper, Anderson, & Cerva, 1977, p. 2; see also Stallings, 1975, and Stallings & Kaskowitz, 1974), and Direct Instruction was one of these models.

Direct Instruction was a behaviorally oriented educational program using a tightly controlled teaching methodology and highly structured instructional materials. The instruction was programmed, emphasizing children's learning of intelligent behavior through programmed questions and answers provided in a fast-paced fashion. "Teachers present specified questions.… Proper responses are reinforced and incorrect answers are corrected according to specified procedures" (Stebbins et al., 1977, p. 65).

According to Kameenui, Carnine, Darch, and Stein (1986), the model of Direct Instruction, as described by Gersten and Carnine (1984) "employs clearly articulated teaching sequences that contain explicit, step-by-step teacher modeling and a means of assessing student mastery at each step of development" (p. 635). Meyer, Gersten, and Gutkin (1983) describe the component of the "Direct Instruction Model:" "a) consistent focus on academic objectives; b) high allocations of time to small-group instruction in reading, language, and math; c) the tight, carefully sequenced Distar curriculum; … e) a comprehensive system for monitoring both the rate at which students progress through the curriculum and their mastery of the material covered" (p. 243).

In work of the same period, Good and colleagues described and studied what they termed "active mathematics teaching" (see Good & Grouws, 1977, 1979; Good, Grouws, & Ebmeier, 1983). Guidelines for instruction in this program indicate a highly structured and prescribed instructional sequence, including: daily review, development, seatwork, as well as homework assignments and special reviews. The development sequence includes explanations, demonstrations, and illustrations, as well as repetition and elaboration (Brophy & Good, 1986, p. 348). Kameenui et al. (1986) provide details about what the development component of active mathematics teaching involves: "The direct approach to development views the teacher as one who controls the instructional goals and pace, chooses the appropriate materials, and provides immediate and academically oriented feedback to the learner." And, in this same vein, work by Slavin (1980) has examined "focused instruction," which involves a "highly structured schedule of teaching, worksheet work, and quizzes." (As described in Beady, Slavin, & Fennessey, 1981, p. 519).

More recently, reform documents of the past two decades have argued against teacher-directed instruction, not the same very specific, programmed kind of direct instruction of the 1960s and 1970s but rather a more general type of instruction in which the teacher is the primary authority. For instance, the National Council of Teachers of Mathematics *Curriculum and Evaluation Standards for School Mathematics* notes: "In many classrooms, learning is conceived of as a process in which students passively absorb information, storing it in easily retrievable fragments as a result of repeated practice and reinforcement" (National Council of Teachers of Mathematics, 1989, p. 10). It is contrasted with an instructional style that emphasizes a "constructive, active view of the learning processes" (p. 10)—which is not exactly aligned with the student-centered view of the 1960s and 1970s.

In summary, the hallmarks of teacher-directed instructional strategies include clearly prescribed instructional sequences, consistent focus on content objectives, emphasis on explanation, assessment and correction of errors, feedback to students and assignments and review, in which the teacher is doing all of these things. In addition, teacher-directed instruction can be manifested in the way the classroom is organized, and is often associated with whole group instruction. Most important is that the teacher is doing the teaching.

For the purpose of this review, "teacher-directed" instruction is viewed as instruction in which primarily the teacher is communicating the mathematics to the students directly and in which the majority of interactions about the mathematics are between the teacher and the student.

2. What Is Encompassed in "Student-Centered" Approaches to Mathematics Instruction?

More than a century ago John Dewey urged educators to consider the notion of what some have called "child-centered" education: "The teacher is not in the school to impose certain ideas or to form certain habits in the child, but is there as a member of the community to select the influences which shall affect the child and to assist him in properly responding to these influences" (Dewey, 1897, pp. 77–80).

The emphasis on the centrality of the student in education has been interpreted in various ways in mathematics education over several decades, most recently in the standards reform movement of the late 1980s and subsequent extensions. Common to most of these interpretations is the notion that the students' experience, motivation, interest, and knowledge needs to be a central consideration in the teachers' design and implementation of instruction. In addition, the focus on teachers' relationships with students is sometimes central. In recent years, numerous policies and programs have promoted a student-centered emphasis, invoking various theories of learning. Constructivism is one of these theories. See Cobb (2007) for a discussion of the ways in which theoretical and philosophical perspectives influence mathematics education.

At least three of the Project Follow Through instructional strategies can be classified as student-centered. The Tucson Early Education Model (TEEM) was "based on the concept that each child has a unique growth pattern with individual rates and styles of learning" (Stebbins et al., 1977, p. 41). TEEM took as a premise that formal learning should have as its basis the experiences young children bring to the classroom. "Some classroom activities are selected and structured by the teacher, and others are chosen by the children" (p. 41). The second model, the Cognitively Oriented Curriculum model, was a Piagetian developmental model focused on developing children's ability to reason. The goals included helping children sustain independent activity, define and solve problems, assume responsibility for decisions and actions, and work cooperatively (p. 2, p. 89). And, the Education Development Center (EDC) Open Education approach, with its roots in the philosophy of the British Infant Schools and the developmental theories of Piaget, provided children with a wide range of materials for learning.

Another clearly defined approach to student-centered instruction was developed by Flanders and his colleagues, based on the Flanders Interaction Analysis Categories (FIAC) (see Flanders, 1970; discussed in Brophy and Good, 1986). According to Brophy and Good, "Flanders believed that there was too much teacher talk and not enough student talk in most classrooms, so that teachers should be more 'indirect'" (p. 333). This style of instruction involved examining pupil attitudes and emphasized "asking questions, accepting and clarifying ideas or feelings, and praising or encouraging as indirect techniques" (p. 333). The student-centered interventions of this time period, often aimed at primary and early elementary age children, featured elements of free choice and developmental readiness.

Lampert (1990) has compared school mathematics with knowledge in the discipline of mathematics, noting that “few teachers engage students in a public analysis of the assumptions that they make to get their answers” (p. 32). She summarizes the assumptions of reform documents (National Research Council, 1989; National Council of Teachers of Mathematics, 1989) as follows: “Reform documents recommend that mathematics students should be making conjectures, abstracting mathematical properties, explaining their reasoning, validating their assertions, and discussing and questioning their own thinking and the thinking of others” (p. 33). This might be viewed as a description of a “student-centered” approach, although of course such approaches could be in place in a teacher-directed classroom as well. Socratic teaching methods, for example, feature teacher-directed dialogues between teachers and students. Close questioning requires students to justify thinking out loud and to explain the logic behind their arguments and conclusions.

In the National Research Council report *How People Learn* (National Research Council, 2000), the term “learner centered” is used to:

> *…refer to environments that pay careful attention to the knowledge, skills, attitudes, and beliefs that learners bring to the educational setting. The term includes teaching practices that have been ‘culturally responsive,’ ‘culturally appropriate,’ ‘culturally compatible,’ and ‘culturally relevant’ (Ladson-Billings, 1995). The term also fits with ‘diagnostic teaching’ (Bell et al., 1980): attempting to discover what the student is thinking in relation to the problems on hand, discussing their misconceptions sensitively, and giving them situations to go on thinking about which will enable them to readjust their ideas (Bell, 1982). Teachers who are learner centered recognize building on the conceptual and cultural knowledge that students bring with them to the classroom (pp. 133–134).*

To be sure, some depictions of student-centered instruction emphasize a passive role for teachers. The Bureau of Labor Statistics, for example, in its *Occupational Outlook Handbook* describes the job of teaching as follows:

> *Teachers act as facilitators or coaches, using classroom presentations or individual instruction to help students learn and apply concepts in subjects such as science, mathematics, or English. They plan, evaluate, and assign lessons; prepare, administer, and grade tests; listen to oral presentations; and maintain classroom discipline. Teachers observe and evaluate a student’s performance and potential and increasingly are asked to use new assessment methods. For example, teachers may examine a portfolio of a student’s artwork or writing in order to judge the student’s overall progress. They then can provide additional assistance in areas in which a student needs help. Teachers also grade papers, prepare report cards, and meet with parents and school staff to discuss a student’s academic progress or personal problems.*

> *Many teachers use a 'hands-on' approach that uses 'props' or 'manipulatives' to help children understand abstract concepts, solve problems, and develop critical thought processes. For example, they teach the concepts of numbers or of addition and subtraction by playing board games. As the children get older, teachers use more sophisticated materials, such as science apparatus, cameras, or computers. They also encourage collaboration in solving problems by having students work in groups to discuss and solve problems together. To be prepared for success later in life, students must be able to interact with others, adapt to new technology, and think through problems logically (U.S. Department of Labor, Bureau of Labor Statistics, 2008).*

In summary, the elements of student-centered mathematics instruction as described in contemporary treatments include emphasis on student responsibility and independence; acknowledgment of students' experiences, prior knowledge, and interests and motivations in the design of mathematics instruction; and the centrality of students' thinking and students teaching other students in the classroom. Teachers facilitate, encourage, and coach but do not explicitly instruct by showing and explaining how things work.

For the purposes of this review, "student-centered" instruction is viewed as instruction in which primarily students are doing the teaching of the mathematics and that the majority of the interactions about the mathematics occurs between and among students.

The vague and often overlapping ways in which "teacher-directed" and "student-centered" are used in the literature, not to mention in contemporary discourse, present challenges for any attempt to summarize research on the topic. A major source of the ambiguity stems from the use of these adjectives to modify several different nouns. As illustrated in the citations above, by their very nature nouns such as "education," "environments," "practices," or "learning" comprise a collection of activities. The Task Group chose to focus on one element—instruction—and to search for studies that contrast who is doing the teaching—teachers or students? The contrast never exists in an absolute sense, of course, but in degree. All of the studies in our review compare an instructional regime in which teachers do more teaching (and therefore students less) with one in which students do more teaching and teachers less.

This focus was chosen because teachers told the Panel that they understand the expectations of administrators in their districts are that they teach exclusively in teacher-directed ways, essentially as it has been defined here. And, other teachers have said that their administrators are critical unless they are teaching in student-centered ways, again as it has been defined here. Thus, this review was undertaken to highlight these distinctions in ways that will hopefully help policymakers and teachers to engage in practice that is evidence based.

In accordance with the definitions of teacher directed and student centered being used, the focus is on the nature of the mathematics **instruction** (literally the interactions between the teachers and the students about mathematics). Not included are studies in which the nature of the **curriculum** (the materials for learning) might be construed as more or less teacher-directed or student-centered. Note that most current interpretations of what it means to be teacher-directed or student-centered conflate issues of instruction and curriculum. For example, the use of practice worksheets (a curricular device) might be associated with a teacher-directed approach but indeed could be highly student-centered in its design.

Within the review, a number of studies were found that directly compare a form of teacher-directed instruction to a form of student-centered instruction. These studies are discussed in the first section. Later sections address studies that have looked at student-centered classroom organizational approaches of cooperative groups and peer-tutoring approaches.

Methodological considerations specific to this section can be found in Appendix A.

3. Comparisons of Student-Centered and Teacher-Directed Approaches to Instruction

Research Studies. Eight studies meeting the criteria to be considered Category 1 studies were located that compared student-centered and teacher-directed approaches to instruction (see Table 1). The pattern of effects is quite complex. It is not possible to undertake a meta-analysis of these studies because the interventions are all of such distinct types, according to the above categorization, that pooling effect sizes is not meaningful. The specific interventions studied in the Project Follow Through evaluation study (Stebbins et al., 1977) are treated in a separate section.

Table 1: Studies That Investigated the Effects of Teacher-Directed and Student-Centered Instruction on Mathematics Achievement

Study	Design	Sample	Duration/Content	Contrast	Measure	Subgroup	Hedge's g	Standard error	
Brenner et al., 1997[a]	RCT	128 students in six intact pre-algebra classes in three junior high schools in southern California	20 days/ Meaningful thematic contexts used in pre-algebra concepts	Guided discovery approach vs. Traditional textbook	Pooled problem solving outcomes: word problem solving (ES = 0.110), function word problem (ES = 0.393), and equation solving (ES= -0.281) measures	Overall	0.074	(ns)	0.399
					Pooled representation outcomes: function word problems (ES = 0.877*) and word problems (ES = 0.623)	Overall	0.750	~	0.403
Ciccelli, 1982	RCT	64 fifth-grade students	40 minutes per day for 9 days/ Probability and graphing	Direct vs. Nondirect instruction	Math achievement test	Low ability	-0.614	(ns)	0.569
						Medium ability	0.156	(ns)	0.311
						High ability	-0.374	(ns)	0.535
Fuchs et al., 2006[c]	RCT	445 third-grade students in 30 classrooms in seven schools in an urban district	16 weeks/ Mathematical problem solving strategies	Schema broadening instruction vs. Control	Pooled transfer measures	Overall	0.545	(ns)	0.439
				Schema broadening instruction-real life vs. Control	Pooled transfer measures	Overall	1.077	*	0.464
Hopkins et al., 1997	Quasi	34 third-grade and 40 fifth-grade students	1 30-minute session/ Arithmetic	Didactic vs. Constructivist approach	Arithmetic computation test	Boys	0.155	(ns)	0.345
						Girls	1.142	***	0.327
Kameenui et al., 1986 – Study 3	RCT	24 fourth-grade students	11 daily 35-minute sessions/ Division	Direct Instruction (Project Follow Through) vs. Control	Math achievement test	Overall	0.444	(ns)	0.399
Muthukrishna & Borkowski, 1995	RCT	54 third-grade students	14 consecutive class days/ Addition and subtraction word problems	Guided discovery approach vs. Direct strategy instruction	Pooled near transfer outcomes (classifications: ES = -0.006, sequence: ES = -0.346, comparison: ES = -0.383)	Overall	-0.245	(ns)	0.276
					Pooled far transfer outcomes (form: ES =0.576*, context: ES =0.380)	Overall	0.478	~	0.278

Continued on p. 6-19

Table 1, continued

Study	Design	Sample	Duration/Content	Contrast	Measure	Subgroup	Hedge's g		Standard error
Rittle-Johnson, 2006	RCT	85 third- through fifth-grade students in an urban parochial school	1 40-minute session/ Mathematical equivalence	Discovery learning and prompts to explain vs. Direct instruction and prompts to explain	Procedural learning test	Overall	-0.272	(ns)	0.301
					Procedural transfer test	Overall	-0.125	(ns)	0.300
					Conceptual knowledge test	Overall	-0.458	(ns)	0.304
				Discovery learning and no prompts vs. Direct instruction and no prompts	Procedural learning test	Overall	-0.719	*	0.313
					Procedural transfer test	Overall	-0.085	(ns)	0.303
					Conceptual knowledge test	Overall	0.050	(ns)	0.303
Rudnitsky et al., 1995[a]	RCT	401 third- and fourth-grade students in 21 classrooms in six schools	18 days/ Addition and subtraction word problems	Writing and discussion vs. Practice and explicit heuristics	Near transfer posttest	3rd-grade females	-0.073	(ns)	0.428
						3rd-grade males	0.392	(ns)	0.429
						4th-grade females	-0.138	(ns)	0.431
						4th-grade males	0.462	(ns)	0.417
Project Follow Through Evaluation									
Stebbins et al., 1977—Direct Instruction Model[b]	Quasi	316 Project Follow Through and 317 Non-Project Follow Through students enrolled in program from kindergarten through third grade in five districts (New York, NY; Grand Rapids, MI; W. Iron Co., MI; Flint, MI; and Providence, RI)	Kindergarten through 3rd grade/ General elementary school mathematics curriculum	Direct Instruction Follow Through vs. Non-Follow Through	Overall Metropolitan Achievement Test (MAT) outcome: computations (ES = 0.315*), concepts (ES = -0.064), and problem solving (ES = 0.017) measures	Overall	0.105	(ns)	0.142
Stebbins et al., 1977—Cognitive Curriculum Model[b]	Quasi	177 Project Follow Through and 337 Non-Project Follow Through students enrolled in program from kindergarten through third grade in five districts (New York, NY; Okaloosa Co., FL; Greeley, CO; Seattle, WA; and Chicago, IL)	Kindergarten through 3rd grade/ General elementary school mathematics curriculum	Cognitively Oriented Curriculum Follow Through vs. Non-Follow Through	Overall Metropolitan Achievement Test (MAT) outcome: computations (ES = -0.318~), concepts (ES = -0.355*), and problem solving (ES = -0.295~) measures	Overall	-0.357	*	0.167
Stebbins et al., 1977—EDC Open Education Model[b]	Quasi	248 Project Follow Through and 487 Non-Project Follow Through students enrolled in program from kindergarten through third grade in five districts (Philadelphia, PA; Burlington, VT; Lackawanna Co., PA; Morgan Comm. Sch., DC; and Paterson, NJ)	Kindergarten through 3rd grade/ General elementary school mathematics curriculum	EDC Open Education Follow Through vs. Non-Follow Through	Overall Metropolitan Achievement Test (MAT) outcome: computations (ES = 0.052), concepts (ES = -0.081), and problem solving (ES = -0.073) measures	Overall	-0.037	(ns)	0.140

~ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$
[a] Data were adjusted for clustering that occurred within classrooms.
[b] Data were adjusted for clustering that occurred within schools.
[c] These studies use classroom-level analyses.

The studies that produced significant effect sizes in contrasts comparing teacher-directed to student-centered instruction will be discussed. Hopkins, McGillicuddy-De Lisi, and De Lisi (1997) investigated different ways of teaching third- and fifth-grade students how to compute with whole numbers and fractions. The experiment consisted of a single 30-minute session in which students were taught individually. The researchers were interested in determining if didactic or constructivist instruction, in this case instruction emphasizing particular mathematical practices, benefits girls and boys differentially when learning mathematical computation. Two groups of children—matched on a pretest of computation skills, grade, and gender—were formed at third and fifth grades. In a single 30-minute session, conducted with individual students, students were taught how to solve six computation problems involving addition, subtraction, multiplication, and division with whole numbers and fractions. Whole number operations included multiplying three digit by three digit numbers and the items with fractions included addition and subtraction of mixed numbers without a common denominator. One group was instructed using a didactic approach in which the mathematical practices of algorithms, rules, and solution methods were explicitly taught. A constructivist teaching strategy was used with the other group. In that setting, teachers suggested alternative mathematical practices, including ways to organize tasks, recasting children's comments with tag questions requesting clarification, and providing demonstrations that guided students to discover solutions. Both groups were then post-tested on computation. No effect was found for boys, but girls in the didactic instructional groups made statistically significant gains over girls who received constructivist teaching (ES = 1.142). The gains for girls were apparent at both grade levels.

The literature search uncovered two high-quality studies that found evidence of far transfer, under very limited conditions. Both studies investigated how to teach problem solving strategies, and both studies found guided discovery (a particular interpretation of student-centered instruction) more effective than direct instruction methods. Muthukrishna and Borkowski (1995) conducted an experiment consisting of 14 lessons teaching third graders the number family (or part-whole) strategy for solving word problems with addition and subtraction. The strategy involves a part-whole schema in which one larger quantity (known or unknown) is thought of as a whole comprising two smaller quantities (known or unknown) that are parts. For example, a student who knows that 1 + 5 = 6 can conceptualize 6 as a whole made up of 1 and 5 as parts, with two subtraction facts, 6 - 1 = 5 and 6 - 5 = 1, derived from the addition fact. Students were taught that the unknown in a word problem involving addition and subtraction is either a whole or a part. Students were randomly assigned to four conditions for instruction: direct strategy teaching, guided discovery, a combination of direct teaching and guided discovery, and a control condition. The number family schema was not taught to the control group, and students in the control condition primarily received instruction from their regular classroom teacher. The other groups were instructed by the experimenter and an assistant. A typical guided discovery lesson consisted of 20 minutes working in pairs followed by whole class discussion of students' solutions. Students in the guided discovery group worked with a variety of manipulative materials and did not engage in individual paper and pencil activities. The post-test consisted of addition and subtraction word problems assessing both near and far transfer of skills.

Additional aspects of the study may be of use in interpreting the results. Although a 14-day period is relatively brief for studying instructional methods, it is a long time to teach a single problem solving strategy to third-graders. Note also that the skill emphasized here—using the number family strategy—was used to solve mathematics problems that are typically far below the grade level of the students in the study. Students who had mastered solving the problem types used in the study were excluded, as were students who lacked the basic skills to solve addition and subtraction word problems not requiring regrouping or carrying.

Calculation of effect sizes on the near transfer outcomes, contrasting the discovery group and the direct instruction group, revealed no significant effect sizes. On the far transfer problems, the guided discovery group outperformed the direct strategy teaching group on the test for form transfer (problems of a different form than those in instruction) with a significant effect size (ES = 0.576). There was no significant effect on the test for context (problems presented in a context different from what had been used in instruction) when contrasting the same two groups. Nonetheless, the pooled effect size on the far transfer measures approached significance (ES = 0.478), favoring the guided discovery group.

A study by a team of researchers at the University of California at Santa Barbara (Brenner et al., 1997) investigated middle school pre-algebra students who were learning how to represent function problems in multiple formats. Problem representation involves constructing and using mathematical representations in words, graphs, tables, and equations, a difficult task particularly for many students making the transition from arithmetic to algebra. In pre-algebra textbooks, function word problems are typically represented by equations, tables consisting of ordered pairs, and graphs.

The intervention consisted of a 20-day instructional unit taught to 128 seventh- and eighth-graders in three schools. Three teachers each had two pre-algebra classes; one class was randomly assigned to the treatment and the other to the control condition. Students in the treatment groups worked in heterogeneous groups and used manipulative materials. Teachers used a guided discovery approach in which students were encouraged "to explore different representations and to develop their own understanding of each one" (Brenner et al., p. 668). In the control groups, students were taught with textbooks and teachers used traditional direct instructional methods.

Results were mixed. Five outcomes were tested in the study: word problem solving, function word problems, equation solving, function representation, and word problem representation. One effect size reached statistical significance. Students in the student-centered strategies groups were better able, at a significant level, to represent function problems in multiple ways (ES = 0.877). The subtest of the assessment instrument measuring this outcome awarded two points per item. For example, students were given the problem: "Mary Wong just got a job working as a clerk in a candy store. She already has $42. She will earn $7 per hour. How many hours will she have to work to have a total of $126?" (Brenner et al., p. 671). For the subtest for representation, students received one point for drawing a diagram, chart, table, or graph to represent the problem and one for writing the correct equation in the form of $y = mx + b$. Having the correct answer had no bearing on the score for the representation subtest. The effect size on the word problem representation outcome measure (ES = 0.623) did not reach statistical significance, and the pooled far transfer outcomes (using the two representation outcome measures) bordered on statistical significance (ES = 0.750). On the subtests reflecting ability to find problem solutions correctly, effect sizes were not significant, but favored the guided discovery group with effect size 0.393 on function word problem outcome, and the traditional textbook group (ES = -0.281) on equation solving measures.

In a 2006 paper, Rittle-Johnson reported on a comparison of teacher-directed instruction and student-centered instruction. In the comparison of the discovery learning and direct instruction conditions without prompts, there was a significant effect size favoring the direct instruction condition on the procedural learning test.

Her formulation of teacher-directed is based in information-processing and cognitive theories about working memory capacity, while her view of student-centered instruction draws on Piaget and current reformers who emphasize the importance of discovery learning. In this study of third through fifth grade students, children were assigned randomly to one of four conditions, based on two factors: “direct instruction versus discovery learning” and prompts for self-explanation vs. no prompts. Self-explanation is “generating explanations for oneself” (Rittle-Johnson, 2006, p. 1). The mathematical focus was on equivalence—the ideas that equations represent balance and that the same quantity is on both sides. Rittle-Johnson points out the importance of this idea as a precursor to algebra.

Assessments involved a pretest on mathematical equivalence problems, a posttest given immediately after the intervention, and then delayed posttest. The intervention was done in a single session, in which students worked one-on-one with a researcher in a 40-minute session. Children solved problems, reported on their solution strategies, and were provided with feedback. In the direct instruction condition students were told explicitly how to solve the problems. In the student-centered condition no instruction was given, and children were asked to “think of a new way to solve the problem.” Prompts for doing self-explanation were introduced in the two groups in that condition. The posttests measured procedural learning, procedural transfer, and conceptual understanding. Comparison of the discovery learning and prompts group to the direct instruction with prompts group yielded nonsignificant but negative (i.e., favoring the direct instruction group) results on all three outcome measures (note ES = -0.272 and -0.458 on the procedural and conceptual tests, respectively).

Four additional studies involved comparisons, in some form, of teacher-directed and student-centered strategies, in which no significant effect sizes were found. The studies by Cicchelli (1982) and Kameenui et al. (1986) introduced a direct instruction-type treatment and compared to more student-centered instruction. The Rudnitsky et al. (1995) and Fuchs et al. (2006) studies both implemented clearly specified student-centered instruction and compared to more of a “business as usual” condition. Effect sizes in Kameenui et al., while non-significant (ES = 0.444) favored the direct instruction condition. In the Ciccelli study, the effect favored the nondirect instruction condition for low and high ability students (ES = -0.614 and -0.374, respectively) and the direct instruction group for medium ability students (ES = 0.156).

Because this set of studies differs in terms of the nature of the intervention, pooled effect sizes have not been calculated. In summary, note that there is no conclusive evidence from this set of studies to support either a teacher-directed or student-centered approach to mathematics instruction.

4. Project Follow Through Evaluation Studies

Three relevant models of the extensive National Longitudinal Evaluation of Project Follow Through (Stebbins et al., 1977) were located that met the criteria for inclusion. At that point in history, the evaluation of Project Follow Through (FT) was the largest and most expensive evaluation of any intervention conducted in education or any of the social sciences. Follow Through's evaluation involved longitudinal quasi-experiments conducted with kindergarteners through third-graders in low-income schools across the country. The outcome measures testing mathematical achievement were the computations, concepts, and problem solving subtests of the Metropolitan Achievement Test. The three Follow Through models included in this report that tested teacher-directed or student-centered instruction and had equivalent groups at baseline were Direct Instruction, Cognitively Oriented Curriculum, and EDC Open Education. Results are reported in Table 1.

Direct Instruction "use(d) a fast moving series of programmed … (i.e., scripted) ... questions and answers … teachers present specified questions to elicit a verbal child response. Proper responses are reinforced and wrong answers are corrected according to specified procedures" (Stebbins et al., 1977, p. 65). Programmed instruction materials are used, and students work in small homogeneous groups; frequent criterion-referenced tests are provided. The Task Group considers this to be a teacher-directed intervention.

The EDC Open Education approach states, "Children learn at individual rates and in individual ways…" (Stebbins et al., 1977, p. 113). The approach has its roots in British infant schools and in the developmental theories of Piaget. The instruction occurs in an open setting and children are provided with a wide range of materials for learning. This is a student-centered model. The Cognitively Oriented Curriculum model, also a developmental model, was aimed at "developing children's ability to reason." The curriculum is based on the use of learning centers in which "children choose their activities and work with teachers in small groups" (Stebbins et al., p. 89). This too is a student-centered intervention.

Concurrent with the impact evaluation, Stallings (1975), conducted an extensive observational study of the activities in the Follow Through classrooms and their corresponding control group (i.e., business as usual) classrooms. Using a complex, reliable observational system, they were able to predict which classrooms were FT and which were control, and to discriminate between each example (FT) and control classroom with over 80% accuracy. They consistently found significant differences between each FT approach and its control condition, and between the various FT models.

When Stallings (1975) compared the Direct Instruction approach with "non Follow Through" instruction, which would have been a version of "business as usual" at the time they were essentially comparing a highly structured, teacher-directed intervention (based on principles of instructional design and concept development derived from learning theory and applied behavior analysis, to say nothing of the genius of Englemann) with a more general type of teacher-directed instruction.

The 1977 project evaluation (Stebbins et al., 1977) found significant effect sizes favoring Direct Instruction on the Computation subtest only (ES = 0.315). No other effect sizes were significant, nor was the overall effect on the Metropolitan Achievement Test Mathematics composite significant.

Another Project Follow Through model study compared the Cognitively Oriented Curriculum to a non-Follow Through business as usual condition. The effect size on the Concepts subtest was significant (-0.355), and the effect sizes on the Computation and Problem Solving measures were marginally significant (ES = -0.318 and ES = -0.295). Note that all effect sizes were negative, favoring the teacher-directed control condition. The final study in this group involved a comparison of the Open Education program to a business-as-usual control condition. There were no significant effect sizes.

No pooling was done of these studies.

5. Conclusion

The studies produced a mix of significant effect sizes favoring student-centered instruction, and others favoring teacher-directed instruction, together with findings of no significant effects. As a result, the research does not lead to any conclusive result about the value of student-centered instructional strategies in comparison to teacher-directed instructional strategies. Under some conditions, with some groups of students, and for some kinds of outcomes, an isolated study may find that either teacher-directed or student-centered strategies are preferable. In general the evidence does not provide a case for favoring or promoting either strategy over the other. The Task Group points out that in only one of the studies reviewed is “teacher-directed” instruction the experimental treatment.

B. Cooperative Learning and Peer Tutoring

Cooperative learning strategies offer students an opportunity to learn from and with other students. However, the means by which cooperative learning plays out in classrooms varies along many dimensions. For instance, tasks assigned in cooperative learning groups range from practice on teacher-taught skills to learning methods of problem solving. Students can be grouped homogeneously or heterogeneously by ability. Students may be assigned specific roles within a cooperative group or they may decide for themselves how to accomplish a group task. Group and individual accountability operate differently in different cooperative learning settings. Slavin (1991) describe a cooperative learning strategy as when groups work to earn some type of recognition or award based on the individual learning of every group member. Group members’ individual learning is measured by success on assignments, quizzes, and tests. Students are motivated to help each other learn so that individual achievement increases and, as a result, the group receives awards or recognition. Cooperative learning may include individual accountability or group reward structures.

Good, Mulryan, and McCaslin (1992) conclude that small-group instructional approaches are supported by research that indicated students need to be more active. “[Research] suggests that students are too passive and need to become more involved intellectually in classroom activities" (p. 167). They go on to note, "Many writers interpret recent cognitive science research as suggesting the need for less teacher support and for more learner independence (see Nickerson, 1988), [although] what this means in practice is far from clear. For example, does heavy reliance on more talented peers mean less dependency on the teacher? Do different skills and concepts require different amounts of expert modeling and coaching so that simple statements about appropriate practice are highly misleading?" (p. 167).

A number of studies compare cooperative grouping strategies to more traditional whole-class instruction, or in some cases, to individual practice that is part of teacher-directed instruction. Our review is organized by the following categories: studies of very specific approaches to cooperative learning [Team Assisted Individualization (TAI), Student Teams-Achievement Division (STAD), Peer Assisted Learning (PALS)]; studies of other collaborative learning strategies; combination strategies involving cooperative or collaborative learning; and cooperative learning approaches in the context of technology.

Note that our search of the literature and analyses are not concerned with affective outcomes, only on measures of mathematics achievement.

1. Specific Approaches to Cooperative Learning Team Assisted Individualization (TAI)

This strategy combines individualization with cooperative work. In TAI, students are grouped in heterogeneous teams of four or five persons. Each student receives a set of mathematics problems tailored to individual performance on a diagnostic test. Students help each other when needed and check each others' work. Rewards are based on group performance on assignments, quizzes, and tests. Tests at the end of the unit are taken individually.

Six studies, described within four separate papers, met our criteria for review and examined the effect of TAI on some type of mathematical outcome. The pooled effect size for computation outcomes on student-level analyses was significant (ES = 0.377), favoring the TAI condition. Slavin, Leavey, and Madden (1984) report on two studies with elementary school students focused on computation with decimals and fractions, and with word problems. In one randomized controlled trial (RCT) study, TAI was contrasted with whole-class lectures and group-paced instruction. The second study, a quasi-experiment, involved fourth- through sixth-graders, with the same type of control condition. A third study (Slavin, Leavey, & Madden, 1984), involving 1,371 students in Grades 3 to 5, again compared TAI with whole class lectures. The fourth paper in this set (Xin, 1999) involved third-grade students in an RCT focusing on arithmetic topics including basic fact families, calculation, coins, and place value. A large number of mainstreamed special education students (14%) were involved. The TAI treatment was coupled with a CAI component; the control condition was whole class instruction coupled with the same CAI.

All of these studies allowed for student-level analyses to examine effect sizes on an outcome measure of computation, the California Test of Basic Skills-Computations (CTBS) for the Slavin studies, and the Stanford Achievement Test-Math for the Xin study. Five contrasts were examined—the three in the Slavin et al. studies, and results for two groups in the Xin study (regular education, and LD students). No effect sizes were significant separately, except for the significant effect size of 0.595 comparing TAI with computer-assisted instruction (CAI) in the Xin study to students in the regular education group.

Two additional studies, both RCTs, are reported in Slavin and Karweit (1985). One worked with students in Grades 4 through 6, the other with third- through fifth-graders, using TAI over a period of 18 and 16 weeks. The mathematics topics were decimal and fraction arithmetic, introduction to algebra, and word problems. The control conditions were the Missouri Mathematics Program (a form of teacher-directed instruction) and a business as usual control condition. Classroom-level effects (ES = 0.709, 0.562) on computation outcomes were significant and favored the TAI intervention on computation scores. In three of these studies a concept outcomes measure was also included. Slavin et al. (1984b) and Slavin & Karweit (1985)—Studies 1 and 2 included outcomes on the CTBS-Concepts test. Three contrasts were computed, all proving to be nonsignificant; the pooled effect size of the classroom-level analyses (0.018) also was nonsignificant.

The studies are summarized in Tables 2a and 2b. It can be concluded that the implementation of TAI for students in Grades 3 through 6, in comparison to a form of whole class instruction, benefits computation skills. Note that this finding applies only to the very particular cooperative group strategy of TAI and only to computation, not concepts or problem solving.

2. Student Teams-Achievement Division (STAD)

This form of cooperative learning developed by Slavin and colleagues, involves four-to-five member homogeneous teams studying together after teacher presentation. Individual quizzes are taken and rewards are at the team level.

Four studies of STAD, all randomized controlled experiments, met our criteria for review. They are summarized in Table 3. No significant effect sizes were produced in this set, although all effect sizes were positive, favoring the STAD intervention. Jacobs (1996) examined the performance of third- through fifth-graders, content not specified, in a STAD condition and then in an individual student accountability condition, and produced non-significant effects favoring STAD of 0.573, 0.484, and 0.454, for third-, fourth-, and fifth-graders respectively. STAD did not provide any particular advantage to student participants in comparison with more teacher-directed classroom strategies as implemented in these four studies.

3. Peer Tutoring Approaches

The studies in this section examine the impact of a small group instruction approach that features peers learning from and with their peers, in variations of peer-tutoring. One particular version, *Peer-Assisted Learning Strategies* (PALS) (http://kc.vanderbilt.edu/pals/), "is a version of classwide peer tutoring. Teachers identify which children require help on specific skills and [whom] the most appropriate children are to help other children learn those skills. Using this information, teachers pair students in the class, so that partners work simultaneously and productively on different activities that address the problems they are experiencing. Pairs are changed regularly, and over a period of time as students work on a variety of skills, all students have the opportunity to be 'coaches' and 'players'." (http://kc.vanderbilt.edu/pals/). The strategy also creates opportunities for a teacher to circulate in the class, observe students, and provide individual remedial lessons.

Five studies that examine various versions of peer tutoring met our criteria for review, allowing for calculation of effect sizes for 20 different contrasts at the student and classroom levels, and for computation and concepts outcome measures. Three of the studies in this section use PALS. The studies are summarized in Table 4. Pooled effect sizes examining computation outcomes at the classroom level were significant (ES = 0.431), and approached significance at the student level, favoring the peer tutoring interventions. In none of these studies were significant effect sizes produced on concept outcome measures. Nor, when doing student-level analyses, were any individual effect sizes on computation significant.

The pooled effect size on the computation outcomes for the three studies that used student-level analyses was 0.238, which approached statistical significance. In a 15-week randomized controlled study of kindergartners (Fuchs et al., 2001) PALS was compared to a control condition that was described as teacher-directed lessons and demonstrations. Positive but not significant effects on the Stanford Early School achievement test in student-level analyses were detected for special education students, low-achieving students, and medium-achieving students (effect sizes respectively, of 0.431, 0.374, and 0.436). In a study of first-graders comparing PALS with a business-as-usual basal core curriculum (Fuchs, Fuchs, Yazdian, & Powell, 2002), no statistically significant effects were found. Ginsburg-Block and Fantuzzo (1998) implemented an RCT with low-achieving third- and fourth-grade students in an urban elementary school using a reciprocal peer tutoring (RPT) model (Palincsar & Brown, 1984) on the mathematics achievement of low SES students. Results were nonsignificant but favored the peer collaboration condition (ES = 0.590).

The effect sizes of the two studies that included classroom-level analyses were also pooled (Fuchs et al., 1995; Fuchs et al., 1997). These examined the impact of peer-assisted strategies, allowing for effect size calculations on seven different contrasts. The 1995 RCT study was done in second through fourth grade classrooms, over a 23-week period in which two 25–30 minute sessions per week were done using PALS, integrated with regular assessments. The control condition was teacher-mediated instruction, and the topic focus arithmetic operations. The 1997 study, also an RCT, also worked with second- through fourth-grade classrooms using peer-mediated versus teacher-mediated instruction (Fuchs et al., 1997). They investigated the effects of a peer tutoring program in which students received explicit instruction on how to provide elaborated help. Students were taught to provide explanations that would encourage peers to solve problems for themselves (instead of simply giving answers), referred to as "elaborated PMI." These interventions were modeled separately, with students assigned to two experimental treatments—PMI-elaborated and PMI elaborated plus conceptual. Both studies had both computation and conceptual outcome measures.

In both studies contrasts were calculated for LD, low-achieving, and average-achieving students. In the 1997 study there was also a comparison of high-achieving students. The results are interesting and mixed. In the 1995 study, the only significant effect size was for low-achieving students (ES = 0.728), on the computation measure, favoring PALS. Only two other effect sizes approached significance in this set: in the 1997 study, for both LD and low-achieving students, on the computation tests, the effect sizes were appreciable (0.663 and 0.704), on the computation outcome measure only, favoring the peer-assisted condition. No other effect sizes were significant, although all but one (in the 1997 study, LD on the concepts outcome measure) were positive. However, when the seven effect sizes for classroom-level computation outcome measures were pooled for these two studies, the result was a significant effect size (ES = 0.431).

In summary, it appears that peer tutoring strategies may be promising in teaching young children mathematical operations (which may not be exclusively computation oriented). However, this finding must be treated cautiously because the evidence is only suggestive. For the pooled effects of the Fuchs et al. (1995) and Fuchs et al. (1997) studies, there are some important limitations. The two studies were both in Grades 2–4, involve learning whole number operations, were possibly conducted using the same sample of schools, and were conducted by the same research team. The 1995 Fuchs et al. study did not include high-achieving students. Moreover, in the 1995 study, teachers in the peer tutoring condition were regularly provided with formative assessment data to guide instruction, but teachers in the control condition were not. Thus, the study does not provide a clear contrast between peer tutoring and a more teacher-directed form of instruction. The extent to which the positive effects that were detected were produced by formative assessment, by peer tutoring, or by an interaction of the two interventions cannot be determined.

4. Other Collaborative Learning Strategies

Five studies of other collaborative learning strategies met the criteria for inclusion, all of them RCTs. The studies are referred to as “collaborative learning” because they do not feature interventions as structured as the cooperative learning techniques featured above, but they all utilize methods of student grouping that involve student-to-student collaboration in learning. Two of the seven contrasts computed yielded effects that were significant.

A study by Barron (2000) produced statistically significant effect sizes favoring the collaborative condition in solving complex video-based mathematical problems. Sixth-graders enrolled in a public magnet school serving academically talented children were assigned randomly to either a group or individual condition. The task was to solve video-based problems from *The Adventures of Jasper Woodbury* series. Students first viewed the 15-minute episode, “Journey to Cedar Creek,” which describes several dilemmas facing a character who is considering the purchase of a boat. In the second session, students were asked to solve these problems either individually or in teams of three by completing exercises in a workbook. In the third session, students were asked to solve the problems again, this time individually, regardless of the condition to which they had been assigned in the previous session. In the fourth and final session, students viewed a 5-minute video posing a parallel problem to assess near transfer of the acquired skills.

Students who had worked in triads solved more of the problems correctly than students who had worked individually at significant levels (ES = 0.472). The effect on a transfer task approached the level of significance (ES = 0.392).

In a study of the effects of a communal environment on African-American students' learning of mathematical estimation (Hurley et al., 2005), effects were significant. The researchers drew a sample of fifth-grade African-American students from two urban public schools. The students had to fall within the middle 75% for their classroom and school in terms of academics and behavior. The children in the sample came from a low SES background as measured by both their school's participation in Title I and the students' participation in the free and reduced-price lunch program.

Boys and girls were divided equally between two treatments though within gender students were randomly assigned. One experimental condition was highly communal; students learned estimation working in groups of three. Students sat at the same table and shared one set of materials. Each study session included the experimenter reading a communal prompt to the students. Students were asked to hold hands and were reminded that they were members of a group and should work hard and help one another because they were members of the same school and community. The other condition was low-communal; students studied alone in sets of three. Each student had his or her own set of materials and sat at his or her own desk. These study sessions included an individualized prompt to remind students they could earn a reward if their scores increased and to work hard on their own to improve their scores. Before and after their study sessions, all students took a 15-question test on mathematics estimation. The intervention was very brief (20 minutes). The effect size (ES = 0.655) favoring the triads groups was significant.

The studies by Janicki and Peterson (1981), Kramarski and Mevarech (2003), and Peklaj and Vodopivec (1999) all compared some form of cooperative group strategies, with no significant effects (see Table 5).

5. Strategies Combining Collaborative or Cooperative Learning With Other Approaches

Also meeting the criteria were three studies that examined cooperative learning strategies used in conjunction with other instructional practices. Because the cooperative learning elements were mixed with other modifications in practice, the Task Group was not able to isolate the effects of cooperative learning alone.

Only one of the studies yielded a statistically significant effect size, on a computation outcome measure. A quasi-experimental study by Busato et al. (1995) investigated Adaptive Instruction and Cooperative Learning (AGO),[1] a Dutch model that includes curricular adaptations, whole class instruction to introduce a topic, small group cooperation, regular assessments, individual work with the possibility of students helping each other, remedial groups working with direct guidance from a teacher, and whole class reflection. The study involved 572 middle school students in the Netherlands, and the intervention lasted for one month. The mathematical topic was pre-algebra ideas, and the interventions were AGO versus "a more traditional instructional method (mainly without group work)" (p. 671). The effect size for boys was significant (0.681) and for girls approached significance (0.583).

[1] AGO refers to the Dutch model called Adaptief Groeps-Onderwijs.

A quasi-experiment by Stevens and Slavin (1995) involved 1,012 elementary school students in Grades 2–6 in a year-long comparison of cooperative learning vs. whole class instruction. The nonsignificant effect on the California Achievement Test was negative, favoring the whole group instruction, on the applications test. The effect on the CAT computation (ES = 0.120) approached significance favoring the cooperative learning treatment. And finally, the Brenner et al. study (1997) discussed previously, small group instruction was compared to a control condition. The effects, favoring the small group condition, were not significant. The effect size for the pooled problem solving outcomes was 0.074.

6. Cooperative Learning Strategies in the Technology Context

Here the Task Group presents a review of the research that examined the use of cooperative learning strategies in the context of technology-based instruction. Use of some form of collaborative learning in computer-based instruction (CBI) is suggested by several positive reports from preschool to college (e.g., Leron & Lavy, 2004; Light & Blaye, 1990; Nastasi & Clements, 1994; Scardamalia et al., 1992; Schofield, 1995; Strommen, 1993). The following section discusses the eight studies identified that examined the use of cooperative learning strategies in the context of computers; the studies are summarized in Table 7. Only two of the studies produced significant effect sizes.

Seven studies investigated learning on computers in groups versus learning on computers individually (Hooper, 1992; Hooper, Temiyakam, & Williams, 1993; Mevarech et al., 1991; Mevarech, 1993, 1994; Slavin & Karweit, 1984; and Xin, 1999). All participating students were in elementary school. Durations ranged from very short (1 week) to an academic year. The outcomes measured in most studies were limited to computation, but the Hooper (1992, 1993), and the Slavin and Karweit (1984) studies also assessed mathematical concepts.

It was possible to calculate 15 different effect sizes across these studies. Only those that were significant are highlighted. Weiss et al. (2006) studied kindergarten students in Israel learning about numbers and operations. One treatment was a multimedia environment involving cooperation while the other was a multimedia environment involving an individual learning style. The outcome measure was a skills test on numbers and operations. The significant effect size (-0.862) favored the individual teaching style treatment. In the study by Xin (1999) that compared TAI involving CAI to whole-class computer-assisted instruction, the effect size on the Standard Achievement Test-Math for regular education students was significant (ES = 0.595), favoring the small group-based treatment. Effects for all but two studies were positive. A negative effect was found in the Weiss et al. (2006) study and for low- and high-achieving students in the Mevarech (1993) study. None of the other effect sizes calculated reached significance.

In summary, implications for policy and practice do not indicate a simple solution, such as, "Students should work together on computers." Positive effects of cooperative learning in technological contexts can be obtained, but they may be limited in size, especially when using simple CAI programs, and may depend on teachers' management and guidance of positive interactions and collaboration.

C. Summary and Conclusions

At this time a body of scientifically sound research does not exist that will quell the controversies about the best way to instruct students in mathematics. As evident from the review, however, educators are not completely in the dark about effective practice. Three principal findings emerge from the literature. First, some of the limitations in the studies reviewed in this section will be discussed.

Although all of the studies discussed here met the technical criteria for inclusion, there were a number of issues relative to the studies that could be addressed in future research. In some cases, for instance, the treatments were of such brief duration that it is difficult to interpret the conclusions. In many cases the control condition is inadequately specified leaving the reader to make assumptions about exactly what is being compared to what. In the case of teacher-directed and student-centered instruction, the terms are so vaguely defined to begin with that this is a serious problem in using the research literature to draw conclusions that may be useful for policy. Only some of these studies actually measure and document the nature of the intervention, leaving questions about the fidelity and extent of implementation and thus lack of clarity about what might be causing the results. This body of work tends to include examinations of specific groups of students, which is informative relative to specific groups, but needs to be balanced with studies that look at broader populations. Finally, in some cases the team evaluating the effectiveness of the intervention also invented the method (although it is emphasized that all studies included in this report met the stringent criteria for inclusion).

The review does allow us to make some key conclusions. First, Team Assisted Individualization (TAI), a cooperative learning strategy, has been shown to be effective in teaching computation skills. The finding does not extend to problem-solving skills or mathematical concepts. It is critical to note that the strategy involves much more than simply putting students into groups. Students first take diagnostic tests, and teachers utilize the results to prepare individualized sets of worksheets that target weak computation skills. Working in heterogeneous groups of four or five students, students are encouraged to work together to ensure that all students in the group attain mastery. Teachers work with small groups of students, pulled from different teams who are working on the same skill (e.g., division of decimals). Cooperation within groups is reinforced by group rewards given on the basis of final tests (for a more detailed description of TAI, see Slavin, Madden, and Leavey (1984) and Slavin and Karweit (1985)).

Why does TAI work? Researchers of TAI have argued that several elements of the technique may enhance learning: students receive immediate feedback from peers (as compared to delayed feedback from teachers during whole class instruction); materials present mathematical skills in a logical, hierarchical sequence; students' deficient areas are assessed, identified, and targeted with individualized materials; a group reward structure motivates students and encourages teamwork; the intervention blends teacher-directed and student-centered instruction. More research is needed to identify the precise mechanisms of TAI's effectiveness.

A second cooperative learning strategy, generally known as peer tutoring, also showed signs of promise, with a significant pooled effect size favoring the peer-assisted condition. This finding must be treated cautiously, in that it involves only two studies, is limited to the study of whole number operations by students in Grades 2-4, and only reaches statistical significance at the class level. In one of the studies, formative assessment was a key component of the intervention. Studies on which student level effects could be calculated did not produce statistically significant findings. As with TAI, the treatment is highly specific and involves far more than having students work in pairs. In both of these sets of studies, it is important to underscore that significant effect sizes were found only for computation or operations, not for mathematical concepts or problem solving.

The second main finding pertains to problem solving. Three studies were reviewed that documented successful far transfer of problem solving skills after extensive instruction. Muthukrishna and Borkowski (1995) studied third-graders who were taught a part-whole problem solving strategy in 14 lessons. Brenner et al. (1997) investigated middle school pre-algebra students learning how to represent function problems in multiple formats. The program consisted of 20 lessons. In both experiments, students in the guided discovery condition outperformed students in the traditional instruction condition on measures of far transfer. These effects were not statistically significant but they bordered on significance. Educators considering whether to implement these interventions would have to weigh the limitations of the outcomes—problem solving strategies that are restricted to particular topics in mathematics—with the amount of instructional time spent to attain them. Whether the benefits of guided discovery extend to content beyond the areas examined in these studies, or can be accomplished in less time, has not been studied.

In contrast, there were three studies (not counting the cooperative group studies) in which significant effects were found, favoring the teacher-directed instructional approach), for performance on computation outcome measures. Hopkins et al. (1997) found that the "didactic" treatment led to better performance by girls on an arithmetic computation test. And, in the Project Follow Through Evaluation, Stebbins et al. (1977) found significant effects favoring direct instruction, and nearly significant effects favoring the control treatment (in contrast to the Cognitively Oriented Curriculum) on the computation outcome measure. It is possible that under some conditions, with certain mathematical emphases and particular groups of students, the teacher-directed approaches can lead to better performance on computational assessments than more student-centered approaches.

That leads to the final principal finding. Much more research is needed that directly compares the effectiveness of student-centered and teacher-directed instruction, and that provides clear operational definitions for these terms. In particular, research is needed with teacher-centered instruction as the experimental condition. Almost all of the research reviewed here investigated experimental modes of instruction that are student-centered—whether guided discovery, cooperative learning, or peer tutoring—with the control condition described as "teacher-directed" or "traditional" or "direct instruction." Experiments with better specified teacher-directed interventions would enhance our understanding of how to improve classroom instruction in mathematics. A comprehensive program of research might succeed in transforming what has been a clash of ideologies into a search for effective practice.

Table 2a: Studies That Investigated the Effects of Team Assisted Individualization (TAI) on Computation Outcomes

Study	Design	Sample	Duration/ Content	Contrast	Measure	Subgroup	Hedge's g	Standard Error	
Computation Outcomes									
Student-level analyses									
Slavin, 1984—Study 1[b]	RCT	504 students in Grades 3–5, including 6% who were receiving special education services, in 18 mathematics classes in six schools in a suburban Maryland school district	8 weeks/ Addition, subtraction, multiplication, division, numeration, decimals, fractions, and word problems	TAI vs. whole class lectures and group paced instruction	CTBS—Computations	Overall	0.103	(ns)	0.460
Slavin, 1984—Study 2[b]	Quasi	375 students in Grades 4–6, including 27% who were receiving special education services, in 16 mathematics classes in four schools in a suburban Maryland school district	10 weeks/ Addition, subtraction, multiplication, division, numeration, decimals, fractions, and word problems	TAI vs. whole class lectures and group paced instruction	CTBS—Computations	Overall	0.109	(ns)	0.460
Slavin et al., 1984[b]	Quasi	1,371 students in Grades 3–5, including 8% that received special education services, in 59 mathematics classes in eight schools in a suburban Maryland school district	24 weeks/ Unspecified math curriculum (likely same topics as above)	TAI vs. whole class lectures and group paced instruction	CTBS—Computations	Overall	0.147	(ns)	0.331
Xin, 1999	RCT	118 third-grade students in six mathematics classes in three schools	Daily for one semester/ Basic fact families including addition, subtraction, multiplication, and division; coin recognition, place value, concepts, number patterns	TAI w/CAI vs. whole class w/CAI	Stanford Achievement Test—Math	Regular education	0.595	**	0.210
						Learning disability	0.338	(ns)	0.390
Heterogeneity									
Q-value	*df (Q)*	*P-value*	*I-squared*	*Pooled ES (four studies, five effect sizes)*			*Hedge's g*	*Standard Error*	
2.265	*4*	*0.271*	*0.687*				*0.377*	****	*0.145*
Classroom-level analyses									
Slavin & Karweit, 1985—Study 1	RCT	345 students in Grades 4–6 in 15 mathematics classes	18 weeks/ Addition, subtraction, multiplication, division, numeration, decimals, fractions, ratios, statistics, introduction to algebra, and word problems	TAI vs. Missouri Mathematics Program	CTBS—Computations	Overall	0.709	***	0.143
Slavin & Karweit, 1985—Study 2	RCT	480 students in Grades 3–5 in 22 mathematics classes in and around Hagerstown, MD	16 weeks/ Addition, subtraction, multiplication, division, numeration, decimals, fractions, ratios, statistics, introduction to algebra, and word problems	TAI vs. As-Is Control	CTBS—Computations	Overall	0.562	***	0.138
Heterogeneity									
Q-value	*df (Q)*	*P-value*	*I-squared*	*Pooled ES (two studies, two effect sizes)*			*Hedge's g*	*Standard Error*	
0.548	*1*	*0.459*	*0.000*				*0.633*	*****	*0.099*

~ p < .10, * p < .05, ** p < .01, *** p < .001
[a] Data were adjusted for clustering that occurred within classrooms.
[b] Data were adjusted for clustering that occurred within schools.

Table 2b: Studies That Investigated the Effects of Team Assisted Individualization (TAI) on Concepts Outcomes

Study	Design	Sample	Duration/ Content	Contrast	Measure	Subgroup	Hedge's g	Standard Error	
Concepts Outcomes									
Student-level analyses									
Slavin, 1984[b]	Quasi	1,371 students in Grades 3–5, including 8% that received special education services, in 59 mathematics classes in eight schools in a suburban Maryland school district	24 weeks/ Unspecified math curriculum (likely same topics as above)	TAI vs. whole class lectures and group paced instruction	CTBS—Concepts	Overall	0.098	(ns)	0.331
Classroom-level analyses									
Slavin & Karweit, 1985—Study 1	RCT	345 students in Grades 4–6 in 15 mathematics classes	18 weeks/ Addition, subtraction, multiplication, division, numeration, decimals, fractions, ratios, statistics, introduction to algebra, and word problems	TAI vs. Missouri Mathematics Program	CTBS—Concepts	Overall	-0.003	(ns)	0.139
Slavin & Karweit, 1985—Study 2	RCT	480 students in Grades 3–5 in 22 mathematics classes in and around Hagerstown, MD	16 weeks/ Addition, subtraction, multiplication, division, numeration, decimals, fractions, ratios, statistics, introduction to algebra, and word problems	TAI vs. As-Is Control	CTBS—Concepts	Overall	0.038	(ns)	0.134
Heterogeneity									
Q-value	*df (Q)*	*P-value*	*I-squared*	*Pooled ES (two studies, two effect sizes)*			*Hedge's g*	*Standard Error*	
0.045	*1*	*0.832*	*0.000*				*0.018*	*(ns)*	*0.097*

~ p < .10, * p < .05, ** p < .01, *** p < .001
[a] Data were adjusted for clustering that occurred within classrooms.
[b] Data were adjusted for clustering that occurred within schools.

Table 3: Studies That Investigated the Effects of Student Teams-Achievement Divisions (STAD)

Study	Design	Sample	Duration/ Content	Contrast	Measure	Subgroup	Hedge's g	Standard Error	
Alkhateeb & Jumaa, 2002[a]	RCT	111 eighth-grade students in four classes in two schools in the United Arab Emirates	3 weeks/ Algebraic expressions	STAD vs. Whole Class Instruction	Algebra test	Overall	0.108	(ns)	0.479
Jacobs, 1996[a]	RCT	266 students in Grades 3–5 at a large private Christian/ fundamentalist elementary school in the Southeast	9 weeks/ Unspecified 3rd grade curricular unit	STAD vs. Direct Instruction with rewards and student individual accountability	Curriculum specific math test	Third grade	0.573	(ns)	0.492
						Fourth grade	0.484	(ns)	0.487
						Fifth grade	0.454	(ns)	0.486
Madden & Slavin, 1983[a]	RCT	183 third-, fifth-, and sixth-grade students, including 40 special education students, in six math classes in the Baltimore City schools	7 weeks/ Unspecified 3rd-, 5th-, and 6th-grade curricular units	STAD vs. Focused Instruction (whole class lectures, individual practice, quizzes, and individual recognition)	Curriculum specific math test	Overall	0.124	(ns)	0.402
Slavin & Karweit, 1984[a]	RCT	588 ninth-grade students in 25 math classes in 16 inner city Philadelphia junior and senior high schools	One school year/ Unspecified 9th-grade general math curriculum	STAD vs. Focused Instruction (students worked individually and did not receive team recognition)	CTBS, Shortened version (computation, concepts and applications subscales)	Overall	0.113	(ns)	0.221
Heterogeneity									
Q-value	*df (Q)*	*P-value*	*I-squared*	*Pooled ES (four studies, six effect sizes)*			*Hedge's g*	*Standard Error*	
1.384	*5*	*0.926*	*0.000*				*0.227*	(ns)	*0.152*

~ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

[a] Data were adjusted for clustering that occurred within classrooms.

[b] Data were adjusted for clustering that occurred within schools.

Table 4: Studies That Investigated the Effects of Peer Assisted Learning

Study	Design	Sample	Duration/ Content	Contrast	Measure	Subgroup	Hedge's g	Standard Error	
Computation Outcomes									
Student-level analyses									
Fuchs et al., 2001[a]	RCT	168 Kindergarten students in 20 classes in five schools in a Southeastern metropolitan area	15 weeks/ Kindergarten core curriculum	PALS vs. Teacher-directed lessons and demonstrations	Stanford Early School Achievement Test	Special education	0.431	(ns)	0.524
						Low achieving	0.374	(ns)	0.454
						Medium achieving	0.436	(ns)	0.270
						High achieving	-0.162	(ns)	0.380
Fuchs et al., 2002[a]	RCT	327 first-grade students in 20 classrooms in a Southeastern metropolitan public school system	16 weeks/ Addition, subtraction, counting, sets, geometry, and measuring	PALS vs. As-is basal core curriculum	Stanford Achievement Test	Overall	0.055	(ns)	0.223
Ginsburg-Block & Fantuzzo, 1998	RCT	104 low-achieving third- and fourth-grade students in an urban elementary school	Two 30-minute sessions per week for 7 weeks/ Addition, subtraction, multiplication, and division computation and word problems	Peer collaboration dyads vs. Control	Curriculum based computation test	Overall	0.590	(ns)	0.389
Heterogeneity									
Q-value	*df (Q)*	*P-value*	*I-squared*	*Pooled ES (three studies, six effect sizes)*			*Hedge's g*	*Standard Error*	
3.361	*5*	*0.645*	*0.000*				*0.238*	*~*	*0.134*
Classroom-level analyses									
Fuchs et al., 1995	RCT	40 Grade 2–4 classrooms in nine elementary schools in a Southeastern, urban school district	Two 25-30 minute sessions per week for 23 weeks/ Grade level's annual operations curriculum	PALS integrated with regular assessments vs. Teacher-mediated instruction	Acquisition learning: Math Operations Test—Revised	Learning disability	0.260	(ns)	0.311
						Low achieving	0.728	**	0.320
						Average achieving	0.297	(ns)	0.312
Fuchs et al., 1997	RCT	40 Grade 2–4 classrooms in a Southeastern metropolitan public school system	18 weeks/ Number concepts, counting, word problems, charts/graphs, money, measurement, geometry, and computation	Peer-mediated instruction vs. Teacher-mediated instruction	Comprehensive Mathematics Test—Operations	Learning disabilities	0.663	~	0.386
						Low achieving	0.704	~	0.388
						Average achieving	0.177	(ns)	0.378
						High achieving	0.242	(ns)	0.378
Heterogeneity									
Q-value	*df (Q)*	*P-value*	*I-squared*	*Pooled ES (two studies, seven effect sizes)*			*Hedge's g*	*Standard Error*	
2.904	*6*	*0.821*	*0.000*				*0.431*	****	*0.132*

Continued on p. 6-37

Table 4, continued

Study	Design	Sample	Duration/ Content	Contrast	Measure	Subgroup	Hedge's g	Standard Error	
Concepts Outcomes									
Classroom-level analyses									
Fuchs et al., 1995	RCT	40 Grade 2–4 classrooms in nine elementary schools in a Southeastern, urban school district	Two 25–30 minute sessions per week for 23 weeks/ Grade level's annual operations curriculum	PALS integrated with regular assessments vs. Teacher-mediated instruction	Acquisition learning: Math Concepts and Applications	Learning disability	0.199	(ns)	0.311
						Low achieving	0.063	(ns)	0.310
						Average achieving	0.307	(ns)	0.312
Fuchs et al., 1997	RCT	40 Grade 2–4 classrooms in a Southeastern metropolitan public school system	18 weeks/ Number concepts, counting, word problems, charts/graphs, money, measurement, geometry, and computation	Peer-mediated instruction vs. Teacher-mediated instruction	Comprehensive Mathematics Test-Concepts	Learning disabilities	-0.016	(ns)	0.377
						Low achieving	0.515	(ns)	0.383
						Average achieving	0.139	(ns)	0.377
						High achieving	0.099	(ns)	0.377
Heterogeneity									
Q-value	*df (Q)*	*P-value*	*I-squared*	*Pooled ES (two studies, seven effect sizes)*			*Hedge's g*	*Standard Error*	
1.406	*6*	*0.965*	*0.000*				*0.186*	*(ns)*	*0.130*

~ p < .10, * p < .05, ** p < .01, *** p < .001
[a] Data were adjusted for clustering that occurred within classrooms.
[b] Data were adjusted for clustering that occurred within schools.

Table 5: Studies That Investigated Other Cooperative Learning Strategies

Study	Design	Sample	Duration/ Content	Contrast	Measure	Subgroup	Hedge's g	Standard Error	
Barron, 2000	RCT	96 sixth-grade students in a public magnet school for academically talented children	Four 1-hour sessions/ Contextual problem solving	Problem solving collaboratively in triads vs. Problem solving individually	Solutions to story problems—mastery	Overall	0.472	*	0.205
					Solutions to story problems—transfer	Overall	0.392	~	0.204
Hurley et al., 2005	RCT	78 African-American fifth-grade students in two urban public schools	One 20-minute session/ Math estimation	Triads worked together in a high-communal setting vs. Individuals worked in a low-communal setting	Math estimation task	Overall	0.655	**	0.230
Janicki & Peterson, 1981[a]	RCT	117 fourth- and fifth-grade students	2 weeks/ Fractions	Small group direct instruction vs. Individual direct instruction	Researcher developed test on fractions	Overall	-0.041	(ns)	0.188
Kramarski & Mevarech, 2003[a]	RCT	384 eighth-grade students in 12 classrooms in four Israeli junior high schools	2 weeks/ Linear graphing	Metacognitive and cooperative groups vs. Metacognitive and individual	Graph interpretation test	Overall	0.355	(ns)	0.387
				Cooperative groups vs. Individual work	Graph interpretation test	Overall	0.105	(ns)	0.388
Peklaj & Vodopivec, 1999[a]	RCT	373 fifth-grade students in 15 classes from nine primary schools in Slovenia	One lesson per week for seven months/ Basic concepts, measure transformation, calculations, problem solving	Cooperative learning vs. Individual work	Teacher developed math test	Overall	0.317	(ns)	0.251

~ p < .10, * p < .05, ** p < .01, *** p < .001
[a] Data were adjusted for clustering that occurred within classrooms.
[b] Data were adjusted for clustering that occurred within schools.

Table 6: Studies That Investigated Multiple Strategies—Cooperative Learning Combined With Other Instructional Practices

Study	Design	Sample	Duration/Content	Contrast	Measure	Subgroup	Hedge's g	Standard Error	
Brenner et al., 1997[a]	RCT	128 students in six intact pre-algebra classes in three junior high schools in southern California	20 days/ Pre-algebra ideas such as the functional relationship between two variables and contextual translation and application	Anchored instruction using small groups vs. Control	Pooled problem solving outcomes: word problem solving (ES = 0.110), function word problem (ES = 0.393), and equation solving (ES = -0.281) measures	Overall	0.074	(ns)	0.399
Busato et al., 1995[a]	Quasi	572 middle school students in 23 classes in six schools in the Netherlands	Unspecified duration/ Existing Dutch math curriculum	AGO model curriculum vs. Traditional curricula	Test of math achievement	Boys	0.681	*	0.227
						Girls	0.583	~	0.235
Stevens & Slavin, 1995[c]	Quasi	1,012 elementary students in Grades 2–6 in five schools in a suburban Maryland school district. Two of the schools were cooperative elementary schools and three schools were more traditional elementary schools	1 year/ Math computation	Cooperative learning school vs. Whole class instruction	California Achievement Test—CAT computations	Overall	0.120	~	0.068
					California Achievement Test—CAT—application	Overall	-0.050	(ns)	0.068

~ p < .10, * p < .05, ** p < .01, *** p < .001

[a] Data were adjusted for clustering that occurred within classrooms.

[b] Data were adjusted for clustering that occurred within schools.

[c] To be more comparable with other studies, the data presented for this study are data only after the first year, although two years of data were available.

Table 7: Studies That Investigated Cooperative Learning Strategies in the Context of Computers

Study	Design	Sample	Duration/Content	Contrast	Measure	Subgroup	Hedge's g	Standard Error	
Hooper, 1992	RCT	115 fifth- and sixth-grade average or high-ability students from a suburban middle school	1 week/ Calculation of number of sides of a three-dimensional object and classifying objects as examples or nonexamples of a concept	Cooperative learning with computer based instruction (CBI) vs. Individual learning with CBI	Math test including fact, application, generalization, and problem-solving questions	Overall	0.335	~	0.201
Hooper, 1993	RCT	175 fourth-grade average or high-ability students from six classrooms in a suburban middle school	3 weeks/ Calculations using the four basic arithmetic operations using symbols to represent constants and operations	Cooperative learning with computer based instruction (CBI) vs. Individual learning with CBI	Math test including fact, application, generalization, and problem-solving questions	Overall	0.305	~	0.157
Mevarech et al., 1991	RCT	149 sixth-grade students in five classrooms in one Israeli school	One trimester/ Basic operations with positive integers and fractions	Pairs w/CAI vs. Individual w/CAI	TOAM achievement (computerized diagnostic)	Low achieving	0.266	(ns)	0.285
						Medium achieving	0.268	(ns)	0.285
						High achieving	0.120	(ns)	0.289
Mevarech, 1993[a]	RCT	110 third-grade students in two Israeli public schools	One trimester/ Traditional 3rd grade curriculum in arithmetic	Pairs w/CAI vs. Individual w/CAI	Arithmetic achievement test	Low achieving	-0.028	(ns)	0.479
						High achieving	-0.506	(ns)	0.480
Mevarech, 1994[a]	RCT	344 third-grade and 279 sixth-grade students in 19 classrooms in five schools in a suburb of Tel Aviv, Israel	Two 20-minute sessions per week for one academic year/ Basic skills with mathematics operations, comprehension of numerical systems, understanding mathematical rules, solving word problems	Integrated learning system (ILS) in homogenous pairs vs. ILS individually	ILS diagnostic test	3rd grade Low Achievers	0.287	(ns)	0.322
						3rd grade High Achievers	0.123	(ns)	0.319
						6th grade Low Achievers	0.266	(ns)	0.337
						6th grade High Achievers	0.348	(ns)	0.336
Slavin & Karweit, 1984[a]	RCT	588 ninth-grade students in 25 math classes in 16 inner city Philadelphia junior and senior high schools	One school year/ Unspecified 9th grade general math curriculum	STAD vs. Focused Instruction (students worked individually and did not receive team recognition)	CTBS, Shortened version (computation, concepts and applications subscales)	Overall	0.113	(ns)	0.221

Continued on p. 6-40

Table 7, continued

Study	Design	Sample	Duration/Content	Contrast	Measure	Subgroup	Hedge's g	Standard Error	
Weiss et al., 2006	RCT	116 students in six Kindergarten classes of medium to high SES in Israel	28 hours over 5 months/ Mathematical skills about numbers and operations from 1 to 10	Multimedia environment in a cooperative learning teaching style vs. Multimedia environment in an individual learning teaching style	Skills test on numbers and operations	Overall	-0.862	***	0.238
Xin, 1999	RCT	118 3rd-grade students in six mathematics classes in three schools	Daily for one semester/ Basic fact families including addition, subtraction, multiplication, and division; coin recognition, place value, concepts, number patterns	TAI w/ CAI vs. whole class w/ CAI	Stanford Achievement Test-Math	Regular education	0.595	**	0.210
						Learning disability	0.338	(ns)	0.390
Heterogeneity									
Q-value	*df (Q)*	*P-value*	*I-squared*	*Pooled ES (8 studies, 15 effect sizes)*			*Hedge's g*	*Standard Error*	
27.457	*14*	*0.017*	*49.011*				*0.157*	*0.101*	

~ p < .10, * p < .05, ** p < .01, *** p < .001
[a] Data were adjusted for clustering that occurred within classrooms.
[b] Data were adjusted for clustering that occurred within schools.

BIBLIOGRAPHY

Alkhateeb, H.M., & Jumaa, M. (2002). Cooperative learning and algebra performance of eighth-grade students in United Arab Emirates. *Psychological Reports, 90,* 91–100.

Barron, B. (2000). Problem solving in video-based microworlds: Collaborative and individual outcomes of high-achieving sixth-grade students. *Journal of Educational Psychology, 92*(32), 391–398.

Beady, C.L., Slavin, R.E., & Fennessey, G.M. (1981). Alternative student evaluation structures and a focused schedule of instruction in an inner-city junior high school. *Journal of Educational Psychology*, *73,* 518–523.

Becker, W.C. (1977). Teaching reading and language to the disadvantaged—What we have learned from field research. *Harvard Educational Review*, *47*, 18–543.

Bell, A.W., O'Brien, D., & Shiu, C. (1980). Designing teaching in the light of research on understanding. In R. Karplus (Ed.), *Proceedings of the fourth international conference for the psychology of mathematics education*, Berkeley, CA: The International Group for the Psychology of Mathematics. (ERIC Document Reproduction Service No. ED 250 186).

Bell, A.W. (1982). Diagnosing students' misconceptions. *The Australian Mathematics Teacher, 1*, 6–10.

Brenner, M., Mayer, R., Moseley, B., Brar, T., Duran, R., Reed, B.S., & Webb, D. (1997). Learning by understanding: The role of multiple representations in learning algebra. *American Educational Research Journal, 34(4*), 663–689.

Brophy, J.E., & Good, T.L. (1986). Teacher behavior and student achievement. In M.C. Wittrock (Ed.), *Handbook of research on teaching*. New York: Macmillan.

Busato, V., Tendam, G., Vandeneeden, P., & Terwel, J. (1995). Gender-related effects of cooperative learning in a mathematics curriculum for 12–16 year olds. *Journal of Curriculum Studies, 27*(*6*), 667–686.

Cicchelli, T. (1982). Effects of direct and indirect instruction patterns and prior achievement on post course achievement. *Journal of Instructional Psychology, 9(4),* 176–189.

Cobb, P. (2007). Putting philosophy to work: Coping with multiple theoretical perspectives. In F. Lester (Ed.), *Second handbook of research on mathematics teaching and learning (Vol. 1, pp. 3–38).* Greenwich, CT: Information Age Publishing.

Dewey, J. (1897). My pedagogic creed. *School Journal, 54,* 77–80.

Flanders, N.A. (1970). *Analyzing teaching behavior*. Reading, MA: Addison-Wesley Publishing Company.

Fuchs, L.S., Fuchs, D., Finelli, R., Courey, S.J., Hamlett, C.L., Sones, E.M., & Hope, S.K. (2006). Teaching third-graders about real-life mathematical problem solving: A randomized controlled study. *The Elementary School Journal, 106(4),* 293–311.

Fuchs, L.S., Fuchs, D., Hamlett, C.L., & Appleton, A.C. (2002). Explicitly teaching for transfer: Effects on the mathematical problem-solving performance of students with mathematics disabilities. *Learning Disabilities Research & Practice, 17*(2), 90–106.

Fuchs, L.S., Fuchs, D., Hamlett, C.L., Phillips, N.B., Karns, K., & Dutka, S. (1997). Enhancing students' helping behavior during peer-mediated instruction with conceptual mathematical explanations. *The Elementary School Journal*, *97(3)*, 223–249.

Fuchs, L., Fuchs, D., & Karns, K. (2001). Enhancing kindergarteners' mathematical development: Effects of peer-assisted learning strategies. *Elementary School Journal*, *101(5)*, 495–510.

Fuchs, L., Fuchs, D. Phillips, N., Hamlett, C., & Karns, K. (1995). Acquisition and transfer effects of classwide peer-assisted learning strategies in mathematics for students with varying learning histories. *School Psychology Review*, *24*(4), 604–620.

Fuchs, L., Fuchs, D., Yazdian, L., & Powell, S. (2002). Enhancing first-grade children's mathematical development with peer-assisted learning strategies. *School Psychology Review*, *31(4)*, 569–583.

Gersten, R., & Carnine, D. (1984). Direct instruction mathematics: A longitudinal evaluation of low-income elementary school students. *Elementary School Journal*, *84(4)*, 395–407.

Ginsburg-Block, M. & Fantuzzo, J. (1998). An evaluation of the relative effectiveness of NCTM standards-based interventions for low-achieving urban elementary students. *Journal of Educational Psychology, 90(3)*, 560–569.

Good, T.L., & Grouws, D.A. (1977). Teaching effects: A process-product study in fourth-grade mathematics classrooms. *Journal of Teacher Education, 28(3)*, 49–54.

Good, T.L., & Grouws, D.A. (1979). The Missouri mathematics effectiveness project. *Journal of Educational Psychology, 71(3)*, 355–362.

Good, T., Grouws, D. & Ebmeier, M. (1983). Active mathematics teaching. New York: Longman.

Good, T.L., Mulryan, C., & McCaslin, M. (1992). Grouping for instruction in mathematics: A call for programmatic research on small-group processes. In D. A. Grouws (Ed), *Handbook of research on mathematics teaching and learning*. New York: Macmillan.

Hooper, S. (1992). Effects of peer interaction during computer-based mathematics instruction. *Journal of Educational Research, 85*, 180–189.

Hooper, S., Temiyakarn, C., & Williams, M.D. (1993). The effects of cooperative learning and learner control on high and average ability students. *Educational Technology Research and Development, 41(2)*, 5–18.

Hopkins, K.B., McGillicuddy-De Lisi, A.V., & De Lisi, R. (1997). Student gender and teaching methods as sources of variability in children's computational arithmetic performance. *The Journal of Genetic Psychology, 158(3)*, 333–345.

Hurley, E.A., Boykin, A.W., & Allen, B.A. (2005). Communal versus individual learning of a math-estimation task: African American children and the culture of learning contexts. *Journal of Psychology, 139(6)*, 513–527.

Jacobs, J.A. (1996). Gender inequality and higher education. *Annual Review of Sociology, 22(1)*, 153–185.

Janicki, T.C. & Peterson, P.L. (1981). Aptitude-treatment interaction effects of variations in direct instruction. A*merican Educational Research Journal, 18(1)*, 63–82.

Kame'enui, E., Carnine, D., Darch, C., & Stein, M. (1986). Two approaches to the development phase mathematics instruction. *Elementary School Journal, 86*, 633–650.

Kramarski, B. & Mevarech, Z.R. (2003). Enhancing mathematical reasoning in the classroom: The effects of cooperative learning and metacognitive training. *American Educational Research Journal, 40(1)*, 281–310.

Ladson-Billings, G. (1995). Toward a theory of culturally relevant pedagogy. *American Education Research Journal, 35*, 465–491.

Lampert, M. (1990). When the problem is not the question and the solution is not the answer: Mathematical knowing and teaching. *American Educational Research Journal, 27*, 29–63.

Leron, U., & Lavy, I. (2004). The emergence of mathematical collaboration in an interactive computer environment. *International Journal of Computers for Mathematical Learning, 9*, 1–23.

Light, P., & Blaye, A. (1990). Computer-based learning: The social dimensions. In H. C. Foot, M. J. Morgan & R. H. Shute (Eds.), *Children helping children* (pp. 135–147). New York: John Wiley & Sons.

Madden, N.A., & Slavin, R.E. (1983). Effects of cooperative learning on the social acceptance of mainstreamed academically handicapped students. *Journal of Special Education, 17(2)*, 171–182.

Mevarech, Z.R. (1993). Who benefits from cooperative computer-assisted instruction? *Journal of Educational and Computing Research. 9*(4), 451–461.

Mevarech, Z.R. (1994). The effectiveness of individualized versus cooperative computer-based integrated learning systems. *International Journal of Educational Research*. *86* (4), 39–52.

Mevarech, Z.R., Silber, O., & Fine, D. (1991). Learning with computers in small groups: Cognitive and affective outcomes. *Journal of Educational and Computing Research. 7*(2), 233–243.

Meyer, L.A., Gersten, R., & Gutkin, J. (1983). Direct instruction: A project follow through success story in an inner-city school. *The Elementary School Journal, 84*, 241–252.

Muthukrishna, N., & Borkowski, J.G. (1995). How learning contexts facilitate strategy transfer. *Applied Cognitive Psychology, 9*, 425–446.

Nastasi, B.K., & Clements, D.H. (1994). Effectance motivation, perceived scholastic competence, and higher-order thinking in two cooperative computer environments. *Journal of Educational Computing Research, 10(3)*, 249–275.

National Council of Teachers of Mathematics (NCTM) (1989). *Curriculum and evaluation standards for school mathematics.* Reston, VA: National Council of Teachers of Mathematics.

National Research Council (1989). *Everybody counts: A report to the nation on the future of mathematics education.* Washington, DC: National Academy Press.

National Research Council (2000). *How people learn: Brain, mind, experience, and school.* Washington, DC: National Academy Press.

National Research Council. (2001). Adding it up: Helping children learn mathematics. In J. Kilpatrick, J. Swafford, & B. Findell (Eds.), *Mathematics learning study committee, center for education, division of behavioral and social sciences and education*. Washington, DC: National Academy Press.

Nickerson, R.S. (1988). On improving thinking through instruction. In E.Z. Rothkopf (Ed.), *Review of research in education, 15,* 3–57. Washington, DC: American Educational Research Association.

Palincsar, A.S., & Brown, A.L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction, 1(2)*, 117–175.

Peklaj, C., & Vodopivec, B. (1999). Effects of cooperative versus individualistic learning on cognitive, affective, metacognitive and social processes in students. *European Journal of Psychology of Education, 14(3)*, 359–373.

Rittle-Johnson B. (2006). Promoting transfer: Effects of self-explanation and direct instruction. *Child Development, 77(1)*, 1–15.

Rudnitsky, A., Etheredge, S., Freeman, S.J.M., & Gilbert, T. (1995). Learning to solve addition and subtraction word problems through a structure-plus-writing approach. *Journal for Research in Mathematics Education, 26(5)*, 467–486.

Scardamalia, M., Bereiter, C., Brett, C., Burtis, P.J., Calhoun, C., & Lea, N.S. (1992). Educational applications of a networked communal database. *Interactive Learning Environments, 2(1)*, 45–71.

Schofield, J.W. (1995). *Computers and classroom culture*. Cambridge, MA: Cambridge University Press.

Slavin, R.E. (1980). Effects of student teams and peer tutoring on academic achievement and time on-task. *Journal of Experimental Education, 48*, 252–257.

Slavin, R.E. (1991). Synthesis of research on cooperative learning. *Educational Leadership, 48*, 71–82.

Slavin, R.E., & Karweit, N.L. (1984). Mastery learning and student teams: A factorial experiment in urban general mathematics classes. *American Educational Research Journal, 21(4)*, 725–736.

Slavin, R.E. & Karweit, N.L. (1985). Effects of whole class, ability grouped, and individualized instruction on mathematics achievement. *American Educational Research Journal, 22 (3)*, 351–367.

Slavin, R., Leavey, M., & Madden, N.A. (1984). Combining cooperative learning and individualized instruction: Effects on student mathematics achievement, attitudes, and behaviors. *The Elementary School Journal, 84(4)*, 409–422.

Slavin, R.E., Madden, N.A., & Leavey, M. (1984). Effects of Team Assisted Individualization on the mathematics achievement of academically handicapped and non-handicapped students. *Journal of Educational Psychology, 76 (5)*, 813–819.

Stallings, J., (1975). Implementation and child effects of teaching practices in follow through classrooms. *Monographs of the Society for Research in Child Development, 40(7/8)*, 1–133.

Stebbins, L., St. Pierre, R.G., Proper, E.L., Anderson, R.B., & Cerva, T.R. (1977). *Education as experimentation: A planned variation model (Vol. IV A–D).* Cambridge, MA: Abt Associates.

Stevens, R.J., & Slavin, Robert E. (1995). The cooperative elementary school: Effects on students' achievement, attitudes, and social relations. *American Educational Research Journal*, *32(2)*, 321–351.

Strommen, E.F. (1993). Does yours eat leaves? Cooperative learning in an educational software task. *Journal of Computing in Childhood Education, 4(1)*, 45–56.

U.S. Department of Labor, Bureau of Labor Statistics. (2008). Teachers—Preschool, kindergarten, elementary, middle, and secondary. In *Occupational Outlook Handbook.* Retrieved on December 1, 2007, from http://www.bls.gov/oco/oco069.htm#nature.

Weiss, I., Kramarski, B., & Talis, S. (2006). Effects of multimedia environments on kindergarten children's mathematical achievements and style of learning. *Educational Media International*, *43(1)*, 3–17.

Xin, J.F. (1999). Computer-assisted cooperative learning in integrated classrooms for students with and without disabilities. *Information Technology in Childhood Education Annual*, *1(1)*, 61–78.

III. Effective Instruction for Students With Learning Challenges: A Meta-Analytic Review

Between 5% and 10% of students will experience a serious learning disability in mathematics before completing high school (Barbaresi, Katusic, Colligan, Weaver, & Jacobsen, 2005). Many more will have difficulties in learning mathematics at an acceptable level of proficiency. In this section, the Instructional Practices Task Group addresses the rigorous research on instructional methods that can help these students. An overview of the methodological procedures for the Task Group is provided in Appendix A. Throughout this section, the Task Group used these meta-analytic techniques as noted in the methodology statement. Because of the wide array of instructional approaches explored in the research for this section, multiple meta-analyses were performed to analyze this research.

The Task Group chose to review studies of students with LD separately from studies of low-achieving students because the problems experienced by students with LD are consistently more severe than those experienced by other low-performing students (Fuchs, Fuchs, Mathes, & Lipsey, 2000; Murphy, Mazzocco, Hanich, & Early, 2007). Therefore, educators cannot necessarily assume that techniques that are effective for students with learning disabilities are the most effective or efficient means for teaching struggling students. However, the reader will note that many of the same themes and issues recur across these two bodies of research.

A. Characteristics of Students With Learning Disabilities in Mathematics

Most of the research on the nature of learning disabilities in mathematics has been conducted with younger students and typically involves understanding their gaps in whole number arithmetic. Certain findings have been consistently replicated. Because students with LD display problems in so many areas of mathematics, pinpointing the exact nature of the cognitive difficulty has been an intricate process (Geary, 2003).

However, there are several problems that seem particularly chronic. The first is efficient retrieval of basic arithmetic combinations (mathematics facts) (Jordan, Hanich & Kaplan, 2003). A second is delayed adoption of efficient counting strategies. Students with learning disabilities will tend to count on their fingers well after their peers have outgrown this approach and when forbidden by their teachers they may count with the help of visual placemarkers in the classroom (e.g., stripes on the ceiling or the radiator), or give up in frustration. Most typically developing students learn, prior to entering school, what is commonly called a “counting-on strategy.” They learn that if they have to add 7 to 2, this process is mathematically equivalent to adding 2 to 7, and that is much more efficient to make this transformation (i.e., that the most efficient way to find 2 + 7 is to start with the 7 in a mental number line and count up 2, rather than start with the 2 and count up 7). In contrast, students with LD will tend to start at 2 and count up using 7 figures or objects. Thus, they are more likely to make errors by using the tedious procedure. Furthermore, even if their answer is accurate, their strategy for reaching this answer is far from efficient.

It also appears that students with learning disabilities have a very limited working memory (Geary, Hoard, Byrd-Craven, Nugent, & Numtee, 2007; McLean & Hitch, 1999; Swanson & Sachse-Lee, 2001), which affects their ability to keep abstract information in their minds for the purpose of counting and solving specific problems. Finally, students with learning disabilities seem to display problems in many aspects of basic number sense such as comparing magnitudes of numbers by quickly visualizing a number line or transforming simple word problems into simple equations (Jordan, Hanich, & Kaplan, 2003; Fuchs, Compton, Fuchs, Paulsen, Bryant, & Hamlett, 2005). In addition, two studies (DiPerna, Lei & Reid, 2007; Fuchs, 2005) have both found that teachers' ratings of a child's attention span and task persistence, both areas that are often difficult for students with LD, are good indicators of subsequent problems in learning mathematics.

B. Students With Low Achievement in Mathematics

Many more students struggle to learn mathematics than the 5 to 10% who appear to possess a learning disability in mathematics (Badian, 1983; Fuchs et al., 2005; Gross-Tsur et al., 1996; Lewis et al., 1994). Although there are numerous disputes about how to best define and operationalize the general term "learning disabilities," and the more specific term "math disabilities," there is some emerging consensus (Bradley, Danielson, & Hallahan, 2002).

In contrast, there is no consensus as to how to operationalize the term "low achieving," other than students whose performance in mathematics is below grade level expectations. In some cases, e.g., Cardelle-Elawar (1995), all students in a low-income, low-achieving school are considered low achieving. Other studies (e.g., Moore & Carnine, 1989) only select students who perform poorly on a screening test that addresses the topic of the intervention research study.

Several dilemmas and constraints presented themselves when considering studies of low-achieving students not presented in working with studies of students with LD. First, there is no agreed upon operational definition of what is meant by a student struggling to learn mathematics or a low-achieving or "at risk" student (Mazzocco, 2007). Indeed, there is no measurable boundary or cut-off criterion, based on standardized test performance, for considering a student to be math disabled versus experiencing low achievement in mathematics.

Meanwhile, the factors that contribute to the low-achieving designation seem to include some unknown combination of the following:

- deficiencies with previous mathematics instruction and mathematics teachers with limited knowledge of the subject (Sowder, Philipp, Armstrong, & Schappelle, 1998);
- limited experiences at home that informally teach familiarity with number concepts, build and reinforce procedural facility and demonstrate relevance of mathematics to everyday problems (e.g., Griffin, Case, & Siegler, 1994);
- problems with sustaining attention to academic tasks and activities (Fuchs, Compton, Fuchs, Paulson, Bryant, & Hamlett, 2005; DiPerna, Lei, & Reid, 2007; Kolligian & Sternberg, 1987); and,
- weak motivation and maladaptive attribution style (Torgesen, 1994).

C. A Meta-Analytic Review of Research With Students With LD and LA in Mathematics (1976–2007)

There is a dramatically smaller body of research on mathematics instruction compared to reading instruction for students with LD. A recent review of the ERIC literature base (Gersten, Clarke, & Mazzocco, 2007) found that the ratio of studies on reading disabilities to mathematics disabilities and difficulties was 5:1 for the decade 1996–2005. However, this was a dramatic improvement over the ratio of 16:1 in the prior decade.

Despite the limited knowledge of the precise nature of learning disabilities in mathematics, especially in areas such as rational numbers, geometry and pre-algebra, researchers have attempted to develop interventions that can teach students with LD. In fact, in the Panel's literature search, the number of high-quality studies examining the effectiveness of various instructional practices for teaching students with LD far surpasses the number of studies conducted with typically developing students.

The Task Group speculates that there are several reasons for this phenomenon. One important factor has been the consistent support for research in the field of special education; annual research budgets for special education often surpassed budgets for research on the education of nondisabled students in academic areas. Even more importantly, the Office of Special Education Programs (OSEP) at the U.S. Department of Education has consistently supported experimental research since the late 1960s.

In addition to studies of students with LD, the Task Group was also able to locate a small number of studies with low-achieving students that met the criteria for rigorous experimental or quasi-experimental research. The children in these studies were defined as either at-risk or experiencing mathematics difficulties based on performance on a screening measure of mathematics skills or teacher ratings or recommendation. None of the participants was formally diagnosed as math disabled by the researchers.

Thus, a reasonable set of studies exists that investigate the effectiveness of various instructional approaches for teaching students with LD using rigorous experimental or quasi-experimental designs, and a smaller, but still adequate set of studies exists that examine various approaches for teaching students who experience difficulties in mathematics, but were not classified as possessing a learning disability. To conduct a meta-analysis, the Task Group needed to center the analysis around a key research question. The question most consistently posed in the research studies reviewed concerned the effectiveness of explicit systematic instruction on the mathematics performance of this group of students.

D. The Nature of This Report

This document summarizes the meta-analyses conducted for the Panel's Instructional Practices Task Group on the nature of effective mathematics instruction for students with LD and for other low-achieving students. To organize these meta-analyses, the Task Group clustered studies into four categories:

- Studies of the impact of systematic *explicit instruction* on the performance of students with LD in mathematics

- Studies of the impact of systematic *explicit instruction* on the performance of low-achieving students in mathematics

- Other approaches for teaching students with LD
  - — Selection of examples to foster development of more sophisticated strategies for quick retrieval of basic arithmetic facts
  - — Use of visual representations as a key component of instruction
  - — Instruction that encouraged students to think aloud

- Other approaches for teaching low-achieving students that are primarily implicit

The following sections provide study characteristics for each of the studies identified and effect sizes for the Category 1 (high-quality) studies on posttest measures and transfer measures (when available). The effect sizes are pooled for the studies that examined the effects of using explicit instruction for students with LD using common meta-analytic standards. Effect sizes for studies in the remaining categories were not pooled because the interventions varied greatly across studies.

All effect sizes have been adjusted for clustering, when appropriate. The Task Group used the U.S. Department of Education's What Works Clearinghouse default Intra-Class Correlation of .20 for the adjustment. For further details on data analysis, see the footnotes accompanying the tables and the Methodological Procedures section in Appendix A.

E. Explicit Strategies Used for Students With Learning Disabilities

Explicit instruction involves teacher-demonstrated step-by-step plans for solving a problem. The teacher demonstrates a specific plan for a set of problems (as opposed to a general problem-solving heuristic strategy) and students are asked to use the same procedures or steps demonstrated by the teacher to solve the problem. For example, Xin, Jitendra, and Deatline-Buchman (2005) provided explicit instruction for using strategies for identifying and solving various word problem types. Students were given prompt sheets for identifying salient features of the word problem types. Students then were taught to map these features onto a schema diagram that represented the problem structure. Next, students used the schema diagram to formulate the appropriate mathematical equation for solving the problem. Each of these steps toward problem solution were explicitly taught strategies for problem solution.

There were nine studies that looked at the effect of explicit strategies for students with LD, met the inclusion criteria, and were methodologically adequate. Results from these studies are presented in three tables. Each table includes the findings for a separate mathematical outcome: word problem solving, computation, or transfer of learning. Table 8 below presents the results from the six studies that investigated the effects of using explicit strategies on improving word problem-solving outcomes with students with learning disabilities. Table 9 presents results of the three high-quality studies that looked at computation outcomes. Table 10 presents the results of the four studies that included a measure of generalization of training. All tables include the pooled effect size, tests for heterogeneity, and tests for statistical significance for the pooled effect size. The pooled effects reveal significant effects of explicit instruction on solving word problem solving (ES = 1.152), computation (ES = 1.285), and transfer (ES = 0.777).

Table 8: Studies That Investigate Explicit Strategies With Students With Learning Disabilities: Word Problem Outcomes

Study	Design	Sample	Duration/Content	Contrast	Measure	Hedge's g		Standard Error
Word Problem Outcomes								
Hutchinson, 1993	RCT	20 LD students in Grades 8–10 from two junior high schools in suburban Vancouver, Canada	40-minute sessions on alternate days for 4 months/ Algebraic word problem solving (relational problems, proportion problems, & two-variable, two-equation problems)	Metacognitive and solution strategies vs. Regular resource class instruction	Pooled B.C. Mathematics Achievement Test (ES = 0.705), Q2 B.C. Achievement Test (ES = 1.724)	1.215	*	0.484
Jitendra et al., 1998	RCT	34 students in Grades 2–5 from four public-schools in the U.S. 25 students were classified as having mild disabilities (LD, educable mentally retarded, or seriously emotionally disturbed), and the remaining nine students had difficulty in math	17–20 40–45-minute sessions/ Addition and subtraction word problems (including change, group, and compare problems)	Explicit step-by-step strategy vs. Traditional basal strategy	Researcher designed word problem solving criterion test	0.557	(ns)	0.342
Milo et al., 2005[a]	Quasi	36 LD students in three special primary schools. Average age of students was 9.10 years	Two weekly lessons for half the year during regular math class/ Addition and subtraction	Directing vs. Guiding instruction	Addition and subtraction word problem test based on problems from the databank of the National Institute for Educational Measurement	0.303	(ns)	0.447
Owen & Fuchs, 2002[a]	Quasi	24 third-grade students with IEPs from 14 classrooms in six schools (20 students had LD, one had MMR, two had speech disorders, and one had ADHD)	Six lessons/ Word problems that involve finding "halves"	Full-dose acquisition and transfer vs. Traditional	Researcher designed word problem solving test	3.385	***	0.888
Wilson & Sindelar, 1991[b]	RCT	62 LD students from nine elementary schools in a medium-sized school district in northern Florida	Fourteen 30-min lessons over 3 weeks/ Addition and subtraction word problems (four types of two- to three-sentence problems)	Strategy plus sequence vs. Sequence only	Researcher designed word problem test	0.782	~	0.470
Xin et al., 2005	RCT	22 middle school students in one school (18 students had LD, one had severe emotional disorders, and three were at-risk for math failure)	12 1-hour sessions/ Multiplicative compare and proportion problems and mixed word problems	Schema-based instruction (SBI) vs. General strategy instruction (GSI)	Researcher designed word problem solving criterion test	1.866	***	0.497
Heterogeneity								
Q-value	*df (Q)*	*P-value*	*I-squared*	*Pooled ES (five studies, five effect sizes)*		*Hedge's g*	*Standard Error*	
14.754	*5*	*0.011*	*66.110*			*1.152*	***	*0.341*

~ p < .10, * p < .05, ** p < .01, *** p < .001

[a] Data were adjusted for clustering that occurred within classrooms.

[b] Data were adjusted for clustering that occurred within schools.

Table 9: Studies That Investigate Explicit Strategies With Students With Learning Disabilities: Computation Outcomes

Study	Design	Sample	Duration/Content	Contrast	Measure	Hedge's g	Standard Error	
Computation Outcomes								
Schopman & Van Luit, 1996	Quasi	60 students between the ages of 5 and 7 attending schools for special education (primarily LD) who scored less than 45% correct on a test for number sense (likely in the Netherlands)	Thirteen lessons in 3 months/ Preparatory arithmetic skills (number sense, counting skills, and Piagetian operations)	Directing and guiding vs. Control	Utrech test of number sense	1.023	***	0.286
Tournaki, 2003	RCT	42 LD second-grade students attending self contained special education classes in one school in New York	Eight 15-minute sessions on consecutive school days/ Algebra (three problem types: relational problems, proportion problems, & two-variable, two-equation problems)	Strategy instruction vs. Drill and practice	Researcher designed computation test	1.612	***	0.426
Van Luit & Naglieri, 1999[a]	Quasi	42 9–11-year-old LD students from two schools for special education in the Netherlands	Three 45-minute sessions per week for 17 weeks/ Multiplication and division problems	MASTER program (Mathematics Strategy Training for Educational Remediation) vs. Standard instruction	Researcher designed mathematics achievement test	2.174	*	0.991
Heterogeneity								
Q-value	*df (Q)*	*P-value*	*I-squared*	*Pooled ES (four studies, four effect sizes)*		*Hedge's g*	*Standard Error*	
2.221	*2*	*0.329*	*9.956*			*1.285*	***	*0.256*

~ p < .10, * p < .05, ** p < .01, *** p < .001
[a] Data were adjusted for clustering that occurred within classrooms.

Table 10: Studies That Investigate Explicit Strategies With Students With Learning Disabilities: Transfer Outcomes

Study	Design	Sample	Duration/ Content	Contrast	Measure	Hedge's g	Standard Error	
Transfer Outcomes								
Jitendra et al., 1998	RCT	34 students in Grades 2–5 from four public-schools in the U.S. 25 students were classified as having mild disabilities (LD, educable mentally retarded, or seriously emotionally disturbed), and the remaining nine students had difficulty in math	17–20 40–45-minute sessions/ Addition and subtraction word problems (including change, group, and compare problems)	Explicit step-by-step strategy vs. Traditional basal strategy	Researcher designed generalization test	1.010	**	0.357
Milo et al., 2006[a]	Quasi	36 LD students in three special primary schools. Average age of students was 9.10 years	Two weekly lessons for half the year during regular math class/ Addition and subtraction	Directing vs. Guiding instruction	Researcher designed transfer test	-0.073	(ns)	0.445
Tournaki, 2003	RCT	42 LD 2nd-grade students attending self contained special education classes in one school in New York	Eight 15-minute sessions on consecutive school days/ Algebra (3 problem types: relational problems, proportion problems, & two-variable, two-equation problems)	Strategy instruction vs. Drill and practice	Researcher designed transfer test	0.801	*	0.382
Xin et al., 2005	RCT	22 middle school students in one school (18 students had LD, one had severe emotional disorders, and three were at-risk for math failure)	3–4 times per week, for a total of 12 1-hour sessions/ Multiplicative compare and proportion problems and mixed word problems	Schema-based instruction (SBI) vs. General strategy instruction (GSI)	Researcher designed generalization test	1.334	**	0.467
Heterogeneity								
Q-value	*df (Q)*	*P-value*	*I-squared*	*Pooled ES (four studies, four effect sizes)*		*Hedge's g*	*Standard Error*	
5.496	*3*	*0.139*	*45.416*			*0.777*	**	*0.277*

~ p < .10, * p < .05, ** p < .01, *** p < .001
[a] Data were adjusted for clustering that occurred within classrooms.

From these results, one can infer that explicit instruction is an effective means for building performance in problem solving, computational proficiency, and ability to transfer from items on which students received training to items on which students had not received training, for students with LD.

However, the number of high-quality studies is small, and one would not want to overgeneralize from a set of nine studies that, taken together, are limited by a restricted range of study characteristics. For example, many of the studies were of short or moderate duration. Although the set of studies represents a wide range of age levels (seven of the studies examine elementary schools, while two studies examined middle schools) there are a sparse number of studies for any given age level or any given mathematical topic. The studies reviewed almost exclusively used researcher-developed measures, which tend to yield higher effect sizes than norm-referenced measures of generalized mathematics proficiency (Swanson & Hoskyn, 1998).

1. The Evolving Nature of Explicit Systematic Strategy Instruction

Nonetheless, the positive and significant pooled effect sizes for the studies that investigate the effect of explicit systematic instruction study results on word problems, computation, and transfer outcomes suggests that explicit systematic instruction is a desirable approach for at least some critical aspects of mathematics instruction for students with LD. The question becomes, what exactly is explicit systematic instruction? There is no easy answer to this question. In fact, like most educational labels, this term means very different things to different individuals. In addition, the nature of explicit systematic instruction has evolved over time.

Probably the earliest use of this term (at least during the past four decades) was the pioneering work of Bereiter and Engelmann (1966) in providing preschoolers from low-income families with explicit systematic instruction in number concepts, counting, phonological awareness, and the more formal structure of the language used in school. By the 1980s, the external evaluation of Project Follow Through documented the success of this approach for teaching low-income students in the primary grades, particularly in the area of mathematics (Stebbins, St. Pierre, Proper, Anderson, & Cerva, 1977; Gersten & Carnine, 1984). As a result, many advocated the use of this approach, called direct instruction, in teaching mathematics to students with LD (e.g., Hallahan & Kauffman, 1986). In a 1998 meta-analysis, Swanson and Hoskyn (1998) concluded that the combination of direct instruction and strategy instruction was an effective approach for teaching students with LD in all academic areas.

In the 1980s, direct instruction approaches began to incorporate principles gained from cognitive psychology and were increasingly referred to by the terms explicit instruction or explicit strategy instruction. In some cases, strategies were rather broad heuristics meant to teach students how to approach any type of mathematical problem (e.g., Montague, 1992). In other cases, the approach was heavily scripted and detailed precise steps students should take to solve a particular problem type. This latter approach has been criticized for its failure to help students understand underlying concepts and build flexible thinking (e.g., flexible use of a mental number line for estimation, and fluency with number properties such as the commutative and distributive laws; Woodward & Montague, 2002). However, others have argued that high degrees of explicitness and highly systematic instruction are critical for students with LD (e.g., Owen & Fuchs, 2002; Jitendra et al., 1998).

Although the nature of explicit strategic instruction has evolved over time and can vary widely from study to study, there are a number of common features that define this approach. Generally, clear consistent modeling of step-by-step strategies through teacher explanation, modeling and demonstration; careful control of task difficulty; planful sequencing of teaching and practice examples; and specified procedures for providing corrective feedback characterize explicit systematic instruction. The studies reviewed here represent a range of approaches to providing explicit systematic instruction. However, all of them include most of the instructional features described above. In addition, this set of studies also demonstrates how explicit instruction has evolved over time to incorporate more innovative instructional features that support and encourage student interaction, flexibility, and generalization.

In Owen and Fuchs' (2002) research on teaching problems involving fraction concepts, students were shown transfer problems in a careful sequence. Transfer problems referred to those with extraneous information, differing terminology from the practice items (e.g., the terms "one third" and "divide equally into three pieces"), and multistep problems that included one step involving manipulation of fractions. Students received clear feedback on their attempts to apply what they had learned in their practice sets onto the broader spectrum of problems and when they experienced problems, teachers demonstrated the underlying similarities to the previously taught problems.

Van Luit and colleagues (1999) have developed a line of research for teaching students with LD that attempts to synthesize principles of explicit strategy instruction with advances in the understanding of the underlying nature of LD in mathematics (e.g., Geary, 2005; Brown & Campione, 1990; Fuchs & Fuchs, 1998). The approach differs from earlier versions of direct instruction in several important ways. As in traditional models of direct or explicit instruction, students are taught in a quite explicit fashion one problem solving strategy at a time. As with Owen and Fuchs (2002) and Jitendra et al. (1998), teachers explicitly present a series of problem-solving steps to students and model several problems of this type for a small group of students. Students are taught multiple problem solving strategies and practice with an array of problems that use different types of syntax and different types of situations. Teachers actively encourage students to think aloud, to either walk through the steps in their strategy or articulate a reason for their decision to, for example, divide rather than multiply. Most of the intensive instruction is conducted in small groups. Teachers in Van Luit and Naglieri (1999), Jitendra et al. (1998), and Owen and Fuchs (2002) also used visual representations to teach problem solving.

Tournaki (2003) used explicit instruction to help students with LD learn more sophisticated counting strategies. She capitalized on the important insight made by Siegler (1987) that a key milestone in beginning mathematics proficiency for children is the insight that to most effectively solve a simple addition problem, it is invariably easier to start counting from the larger number, rather than the first number. For 3 + 8, it is far more efficient to count 3 up from 8 than to begin with the 3 and "count up" 8. This insight requires students to have some grasp of the commutative law and also a reasonable sense of magnitude comparison—two essential components of number sense.

The goal of this study was to examine whether the counting on strategy could be successfully taught to elementary students with LD. Instruction was quite clear and explicit. An example of the Strategy Instruction condition from Tournaki (2003, p. 458) is as follows:

> *Q (Teacher): When I get a problem, what do I do?*
> *A (Desired student response, i.e., repeat of the rule): I read the problem: 5 plus 3 equals how many. Then I find the smaller number.*
>
> *Teacher points to the smaller number and says, 3. Now I count the fingers.*
> *Q (Teacher): So how many fingers am I going to count?*
> *A (Desired student response): 3.*

After a few problems, the teacher had students solve problems while thinking aloud, i.e., repeating the steps and asking themselves the questions. Teachers always provided clear, immediate feedback when students made errors.

Note how closely this approach aligns to the depiction of explicit instruction presented earlier. Yet note how the target goal is to intentionally propel students into use of a more sophisticated counting strategy than just adding two numbers together, based on the finding from cognitive psychology (Siegler & Shrager, 1984; Geary, 1993) that students with LD tend to solve a problem such as 3 + 8 by starting at 3 and counting "up" 8 objects, whereas nondisabled students quickly learn that since 3 + 8 is the same as 8 + 3, it is much more efficient to start with 8 and count up 3 more objects.

An interesting pattern emerges in the research of Tournaki (2003) on explicitly teaching students to use the counting on strategy. There is a significant impact on the immediate computation posttest (ES = 1.612). In other words, students with LD do better when taught a strategy than when they are simply given a set of addition problems and told to do them as fast as they can. However, the significant effect measured by the transfer test (ES = 0.801) indicates that strategy-based approaches that teach students about number families and number bonds pay dividends in terms of other important areas of mathematics such as estimation.

2. Contemporary Adjustments to Explicit Strategy Instruction

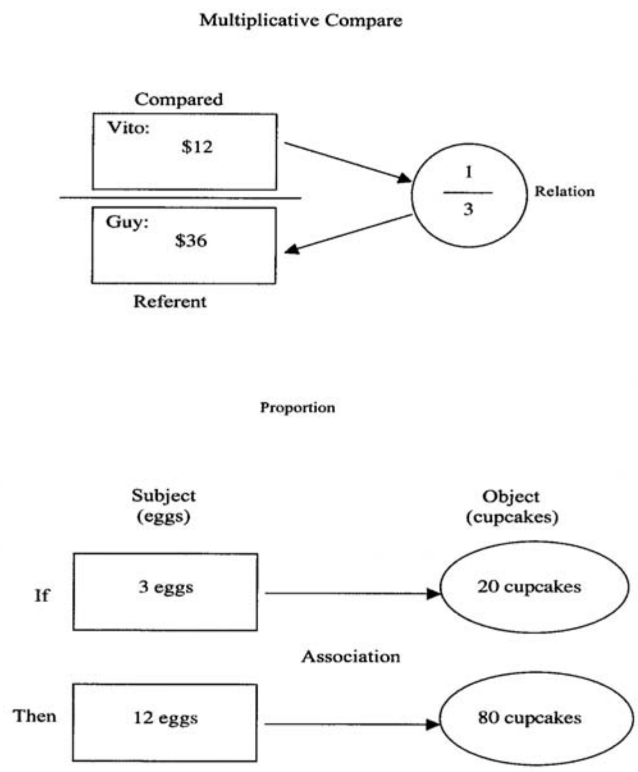
There are several additional important characteristics of most contemporary approaches to explicit strategy instruction. Van Luit and Naglieri (1999) provide a concise description. In their view, strategy instruction is when "students are taught to flexibly apply a small repertoire of strategies that reflect the processes most frequently evidenced by skilled students" (p. 99). They also stress the importance of a good deal of small group interaction in which students are encouraged and prompted to think aloud as they do mathematics, and peers provide feedback on their strategy selection and execution.

Van Luit and Naglieri (1999) began instruction with use of concrete objects but then expeditiously moved into mental solutions that entailed a good deal of thinking aloud. The final phase of each instructional cycle included a "phase of control, shortening, automization and generalization" (p. 101). What is similar between these two methods is that transfer and practice for automaticity are not assumed. Nor are students expected to develop these proficiencies by doing homework problems or by informal discussions with peers. Significant blocks of instructional time are dedicated to these goals, and teachers closely monitor student progress toward independent performance. Whereas the goal of automaticity and clear, explicit modeling remains central, teaching students how to transfer the knowledge they obtained is a major focus, and is characteristic of the more contemporary explicit strategy instruction studies.

Another strand in this research seems particularly relevant for students with LD who struggle with story problems. In related streams of research, Hutchinson (1993), Jitendra et al. (1998) and Xin et al. (2005) taught students in a systematic fashion a graphic representation to help them analyze the contents of a story problem. The three studies have addressed a) simple arithmetic word problems involving addition and subtraction ("change, combine, compare" following the Riley, Greeno, and Heller (1983) representational system) b) comparative problems involving multiplication (Xin et al., 2005), and c) word problems that typically are taught in beginning algebra.

The goal is to help students grasp the nature of word problems that involve an operation (either addition or multiplication of whole numbers) and its inverse operation. Rather than focusing on tricks such as "key words" students learn to use a visual representation to analyze the question and then discern how to handle relevant information. Exposure to all aspects of each of the problem types is deliberate and explicit. Practice is extensive, including opportunities for students to think aloud as they complete their graphic organizers. The instructor carefully highlights the key aspects of each problem type and provides a good deal of discrimination practice. Figure 1 below is an example of the graphic representation used to teach students a way to analyze multiplication problems involving comparisons (Xin, Jitendra, & Deatline-Buchman, 2005, p. 185).

Figure 2: General Problem-Solving Steps Employed in the Schema-Based Instruction and General Strategy Instruction Conditions



Source: Xin et al., 2005, p. 185.

Upon examining the full array of studies, one is struck by several features. The first is that all these studies address topics that are particularly problematic for students with LD, particularly those with difficulties in both mathematics and reading (Jordan, Hanich, & Kaplan, 2003). The second is that the pooled effects on word problems, computation, and transfer outcomes are all significant. The third is that the instructional strategies in the interventions do borrow from both the mathematics education research and the cognitive development research in mathematics. This seems an advance over the very general heuristics that comprised much of the mathematics intervention research in the special education literature 15 to 20 years ago. Those generic strategies were often borrowed from the research on reading comprehension or writing, and failed to capitalize on the advances made in research on the teaching and learning of mathematics. In fact, Xin et al. (2005) intentionally used the older, generic problem solving approach as the control group condition and found large effects favoring the more innovative approach for helping students understand the mathematical nature of the story problem.

Because in explicit strategy instruction students are invariably taught how to approach the problem type or types and are usually given precise wording to use as they think aloud, development of mathematical insight rarely plays a role in the design of the interventions (with the possible exception of Van Luit and Naglieri, 1999, in which multiple strategies are highlighted, potentially allowing such insights to develop). Therefore, there is not much known about the extent to which explicit instruction helps support students in developing such insights or understandings since proficiency and conceptual knowledge are always related in an integral fashion (Rittle-Johnson, Siegler, & Alibali, 2001).

In summary, this body of research on explicit instruction suggests that the field has made reasonable strides in understanding at least one type of intensive mathematics instruction that will help students with LD become more proficient in solving relatively basic grade level word problems and at least make some gains toward understanding how to translate stories or written problems into appropriate symbols, representations, and mathematical expressions.

This approach for providing explicit systematic instruction should also help inform the development and implementation of the type of preventative small group interventions that are increasingly used to help students who are struggling to acquire proficiency in mathematics in general classroom instruction. Preventative small group interventions provide students, who are identified as struggling in the Tier 1 core curriculum, with specific skill instruction in small groups. A major goal of preventative small group interventions is that they will reduce inappropriate referrals to special education, because if students benefit from a relatively low cost small group mathematics intervention in their general classroom, they are unlikely to require the intensive instruction that special education is intended to provide.

3. Studies Evaluating the Impact of Explicit Instruction for Low-Achieving Students

As described earlier, explicit instruction requires the teacher to be the provider of knowledge and to provide a great deal of structure and control concerning how content is learned, including the specific strategies or steps used by the children to solve the problems. Table 11 summarizes the results from the studies that investigated the effects of various strategies on the math achievement of low-achieving students. The four studies that investigated explicit instruction as a means for teaching low-achieving students are: Darch, Carnine, and Gersten (1984); Kroesbergen, Van Luit, and Maas (2004); Moore and Carnine (1989); and Woodward and Brown (2006). All but one of the effect sizes for the explicit instruction studies (i.e., Woodward & Brown, 2006) are significant.[2]

Table 11: Studies That Investigate the Effects of Various Instructional Strategies on Math Achievement for Low-Achieving Students

Study	Design	Sample	Duration/ Content	Contrast	Measure	Hedge's g	Standard Error	
Darch et al., 1984	RCT	73 low-achieving fourth-grade students in one school	Eleven 30-minute lessons/ Math story problems	Explicit Method with Fixed Time vs. Basal Instruction with Fixed Time	Researcher designed word problem test	1.914	***	0.408
Fuchs et al., 2005	RCT	139 first-grade students at risk for the development of math difficulty in 41 classrooms in 10 schools	48 sessions, 3 times weekly for 16 weeks/ Identifying numbers, more and less, addition and subtraction	Tutoring based on CRA vs. No Tutoring	Pooled computation measures (includes two tests)	0.441	**	0.179
					Pooled fact fluency measures (includes two tests)	0.180	(ns)	0.177
					Pooled conceptual and application measures (includes three tests)	0.414	*	0.179
Kroesbergen et al., 2004	RCT	265 students aged 8–11 years old from 13 general and 11 special elementary schools for students with learning and/or behavioral disorders in the Netherlands	Thirty 30-minute lessons, twice weekly, over 4 to 5 months/ Multiplication	Explicit vs. Traditional Instruction	Pooled Problem Solving Measures	0.569	*	0.262

Continued on p. 6-61

[2] An obvious outlier that was not included in the table was Cardelle-Elawar's (1995) study. The effect size was equivalent, for example, to the average control classroom being at the 3rd percentile and the average experimental classroom being at approximately the 84th percentile. The effect size for that study is extraordinarily large. This may, in part be due to the fact that the unit of analysis was the classroom, not the individual child. Effect sizes are larger when analysis is based on means of classrooms because individual differences among children within classrooms are minimized. However, there are likely to be other factors relating to the alignment of test to the intervention that lead to the study's extraordinary high effect size.

Table 11, continued

Study	Design	Sample	Duration/ Content	Contrast	Measure	Hedge's g	Standard Error	
Moore & Carnine, 1989	RCT	29 students in Grades 9–11 from three math classes for low-performing students in a high school in a medium-sized city in the Northwest	Twenty 50-minute lessons/ Ratio and proportion word problems	ATCD (Active teaching with empirically validated curriculum design) vs. ATB (Active training with basals)	Researcher designed criterion referenced test to assess student mastery of specific mathematics skills	0.994	**	0.386
Pasnak et al., 1991[a]	RCT	85 low-performing students from 17 Kindergarten classes in six neighboring Northern Virginia schools	3–4 sessions per week over three months/ Introductory mathematical concepts	Piacceleration vs. Control	SESAT math subtest	0.520	(ns)	0.348
Thackwray et al., 1985	RCT	60 third- and fourth-grade children with teacher perceived academic problems from three urban public schools	Four 45-minute sessions/Addition	Specific self instruction vs. Didactic	Pooled math quiz (ES = 0.501) and Peabody Individual Achievement Test (ES = 0.780)	0.641	*	0.325
Woodward & Baxter, 1997[b]	RCT	38 low-achieving third-grade students in nine classes in three schools located in the Pacific Northwest	One school year/ Third grade math	Everyday Mathematics vs. Heath Mathematics Program	ITBS including computation (ES = -0.176), concepts (ES =0.199), and problem solving skills (ES = -0.085) subtests	-0.223	(ns)	0.635
Woodward & Brown, 2006[b]	Quasi	53 students in two middle schools in nearby medium-sized suburban school districts. Students had been identified as low-achieving in mathematics by elementary school teachers. No student had an IEP for mathematics	One school year/ Both curricula emphasized core NCTM strands: numbers, operations, measurement, geometry, data analysis and probability	Transitional Math Curriculum vs. Connected Math Program	Pooled standardized (ES = 0.797) and researcher developed test (ES = 1.435)	1.116	(ns)	0.688

~ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$
[a] Data were adjusted for clustering that occurred within classrooms.
[b] Data were adjusted for clustering that occurred within schools.

Two of the studies (Darch et al., 1984; Moore & Carnine, 1989) used a highly explicit approach based on the Direct Instruction model articulated by Silbert, Carnine, and Stein (1989) and Engelmann and Carnine (1982). This is a traditional approach to explicit instruction, which has been widely used in the field of special education with students with LD, especially in the 1980s and early 1990s. With direct instruction, teachers model how to solve a specific problem type, and spell out the necessary steps. Students learn the steps and through careful sequences of examples, practice solving problems in the precise fashion that they were taught. Another one of the studies, Kroesbergen et al. (2004) also employed a highly explicit instructional approach. That is, students (8–11 years old) were instructed via directions and modeling by their teacher how and when to apply specific strategies for solving multiplication computation problems. Students were directed to only use the strategy taught by the teacher. Although highly explicit instruction has been shown to lead to enhanced academic outcomes for students with learning disabilities, and other students considered at risk for experiencing difficulties in mathematics (e.g., Gersten & Carnine, 1984; Baker, Gersten, & Lee, 2002) some have questioned the extent to which students actually learn the underlying rationale behind the strategies that are explicitly taught (e.g., Woodward & Howard, 1994; Woodward & Montague, 2002).

The two studies (Kroesbergen et al., 2004; Woodward & Brown, 2006) could be characterized as teaching students a variety of heuristics for problem solving but with significant segments of instruction following the highly explicit nature of classic direct instruction. In fact, one of the goals in the framing of some of the research studies we discuss below is an attempt to ponder and define the nature of explicit instruction for low-achieving students. Their thinking is helpful in beginning to unpack this construct.

Woodward and Brown (2006) found that despite statements by the National Council of Teachers of Mathematics (NCTM) (2000) indicating that students experiencing difficulties in mathematics benefit from a challenging curriculum, they could not locate any research to support this claim. They note, "In-depth examinations of this population indicate that without substantive modifications, these students do not exhibit high levels of success on either academic measures or everyday activities" (e.g., Baxter, Woodward, Wong & Voorhies, 2002; Woodward & Baxter, 1997, p. 151). Their analysis of the relevant research, with which we concur, notes that effective components of instruction for low-achieving students in mathematics supports the use of both concrete and visual representations of concepts, carefully orchestrated practice activities with feedback on all aspects of mathematics and high, but reasonable, expectations.

Woodward and Brown (2006) evaluated an intervention, written by Woodward, called Transitional Mathematics, and attempted to put these components into practice in six intensive, remedial middle school classrooms. The curriculum included numerous visual models for representing mathematical procedures in a meaningful way. They present, for example, difficult concepts such as place value in three-digit addition by both using a written algorithm and a visual model that depicts the algorithm. Regrouping was taught via systematic use of expanded algorithms as well as visual models of the expanded algorithm. Practice on relevant mathematics facts, and factoring was part of each daily lesson. The teacher explicitly introduced the concepts, and worked problems with the group that exemplified the concept before students broke into pairs. Because of the reading difficulties of many of the students, the teacher most often read the problem to the students. Guided practice consisted of approximately five problems worked on by students and reviewed with the teacher. This part of the lesson also included a good deal of checking for understanding (Good & Grouws, 1977) and attempts to explore any student misconceptions. Practice was typically done in pairs and included opportunities for students to explain their reasoning to each other and with the class. Students in the comparison classroom were taught using the Connected Mathematics Program, a commonly used middle school curriculum. Woodward and Brown characterize this as follows: The core emphasis of this program is problem solving, and students typically read descriptions of problems as part of each lesson. Connected Mathematics is much more contextualized in elaborate "real-world" problems, and has a more peripheral attention to skill development, in contrast with Transitional Mathematics that integrates the latter with distributed practice. This quasi-experimental study involved two schools, one the intervention school and one the comparison school—no mention was made concerning how schools were designated.

Regarding differences in achievement between groups, the effect size on the Terra Nova, a standardized mathematics achievement test was not significant but indicative that the Transitional Mathematics treatment is a promising approach (ES = 1.116). Note how this study, like the others in the explicit instruction set also includes an array of other practices deemed to be beneficial—use of guided practice, intensive use of visual models so that students can represent problems in multiple ways (Donovan & Bransford, 2005), clear and explicit instruction in use of the concepts and provision of heuristics for problem solving.

Kroesbergen et al. (2004) and Darch, Carnine, and Gersten (1984) used an approach that was even more explicit than the Woodward and Brown (2006) model. In these studies, teachers modeled an approach for solving problems and students were expected to follow the teachers' model. Teachers did explain when the strategy was appropriate, and provided examples of occasions when it was not appropriate. In both cases, the degree of structure was higher than in Woodward and Brown (2006) and students were not given a chance to talk through their approach for solving the problem with a partner or the teacher. Both studies (Kroesbergen et al. and Darch, Carnine, & Gersten) demonstrated significant effects on researcher-developed measures that were aligned with curricula taught, favoring the explicit instruction groups. In the Kroesbergen et al. (2004) study, the effect size in the area of word problems involving multiplication was significant (ES = 0.569) and in Darch, Carnine, and Gersten (1984) there was also a significant effect size in the area of word problems (ES = 1.914) that cut across all four basic arithmetic operations.

Moore and Carnine (1989) explored the degree of explicitness within the context of highly interactive, teacher-directed instruction. In this study, high school students were taught how to solve ratio and proportion problems. Whereas control group students were taught to ask themselves "Is this information important?" students in the experimental condition were taught in a much more step-by-step fashion and were taught strategies for each of four types of problem sets. For this study, the effect size was significant (ES = 0.994) on a test of mastery of specific skills. However, one must take into account that this study measured only the topic covered, in contrast to Woodward and Brown (2006), which measured all aspects of mathematics covered in a typical school year.

F. Other Approaches for Teaching Students With Learning Disabilities

This next section addresses other instructional approaches for teaching mathematics to students with LD. The findings are organized by three major themes:

- Selection of examples to foster development of more sophisticated strategies for quick retrieval of basic arithmetic facts
- Emphasis on visual representation
- Emphasis on encouraging students to think aloud

It is interesting to note that virtually all of the studies in these categories also have at least a reasonably strong degree of explicitness in the design of their instruction—a feature that is consistent across the body of studies reviewed for this section.

1. Strategies for Quick Retrieval of Basic Arithmetic Facts

Quick retrieval of basic arithmetic facts or combinations has been assumed by virtually the entire mathematics education community as critical for success in more advanced mathematics. It is considered a necessary, though not a sufficient requirement for emerging mathematical competence. Researchers in the field of LD have found for several decades that slow and inaccurate retrieval of basic combinations is a clear, consistent early indicator of persistent serious difficulties in mathematics (Gersten, Jordan, & Flojo, 2005; Geary, 2005; Jordan, Hanich, & Kaplan, 2003; Goldman & Pellegrino, 1987; Hasselbring, Goin, & Bransford, 1988).

Two studies with students with LD were included in this classification, and are summarized in Table 12. Beirne-Smith (1991) attempted to examine whether sequencing of examples could enhance facility with basic addition combinations for students with LD. She used an array of facts developed by Carnine and Stein (1981) that was geared toward helping students see that to compute, for example, 8 + 2, all they needed to do was count up by 2. Examples of the array are 2 + 4, 2 + 5, and 2 + 6. The impact of the sequence did not lead to significant improvement over simple rote practice.

Table 12: Studies That Investigate the Use Of Strategies With Students With Learning Disabilities to Develop the Ability to Quickly Retrieve Arithmetic Facts

Study	Design	Sample	Duration/ Content	Contrast	Measure	Hedge's g	Standard Error	
Beirne-Smith, 1991	RCT	30 students with LD aged 6 to 10 years old from four schools in two adjacent southeastern school districts were tutored. 20 students with no learning disabilities in Grades 3–6 served as tutors	30-minute tutoring sessions for four weeks/ Single-digit addition facts	Counting-on procedure vs. Rote memorization	Oral test on addition facts	0.165	(ns)	0.448
Woodward, 2006	RCT	15 fourth-grade LD students from two "mainstreamed" classrooms in a school in a suburban school district in the Pacific Northwest	20 25-minute sessions daily over four consecutive weeks/ Multiplication facts	Strategy and timed practice vs. Time practice via direct instruction	Pooled researcher designed computation measures	0.377	(ns)	0.509

~ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

Woodward (2006) extended this line of research to much more complex multiplication combinations, which require many students to rely on some variant of multiplication tables and sheer rote practice. He developed an intervention containing two components. The first involved explicit instruction in an array of strategies that can help with quick retrieval of multiplication combinations. These included numerous shortcuts based on properties of numbers. One is counting backward for combinations of 9, i.e., knowing that 8×9 achieves the same answer as $8 \times 10 - 8$. Another is use of the distributive law, e.g., 37×5 equates to the same product 35×5 plus 2×5 An advantage of this strategy's approach is that students could not only learn more efficient ways to compute these multiplication facts but also develop their facility with using properties of numbers to solve problems.

However, Woodward (2006) noted that strategy instruction will not, in and of itself, promote quick retrieval of mathematical combinations for all students with LD (see Hasselbring, Goin, & Bransford, 1988). He therefore combined the strategy instruction and practice with timed practice drills. He compared students taught with a combined strategy instruction and timed practice approach to students taught only with timed practice.

Results were positive favoring the strategy group and nonsignificant. However, given the small sample size, and inconsistent findings across mathematics domains, one can only infer that this approach—or aspects of this approach—might be worth exploring in terms of development of more fluent retrieval as well as in helping students understand more about number families and increasing their ability to estimate. However, the set of studies on building fluency in computing mentally or retrieving arithmetic combinations indicates that there is a good deal more to be learned about how to improve students' proficiency in this critical area.

2. Use of Visual Representations, Visualization, and the Concrete-Representation-Abstract Approach

Adding It Up, the 2001 National Research Council report on the teaching of mathematics, eloquently describes the role of representations in the teaching and learning of mathematics, a role that has not always been adequately highlighted until recently in the instructional research on LD.

> Mathematics requires representations. In fact, because of the abstract nature of mathematics, people have access to mathematical ideas only through the representations of those ideas. … *Much of the real intellectual work in mathematics concerns the interpretation and use of representations of mathematical ideas* (pp. 94–95, emphasis added).

The authors explain that mathematical ideas are often metaphorical, and thus, a representation or multiple representations are excellent means for conveying mathematical ideas. This section summarizes a set of five recent studies on the role of visual representations as a key means for teaching mathematical ideas, strategies, and procedures to students with LD. Each researcher approaches the use of representations somewhat differently.

Table 13 presents information on each of the studies, as well as the outcomes of the studies. Because the instructional approaches are so different, the Task Group did not pool effect sizes across the set of studies. However, taken together, these approaches reflect a trend toward serious thinking about instructional uses of representations that include physical models using manipulatives, pictorial representations, abstract representations using geometric shapes, as well as increasingly abstract representations, such as number lines and graphs of functions and relationships. It should be noted, however, that the effect sizes across studies are quite variable.

Table 13: Studies That Investigate the Use of Concrete Instruction and Visual Representations Used for Students With Learning Disabilities

Study	Design	Sample	Duration/Content	Contrast	Measure	Hedge's g	Standard Error	
Butler et al., 2003[a]	RCT	50 students in Grades 6–8 with mild–moderate disabilities (42 students with specific learning disabilities in math and eight with other disabilities) from a public middle school located in a large urban area of the Southwest	Ten 45-minute lessons/ Fraction concepts and procedures	Concrete-representational-abstract (CRA) vs. Representational-abstract (RA)	Area Fractions, Quantity Fractions, and Improper Fractions subtests provided measures of conceptual understanding of fraction equivalency and Abstract Fractions and Word Problems subtest provided a measure of application	-0.095	(ns)	0.526
Manalo et al., 2000—Experiment 1	RCT	29 From three students (equivalent to eighth grade) with learning disabilities from two schools in the Palmerston North area of New Zealand	Five 25-minute sessions twice per week/ Addition and subtraction	Process mnemonics vs. Demonstration imitation	Immediate Posttests - Addition and subtraction computation skills tests	-0.043	(ns)	0.477
					1 week follow-up tests	0.076	(ns)	0.477
					6 week follow-up tests	0.956	~	0.506
Manalo et al., 2000—Experiment 2	RCT	28 From three students (equivalent to eighth grade) with learning disabilities from two schools in Auckland, New Zealand	Ten 25-minute sessions twice per week/ Addition, subtraction, multiplication, and division	Process mnemonics vs. Demonstration imitation	Immediate Posttests - Addition, subtraction, multiplication, and division computation skills tests	-0.153	(ns)	0.475
					1 week follow-up tests	0.180	(ns)	0.472
					8 week follow-up tests	1.876	**	0.579
Walker & Poteet, 1989[a]	RCT	70 sixth- and eighth-grade LD students receiving mathematics instruction in resource room programs in four Indiana school districts	Seventeen 30-minute lesson/ Problem solving strategies for simple word problems involving addition and subtraction	Instruction using diagrammatic representations vs. Traditional instruction	One-step story problem-solving test	0.349	(ns)	0.330
Witzel et al., 2003[a]	RCT	34 matched pairs of sixth- and seventh-grade students with learning disabilities or at-risk for difficulties in math (41 LD students and 27 at-risk) in 12 inclusive classrooms in an urban county in the Southeast	Nineteen 50-minute lessons/ Algebraic transformation equations	Concrete-representational-abstract (CRA) vs. Abstract	Algebra transformation equations test	0.826	*	0.346

~ p < .10, * p < .05, ** p < .01, *** p < .001
[a] Data were adjusted for clustering that occurred within classrooms.

The intriguing set of two experiments by Manalo and colleagues (2000) examines the use of easy-to-imagine visual stories and schema to help students remember rules, principles, and procedures. These studies address a potentially important issue in the practice of teaching mathematics: how to provide prompts or facilitators to help students create visual representations. In this study, explicit teacher-directed instruction (modeling followed by guided practice with clear feedback) is a constant. The variable is visualization.

Manalo et al. (2000) adapted an approach from a Japanese educator, Nakane. The goal of this approach is to “summarize the organization and the process of problem solving … using familiar metaphors expressed in familiar ways” (p. 138) and thus to teach mathematical operations in a clear, comprehensible fashion. The goal of the researcher was to present mathematical problems as interesting, easy to visualize narratives that would engage the students, and thus enhance their interest in the process, their memory of the procedures taught, and the questions that children must ask themselves before deciding on a strategy for solving a problem.

For Study 1, the topic was basic arithmetic computation problems; all participants were screened to ensure that they were not proficient in use of standard algorithms for multidigit operations involving regrouping, even though they had been taught this material before, often many times before. Numbers were presented as characters and operations as stories. For example, to teach subtraction, students were asked to visualize warriors with numbers on their uniforms, and to visualize that the bigger the number on the uniform, the stronger the warrior. The teacher used simple drawings to demonstrate the procedure or story. For subtraction, the top number represented the attackers and the bottom numbers the defenders, and students were told that the attackers weakened during the battle. The number on the uniform of a defender told a student how much strength was sapped from the warrior. In cases involving regrouping (e.g., 33-5), students were told that for example, a warrior with strength of 3 would not have adequate strength to sustain a battle with a defender with strength of 5. Thus, the army would need to regroup and borrow some strength from the warrior with strength of 30. The teacher used pictures to demonstrate the process of regrouping.

Similar stories were developed for multiplication and division. The approach used to teach students in both the experimental and control conditions was a combination of model-demonstration with guided practice and feedback. Two experiments were conducted. The first entailed the researcher as the teacher; the second used two different teachers. For both studies, the pattern of findings was similar. No significant effects were found on the immediate posttest or a test administered one week later. Yet, on the six-week and eight-week follow-up tests, the effect for Experiment 1 (in which the researcher did all the teaching) was 0.956, which bordered on statistical significance, and for Experiment 2, which used teachers other than the researcher, the effect was larger and statistically significant (effect size = 1.876).

Given the problems with maintenance of knowledge for many students with LD, these results seem worth noting. The use of consistent visual representations and stories to help students think through their decisions about appropriate computational processes is an important fact to note. One wonders about its impact on helping students with LD translate more complex mathematical problems and work with more complex mathematical concepts.

3. Visual Representations and Helping Students Understand Visual Representations by Use of the Concrete-Representational-Abstract (CRA) Method

Two of the studies in this section (Butler et al., 2003; Witzel et al., 2003) examine the use of concrete-representational-abstract (CRA) instruction for students with learning disabilities. This sequence of instruction begins at a concrete level, with students manipulating objects. Once students understand a topic concretely, they work with the topic using visual representations. Once the students are comfortable with how the topic can be represented in multiple ways, they work with the concepts at a more abstract level. The Walker and Poteet study (1989–1990) are also included in this section because their work can be seen as a precursor to the more complex CRA model.

In the earliest study in this subcategory, Walker and Poteet (1989–1990) compared a diagramming method of problem solving with a keyword approach. Subjects were middle school students with LD. In both conditions in this study, explicit instruction was a constant and not a variable. The experimental variable using a visual representation to help students organize information from one- and two-step story problems involving basic addition and subtraction, then to translating the pictures into numerical expressions, and ultimately to computing the answer.

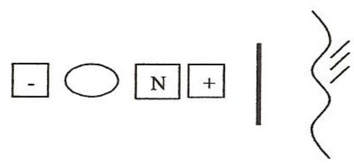
Although this skill seems exceptionally easy for middle school students, poor performance on word problems is prevalent with this group. In fact, on the pretest, the average score for students was equivalent to 16.43 correct (out of 32 possible problems, 51.34% correct). Students in the diagramming group were taught to create diagrams that bear similarity to those used by Xin et al. (2005) and Jitendra et al. (1998). Finally, students were asked to compute the actual solution. The comparison group was taught to identify keywords in the problem that could then be directly translated to specific numerical operations. Despite not reaching statistical significance, the effect size (0.349) suggests that, given the difficulty that students with LD have in developing proficiency in this area; the approach could well be labeled promising.

Butler et al. (2003) used the CRA to teach middle school students with LD basic concepts and procedures involving fractions. The topics included concepts and procedures related to equivalence of fractions and computations involving fractions. The authors note that researchers (e.g., Woodward & Montague, 2002) have suggested that many students with LD lack any real understanding of the concepts underlying various procedures that they can perform and that these problems truly surface once students begin to work with rational number concepts and operations. In this study, as in Walker and Poteet (1989–1990), the instructional methodology was similar for experimental and control students in that explicit instruction was used in both conditions. The major difference was that CRA students spent three days working with concrete objects, three days with visual representations, and only then moved on to abstract, symbolic notation. The control condition began with visual representations for three days. Major emphasis in both conditions was placed on fractions as part of a set, as opposed to fractions as area. As one can see in the first study in Table 13, the effect size for this method was nonsignificant. Thus, the CRA intervention implemented in this study was not more effective than the control condition in teaching fractions to middle school students with LD.

The other CRA study, Witzel et al. (2003), also conducted with middle school students, differs in several important ways from Butler et al. (2003). The first is that the topic was a more difficult one, algebraic transformation equations. The second is that the researchers used CRA quite differently. For example, Witzel et al. progressed more fluidly from concrete, to visual, to abstract. The third difference is that, in this case, a researcher-developed measure was used rather than the standardized measure used by Butler et al. Finally, as can be seen in the last study in Table 13, effect size (ES = 0.826) is statistically significant.

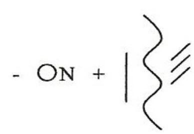
The authors note that because of the abstract nature of algebra, building a mathematically accurate concrete representation is much more of a struggle. Figure 3 presents an example of the instructional materials used and how the researchers grappled with representation of a variable (*x*) with concrete objects when x can represent any real number (Witzel et al., 2003, p. 127).

Figure 3: Concrete, Representational, and Abstract Examples of an Inverse Operation



To solve a concrete problem, students manipulate objects at each step towards the solution.

A pictorial representation would closely resemble the concrete objects but could be drawn exactly as it appears here.



To solve a representational problem, students draw each step towards the solution.

An abstract problem is written using Arabic symbols as displayed in most textbooks and standardized exams.

$$-1N + 10 = 3$$

To solve an abstract problem students write each step in Arabic symbols.

Source: Witzel et al., p. 127.

G. Strategies That Encourage Students to Think Aloud

The Task Group identified two studies that examined strategies that encouraged students with LD to think aloud (Ross & Braydon, 1991; Schunk & Cox, 1986). Table 14 below summarizes characteristics for the two studies, and presents the effect sizes.

Table 14: Studies That Investigate the Impact of Think Aloud Strategies With Students with Learning Disabilities

Study	Design	Sample	Duration/ Content	Contrast	Measure	Hedge's g	Standard Error	
Ross & Braden, 1991[a]	RCT	94 elementary school students with LD in nine intact special education resource rooms classified as learning disabled in math	Nineteen 60-minute sessions over four weeks/ Addition and subtraction	Cognitive behavior therapy in which students are instructed to talk aloud vs. Direct instruction	Stanford Diagnostic Mathematics Test - computations	0.135	(ns)	0.434
Schunk & Cox, 1986	RCT	90 students classified with LD in math from six middle schools	Six 45-minute sessions/ Subtraction with regrouping	Continuous verbalization vs. No verbalization	Subtraction test	1.005	***	0.271

~ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$
[a] Data were adjusted for clustering that occurred within classrooms.

As previously mentioned, asking students to think aloud was a major component in many of the explicit instruction studies. What differentiates these two studies from those in the explicit instruction set (e.g., Manalo, 2000; Tournaki, 2003) is that in these studies, verbalization was the sole independent variable. In contrast, in the other explicit instruction studies, students thinking aloud was but one of several instructional components. Thus, these two studies suggest that: encouraging students to think aloud as they work on arithmetic problems shows promise as one component of a mathematics intervention.

Both of these studies were influenced, to some extent, by the research of Donald Meichenbaum (1985), which suggested that students with learning disabilities, behavior disorders and, in all likelihood attention deficit disorders, could be helped in many areas of both academic and social development by being taught to verbalize. The Meichenbaum approach targets one of the key characteristics of students with LD—Geary's (2005) concept of impulsivity and Kolligian and Sternberg's (1987) concept of lack of task persistence. By actively encouraging students to speak to themselves about the strategies they are using to solve a problem, the researchers felt that students would be inhibited from quickly, almost recklessly proceeding forward without serious thought. In addition, the focus on active encouragement of thinking aloud is integrally linked to Vygotsky's notion that thought is inner speech, and that students may well need to go through a period of actually thinking out loud, especially those students with learning difficulties.

Schunk and Cox (1986) examined the effectiveness of verbalizing the steps of problem solving with middle school students with LD working two- to six-column subtraction problems with and without regrouping. As in the Tournaki study, the teacher in the treatment group talked through the steps of solving the subtraction problem and then students worked several problems while verbalizing the steps.

Students in the comparison group learned the same procedures but were in no way encouraged to verbalize during problem solving. This study is different from both the Friedman (1992) and Lambert (1996) studies in that students were solving computation problems not word problems, although some of the computations were rather complex. A statistically significant effect size of 1.005 was found on a subtraction computation test that was closely aligned to the types of problems used during instruction. Effects were modest and not statistically significant for the Ross and Braden (1991) study (ES = 0.135), although this could be due to the fact that the measure lacked the tight alignment to the intervention of the Schunk and Cox study. In any case, one study (Schunk & Cox, 1986) but not the other (Ross & Baden, 1991) suggests that for students with LD, merely encouraging self-verbalization or thinking aloud can have beneficial effects in terms of learning mathematics.

H. Other Approaches for Teaching Low-Achieving Students

Unlike studies of other approaches to teaching students with LD, most of the other studies reviewed with low-achieving students can be characterized as providing primarily *implicit* instruction (with the exception of Fuchs et al., 2005). Implicit instruction refers to the teaching approaches that provide students with broad guidance in terms of general procedures for solving problems, including relatively broad questions to ask themselves. However, there is little in the way of specific guidance in how students construct knowledge, and these approaches do not necessarily include any mathematics in them. Students are provided strategies that are used to solve math problems, such as teaching students to think aloud, or use visual representations with strategic use of manipulatives. For example, Thackwray, Meyers, Schleser, and Cohen (1985) taught students five specific self-instructions to say out loud while solving addition word problems. Presumably, these self-instructions were intended to enhance student's ability to construct accurate representations of the problem features and solution strategies.

The four studies that are included in this category are Fuchs et al. (2005); Pasnak, McCutcheon, Holt, and Campbell (1991); Thackwray et al. (1985); and Woodward and Baxter (1997). Table 11 summarizes the characteristics of these four studies. It is important to note that these studies do provide various degrees of teacher direction, so we prefer the term "primarily implicit instruction" rather than implicit instruction because all studies reviewed below seem to provide instruction primarily, but not necessarily exclusively, via an implicit instructional approach. We treat each study separately as we did for studies considered in this section for evaluating the impact of explicit instruction for low-achieving students.

One study, Thackwray et al. (1985), examined the effects of encouraging students to think aloud as they worked using the cognitive behavioral model developed by Donald Meichenbaum (1985), which was a common special education technique in the 1970s and 1980s. This technique was based on the premise that thinking aloud consistently will increase students' ability to reflect on their actions and help dissipate some of the impulsivity that is typical of low-achieving students in mathematics. Students were taught five steps. The first step involved orienting students to solve the problem. The next two steps appear below:

Step 2: First, I have to look at the problem very slowly to determine if it is addition, subtraction, multiplication or division.

Step 3: This one is addition. I can tell by the sign. (Thackwray et al., 1985, p. 301).

First, the instructor (a graduate student) modeled the steps; gradually, the student performed the five steps independently with no prompting. This study exemplifies implicit instruction because teachers provided minimal control over how students solved the problems. Rather, students were allowed to verbalize as they wished. In the specific self-instruction condition, the experimenter modeled the self-instructions (verbalizations) while the teacher performed two, three, and four digit addition problems. Using Meichenbaum's (1975) five-step fading procedure, the experimenter gradually required the child to verbalize while performing each step toward solution alone while solving the math problem. In the didactic condition, children were simply provided instructions concerning what to verbalize during problem solving. However, no modeling of the verbalizing process was provided.

Thackwray et al. (1985) investigated the effectiveness of this approach in a study involving 60 third- and fourth-graders who were perceived as experiencing difficulties in mathematics by their teachers. Although teacher judgment is no substitute for a mathematics performance measure, it often is reasonably accurate (Hoge & Coladarci, 1989). The intervention was quite short: four 45-minute lessons. The content was problems involving whole number addition. The control group received typical lecture-demonstration-practice with feedback instruction. The effect size (ES = 0.641) was significant, suggesting evidence of efficacy for this approach. The outcome was a composite of a standardized test: the Peabody Individual Achievement Test (PIAT) and a 20-item addition test. Based on the one study, there appears to be evidence of the effectiveness of promise in this general approach for problem solving, though replication of these findings in other studies would seem important for further research.

1. Instruction in Piagetian Cognitive Operations (Classification, Seriation, and Number Conservation)

The writings of Jean Piaget have always played a role in instructional research in mathematics, most recently in Griffin, Case, and Siegler (1994). Pasnak et al. (1991) examined the impact of small group instruction on Piagetian cognitive operations on kindergartners' performance on the Stanford Early School Achievement Test (SESAT)-Mathematics. The SESAT is essentially a readiness test, as opposed to a mathematics achievement test.

The sample was also selected by the kindergarten teachers as students who were having difficulty learning the basics of mathematics in kindergarten. The researchers used the students' scores on the Otis Lennon School Ability Index to confirm that they were in the at risk category. On average, these students were at least .5 standard deviation units below the school mean.

Instruction focused on the three Piagetian concrete operations that many students acquire informally before kindergarten (classification, seriation, and conservation). Many types of manipulatives were used (bolts, cups, lima beans, dominoes etc.). The amount of time devoted to this instruction was appreciable, three months of 15–20 minute small group lessons, delivered three to four times a week. Control group students received typical kindergarten instruction in numbers and number concepts.

The effect size (0.520) was not statistically significant, when corrected for classroom level clustering. Nonetheless, the magnitude of the effect size, especially given the fact that a standardized achievement test was used which was not closely aligned to the specific content taught, suggests there may well be some promise to this approach.

2. Evaluation of the Effects of 'Reform' Curricula on Low-Achieving Students

Woodward and Baxter (1997) conducted a small, but oft-cited, quasi-experiment that examined the impact of *Everyday Mathematics*, one of the curricula assumed to be consistent with the 1989 NCTM *Curriculum and Evaluation Standards for School Mathematics*. The study involved 38 low-achieving third-grade students in nine classes. The researchers assessed the impact of the reform curriculum versus a more mainstream commonly used core mathematics text. Note that students in neither condition received any additional support in mathematics from either special education or Title I. Results were nonsignificant (when adjusted for within-school clustering) and favored the control group (ES = -0.223). The reform curriculum produced one positive effect on the Concepts section of the Iowa Test of Basic Skills (ITBS) and two negligible negative effects: on the ITBS subtests for Computation and Problem Solving. None of these effects was significant. One reasonable conclusion is that low-achieving students require additional support and intensive work on foundational skills and that use of an innovative curriculum will not lead to any serious benefit unless such support is provided above and beyond the students' classroom mathematics instruction. One notes that more recent research, including research by Woodward, adopts approaches that combine interest in teaching concepts along with procedures to build conceptual knowledge with the use of explicit instruction.

3. Response to Intervention: Evaluation of a Preventative Small Group Intervention for First-Graders at Risk for Experiencing Difficulties in Mathematics

We identified only one study that investigated the effects of tutoring using concrete-representational-abstract (CRA) instruction with low-achieving students.

Fuchs et al. (2005) screened first-grade students in 41 classrooms in 10 schools using a set of screening measures that are known to be valid and reliable (see for example Gersten, Jordan, & Flojo, 2005). These students received small group instruction three times per week, a typical procedure for preventative small group (Tier 2) interventions. Core components of the intervention included strategic use of manipulatives to ensure students understood more abstract visual representations and mathematical symbols, heavy emphasis on problem solving and discussion of solutions, and use of technology to provide individualized practice on basic addition and subtraction combinations to increase quick and fluent retrieval.

Fuchs et al. (2005) used a wide array of both researcher-developed and standardized measures, of computation and concepts, applications, or word problems, as well as addition and subtraction fact fluency. Effect sizes were 0.414 and significant, favoring the tutoring group for concepts or problem solving, 0.441 and significant for the combined computation measures and 0.180 but not significant for the two fact fluency measures. The effects were stronger for the computation and concepts measures than the fact fluency measure, indicating that the technology component appeared to be the weakest facet of the intervention. In interpreting the effect sizes, the reader should note that the control group students received no additional instruction. Thus the independent variable is receiving tutoring using a CRA-based instructional model versus receiving no additional support whatsoever.

In general, this appears to be an effective preventative small group early intervention for students who exhibit problems in mathematics at the beginning of the first grade. It also is a solid example of how both concepts, procedures, and problem solving can be taught and practiced in an intense, integrated fashion. It should be noted that beyond Bruner's concrete-pictorial-symbolic sequence, no information is provided about how the tutors interacted with the children about the mathematics.

I. Summary and Conclusions

The Task Group was able to locate a reasonable number of high-quality experimental and quasi-experimental studies that investigated the effectiveness of various mathematics interventions in teaching mathematics to students with LD and LA. These studies provide a great deal of guidance concerning some defining features of effective instructional approaches for students with learning disabilities as well as low-achieving students. These features, many of which are associated with explicit systematic instruction, can be roughly categorized as follows:

1) Concrete and visual representations (mathematical drawings)
2) Explanations by teachers
3) Explanations and math talk by students in whole class discussion
4) Students working together
5) Carefully orchestrated practice activities with feedback
6) High but reasonable expectations

Some additional features of this research are noteworthy beyond the generally consistent effectiveness of both explicit and primarily implicit instructional approaches (interestingly Kroesbergen et al. (2004) actually compared and found no differences in multiplication outcomes between these two approaches). The first is that studies varied widely in terms of mathematical skills that were targeted. Most included a focus on computation skills, while others included specific attention to word problem solving. This focus on problem solving in research on students with learning disabilities and low-achieving students is a relatively recent trend, and an important one, because students with LD and LA struggle, in particular, with word problems.

The second is that a small but important set of studies examined best methods for building quick retrieval of arithmetic combinations. Mathematics educators have long been aware of the importance of quick retrieval of basic combinations so that students can focus on the problem at hand. Retrieval is stressed in the NCTM's *Curriculum Focal Points for Prekindergarten through Grade 8 Mathematics: A Quest for Coherence* (2007). In addition, research on learning disabilities has consistently documented that inefficient and ineffective retrieval of combinations is typical of a student with learning disabilities in mathematics. Addressing this issue has been more of a struggle. Programs have been developed that orchestrate practice sets for each student and try to teach similar combinations together. However, Hasselbring et al. (1988) noted that even these programs are not successful with many students with LD.

The studies by Tournaki (2003) and Woodward (2006) are important because they demonstrate that there is wisdom in teaching students strategies about computation as a means of increasing speed and accuracy of retrieval. If nothing else, this type of instruction is more interesting and potentially engaging for students and more likely to build a deeper understanding of the number system than pure rote memorization. Note that Woodward intentionally paired strategy instruction with 15 minutes of timed practice. This mixture is one that seems to show promise.

Additionally, many studies examined approaches to instruction that, based on the description, included coverage of conceptual understanding. In some studies, students were provided visual models so that students could use a visual representation to either compute or solve a word problem. Others used strategies that encouraged children to analyze word problem structure, so that meaningful patterns could emerge such as via explicit instruction or verbalizing. Mathematics educators have long been aware of the importance of developing an understanding of number operations, and patterns in problem solving, and this emphasis on meaningful understanding of operations is stressed in the NCTM *Focal Points*. In addition, research on mathematical learning in general, and mathematical disabilities and low-achievement, is associated with the nature of development of areas of number sense, including conceptual understanding of mathematical procedures and strategies for obtaining solution (Gersten & Chard, 1999; Hecht, Vagi, & Torgesen, 2007). Programs have been developed that orchestrate practice sets for each student and try to teach meaningful understandings of numbers and number operations. However, Fuchs et al. (2005) remind us that future work is needed to increase the power of classroom as well as tutorial treatments in low-achieving (at-risk) children.

1. Quality of Mathematics Taught in the Studies

In order to obtain an independent review of the quality of the actual mathematics taught in this set of studies, two research mathematicians involved in mathematics education, and one prominent mathematics educator were asked to examine the mathematical content and the nature of instruction (as opposed to the research design and technical details) of a small subset of studies. They looked at studies that described the mathematics content and instructional procedure with some amount of detail because we saw no benefit in, for example, asking a mathematician to evaluate a study in which it said, “Students learned the material in the third-grade mathematics state standards,” or studies that focused on very simple algorithms. Thus, we tended to choose the studies that bit off the most ambitious mathematical material and developed seemingly effective means for teaching the materials to students with learning problems and learning disabilities. We selectively summarize some results in this section.

The reader will recall that Woodward (2006) employed a combination of individualized fact practice with instruction that involved work with number families and applications of the distributive law to ease mental computation fluency. (For example, students learned that it is usually easier to calculate 8×9 by remembering that since 9 is the same as $10 - 1$, this problem has the same answer as $8 (10 - 1)$ or 8×10 minus 8×1. Or that 9×8 is the same as 8×9, so if you know one, you know the other because they are equivalent.)

The attempt to link computation to number properties is admirable, but several problems were noted. One mathematician observed that students should not be taught that 9×3 is the same as 3×9. They are, in fact two different problems with the same answer. One refers to 9 sets of 3 units, the other to 3 sets of 9. For students to succeed in algebra, they must understand this difference and remember that two things may look very different and represent very different type of problem types but still have the same answer. The work on multiplication combinations could have resulted in intense work on applications of the commutative, associative and distributive properties of numbers, but based on the text of the article, it did not appear to do so. The importance of doing so for students with LD and other students with learning problems is critical. In contrast, the treatment in Woodward and Brown (2006), developed by the same author, appeared to offer a much richer mathematical menu to students.

A similar concern was expressed about the pre-algebra material used in Witzel et al. (2003). Algebra was taught only on the procedural level. The importance of understanding the nature of a defining variable appeared to be underdeveloped, as did the potential richness of the concrete and visual representations and their link to sets of story problems. Similar concerns were raised about the CRA research of Butler et al. (2003), where numerous opportunities to explore rich mathematical ideas were lost.

These are among the more ambitious studies in the set reviewed, and among the few that really try to delve into complex mathematical topics and concepts. Each of the studies demonstrated some success in reaching the population. However, more intensive collaboration with research mathematicians who know the underlying mathematics in the K–8 curriculum can result in even richer, more effective intervention research for these students.

The research mathematicians also noted that although the two studies that attempted to teach story problems to students (Xin et al., 2005; Fuchs et al., 2005) did not really teach problem solving in the sense that NRC (2001) defined it. However, the studies seemed to be solid attempts to help students understand how to use the mathematics they already knew in an increasing array of applications.

2. Conclusions

On a positive note, many of the studies seriously address two areas of extreme difficulty for students with LD and low-achieving students: application of mathematics to word problems and building of quick retrieval of basic arithmetic combinations. These two areas are essential components of any serious mathematics intervention for these students and we now possess several evidence-based approaches for addressing these areas.

It becomes difficult to conclude easy generalizations about the set of studies. A terse summary would be that explicit instruction is effective (often highly effective) in both domains. In addition, more implicit instructional approaches such as strategic use of concrete objects and visual representations shows some promise, although the number of studies supporting this approach is small, and results are not consistent. Finally, approaches that encourage students to think aloud as they solve problems seem to produce significant positive effects. The drawback in these generalizations is that these terms mean different things to different people. Thus, in translating these findings into practice, effects may be highly dependent on how these instructional principles are conceptualized and how carefully they are incorporated into instruction.

One process that might ease the transition is that many of the interventions included specific scripts for teachers to use for their lessons (although they were usually told to use these as a guide rather than an ironclad script). These could serve as templates for lessons as districts develop various professional academies and training institutes.

Moreover, some other barriers to translating the research findings into classroom practice are as follows including the lack of sufficient specificity concerning the actual content of the mathematics instruction that was provided, which makes replication and extension of the current studies difficult. From a pragmatic standpoint, this is understandable given the need for authors to both describe the instructional sequence and content while leveraging page length. Also, the reviewed studies tend to use either a criterion-referenced test with items that are not presented or comprehensively defined or a standardized achievement test. A notable problem with standardized achievement tests is that they are composed of items from many domains of mathematics skill (e.g., basic computation, long division, fractions) and therefore tend to provide limited specificity concerning the actual mathematics content that students have mastered (Geary, 2005; Hecht, 1998). Finally, criteria for identifying and including low-achieving students examined in these studies were not consistent, which makes generalization of findings uncertain. Most of the studies utilize quite systematic instruction, with high degree of structure and a deliberate pace. This degree of explicitness and detail seems critical for this group of students. Our hope is that research and development efforts will continue to incorporate these elements into instructional materials that can be used with students with learning disabilities and low achievement in mathematics.

BIBLIOGRAPHY

Badian, N.A. (1983). Dyscalculia and nonverbal disorders of learning. In H.R. Myklebust (Ed.), *Progress in learning disabilities* (pp. 235–264). New York: Stratton.

Baker, S., Gersten, R., & Lee, D. (2002). A synthesis of empirical research on teaching mathematics to low-achieving students. *The Elementary School Journal*, 103, 51–73.

Barbaresi, W.J., Katusic, S.K., Colligan, R.C., Weaver, A.L., & Jacobsen, S.J. (2005). Math learning disorder: Incidence in a population-based birth cohort, 1976–82, Rochester, Minn. *Ambulatory Pediatrics*, *5*, 281–289.

Baxter, J., Woodward, J., Wong, J., & Voorhies, J. (2002). We talk about it, but do they get it? *Learning Disabilities Research and Practice*, *17(3)*, 173–185.

Beirne-Smith, M. (1991). Peer tutoring in arithmetic for children with learning disabilities. *Exceptional Children*, *57(4)*, 330–337.

Bereiter, C., & Engelmann, S. (1966). *Teaching disadvantaged children in the preschool*. Englewood Cliffs, NJ: Prentice Hall.

Bradley, R., Danielson, R.L., & D.P. Hallahan, D.P. (2002). *Identification of learning disabilities: Research to practice.* Mahwah, NJ: Erlbaum.

Brown, A.L., & Campione, C. (1990). Communities of learning and thinking: A context by any other name. *Contributions to Human Development, 21*, 108–126.

Butler, F.M., Miller, S.P., Crehan, K., Babbitt, B., & Pierce, T. (2003). Fraction instruction for students with mathematics disabilities: Comparing two teaching sequences. *Learning Disabilities Research & Practice, 18(20),* 99–111.

Cardelle-Elawar, M. (1995). Effects of metacognitive instruction on low achievers in mathematics problems. *Teaching and Teacher Education, 11(1),* 81–95.

Carnine, D.W., & Stein, M. (1981). Organizational strategies and practice procedures for teaching basic facts. *Journal of Research in Mathematics Education, 12(1),* 65–69.

Darch, C., Carnine, D., Gersten, R. (1984). Explicit instruction in mathematics problem solving. *Journal of Educational Research, 77(6),* 351–359.

DiPerna, J.C., Lei, P., & Reid, E.E. (2007). Kindergarten predictors of mathematical growth in the primary grades: An investigation using the early childhood longitudinal study-kindergarten cohort. *Educational Psychology, 99(2),* 369–379.

Donovan, M.S., & Bransford, J.D. (2005). Introduction. In the National Research Council (Ed.), *How students learn: Mathematics in the classroom* (pp. 1–30). Washington, DC: The National Academy Press.

Engelmann, S., & Carnine, D.W. (1982). *Theory of instruction: Principles and applications*. New York: Irvington.

Friedman, F. (1992). An evaluation of the integrated cognitive-behavioral model for improving mathematics performance and attentional behavior of adolescents with learning disabilities and attention-deficit hyperactivity disorder. Unpublished doctoral dissertation, Columbia University Teachers College.

Fuchs, D., Fuchs, L.S., Mathes, P.G., & Lipsey, M.W. (2000). Reading differences between low-achieving students with and without learning disabilities: A meta-analysis. In R. Gersten & S. Vaughn (Eds.*), Issues and research in special education* (pp. 81–104). Hillsdale, NJ: Erlbaum.

Fuchs, L.S., Compton, D.L., Fuchs, D., Paulsen, K., Bryant, J.D., & Hamlett, C.L. (2005). Responsiveness to intervention: Preventing and identifying mathematics disability. *Teaching Exceptional Children*, *37(4)*, 60–63.

Fuchs, L.S., Compton, D.L., Fuchs, D., Paulsen, K., Bryant, J.D., & Hamlett, C.L. (in press). The prevention, identification, and cognitive determinants of math difficulty. *Journal of Educational Psychology*.

Fuchs, L.S., & Fuchs, D. (1998). Peer-assisted learning strategies in math for kindergarten through grade 6. Unpublished article available from Lynn Fuchs, 328 Peabody, Vanderbilt University, Nashville, TN 37220 (lynn.fuchs@vanderbilt.edu).

Geary, D.C. (1993). Mathematical disabilities: Cognitive, neuropsychological, and genetic components. *Psychological Bulletin, 114(2)*, 345–362.

Geary, D.C. (2003). Learning disabilities in arithmetic: Problem solving differences and cognitive deficits. In H.L. Swanson, K. Harris, & S. Graham (Eds.), *Handbook of learning disabilities* (pp. 199–212). New York: Guilford Press.

Geary, D.C. (2005). Role of cognitive theory in study of learning disability in mathematics. *Journal of Learning Disabilities, 38*, 305–307.

Geary, D.C., Hoard, M.K., Byrd-Craven, J., Nugent, L., & Numtee, C. (2007). Cognitive mechanisms underlying achievement deficits in children with mathematical learning disability. *Child Development, 78(4)*, 1343–1359.

Gersten, R., & Carnine, D. (1984). Direct instruction mathematics: A longitudinal evaluation of low-income elementary school students. *Elementary School Journal, 84(4)*, 395–407.

Gersten, R., & Chard, D. (1999). Number sense: Rethinking arithmetic instruction for students with mathematical disabilities. *Journal of Special Education, 33(1)*, 18–29.

Gersten, R., Clarke, B., & Mazzocco, M. (2007). Historical and contemporary perspectives on mathematical learning disabilities. In D. B. Berch, & M. M. Mazzocco (Eds.), *Why is math so hard for some children? The nature and origins of mathematical learning difficulties and disabilities* (pp. 7–29). Baltimore, MD: Paul H. Brookes Publishing Company.

Gersten, R., Jordan, N.C., & Flojo, J. (2005). Early identification and interventions for students with mathematics difficulties. *Journal of Learning Disabilities, 38(4)*, 293–304.

Goldman, S.R., & Pellegrino, J.W. (1987). Information processing and educational microcomputer technology: Where do we go from here? *Journal of Learning Disabilities, 20(3)*, 144–154.

Good, T., & Grouws, D. (1977). Teaching effects: A process-product study in fourth grade mathematics classrooms. *Journal of Teacher Education, 28(3)*, 49–54.

Griffin, S., Case, R., & Siegler, R. (1994). Rightstart: Providing the central conceptual prerequisites for first formal learning of arithmetic to students at risk for school failure. In K. McFilly (Ed.), *Classroom Lessons: Integrating cognitive theory and classroom practices* (pp. 25–49). Cambridge: MIT/Bradford Books.

Gross-Tsur, V., Manor, O., & Shalev, R. S. (1996). Developmental dyscalculia: Prevalence and demographic features. *Developmental Medicine and Child Neurology, 38,* 25–33.

Hallahan, D.P., & Kauffman, J.M. (1986). *Exceptional children* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.

Hasselbring, T.S., Goin, L.I., & Bransford, J.D. (1988). Developing math automaticity in learning handicapped children: The role of computerized drill and practice. *Focus on Exceptional Children, 20(6),* 1–7.

Hasselbring, T.S., Lott, A., & Zydney, J. (2005). Technology supported math instruction for students with disabilities: Two decades of research and development. Retrieved on November 1, 2007 from http://www.citeducation.org/mathmatrix/default.asp.

Hecht, S.A. (1998). Toward an information processing account of individual differences in fraction skills. *Journal of Educational Psychology, 90(3),* 545–559.

Hecht, S.A., Vagi, K.J. & Torgesen, J.K. (2007). Fraction skills and proportional reasoning. In Daniel B. Berch & Michele M.M. Mazzocco (Eds.), *Why is math so hard for some children?* Baltimore, MD: Brookes Publishing Company.

Hoge, R.D., & Coladarci, T. (1989). Teacher-based judgments of academic achievement: A review of literature. *Review of Educational Research, 59(3)*, 297–313.

Hutchinson, N.L. (1993). Effects of cognitive strategy instruction on algebra problem solving of adolescents with learning disabilities. *Learning Disability Quarterly, 16(1),* 34–63.

Jitendra, A.K., Griffin, C.C., McGoey, K., Gardill, M.C., Bhat, P., & Riley, T. (1998). Effects of mathematical word problem solving by students at risk or with mild disabilities. *The Journal of Educational Research, 91(6)*, 345–355.

Jordan, N.C., Hanich, L.B., & Kaplan, D. (2003). A longitudinal study of mathematical competencies in children with mathematics difficulties versus children with co-morbid mathematics and reading difficulties. *Child Development, 74(3*), 834–850.

Kolligian, J., & Sternberg, R.J. (1987). Intelligence, information processing, and specific learning disabilities: A triarchic synthesis. *Journal of Learning Disabilities, 20(1),* 8–17.

Kroesbergen, E.H., Van Luit, J.E.H., & Maas, C.J.M. (2004). Effectiveness of explicit and constructivist mathematics instruction for low-achieving students in the Netherlands. *The Elementary School Journal, 104*(3), 233–251.

Lambert, M.A. (1996). Teaching students with learning disabilities to solve word-problems: A comparison of a cognitive strategy and a traditional textbook method. Unpublished doctoral dissertation, Florida Atlantic University.

Lewis, C., Hitch, G.J., & Walker, P. (1994). The prevalence of specific arithmetic difficulties and specific reading difficulties in 9- to 10-year-old boys and girls. *Journal of Child Psychology & Psychiatry & Allied Disciplines, 35(2),* 283–292.

Manalo, E., Bunnell, J.K., & Stillman, J.A. (2000). The use of process mnemonics in teaching students with mathematics learning disabilities. *Learning Disability Quarterly, 23*(2), 137–156.

Mazzocco, M.M.M. (2007). Defining and differentiating mathematical learning disabilities and difficulties. In D. Berch, & M.M.M. Mazzocco (Eds.), *Why is math so hard for some children: The nature and origin of mathematics learning difficulties and disabilities* (pp. 29–47). Baltimore: Brookes Publishing Co.

McLean, J.F., & Hitch, G.J. (1999). Working memory impairments in children with specific arithmetic learning difficulties. *Journal of Experimental Child Psychology, 74(3)*, 240–260.

Meichenbaum, D. (1985). Teaching thinking: A cognitive-behavioral perspective. In S. Chipman, J. Segal, & R. Glaser (Eds.), *Thinking and learning skills (Vol. 2) Research and open questions* (pp. 407–428). Hillsdale, NJ: Lawrence Erlbaum Associates.

Milo, B.F., Ruijssenaars, A.J.J.M., & Seegers, G. (2005). Math instruction for students with special education needs: Effects of guiding versus directing instruction. *Educational and Child Psychology, 22(4)*, 68–80.

Montague, M. (1992). The effects of cognitive and metacognitive strategy instruction on the mathematical problem solving of middle school students with learning disabilities. *Journal of Learning Disabilities, 25(4)*, 230–248.

Moore, L.J., & Carnine, D. (1989). Evaluating curriculum design in the content of active teaching. *Remedial and Special Education, 10(4)*, 28–37.

Murphy, M.M., Mazzacco, M., Hanich, L.B., & Early, M.C. (2007). Cognitive characteristics of children with mathematics learning disability (MLD) vary as a function of the cutoff criterion used to define MLD. *Journal of Learning Disabilities, 40(5)*, 458–478.

National Council of Teachers of Mathematics. (2006). *Curriculum focal points for prekindergarten through Grade 8 mathematics: A quest for coherence*. Reston, VA: National Council of Teachers of Mathematics.

National Council of Teachers of Mathematics. (2000). *Principles and NCTM standards for school mathematics.* Reston, VA: Author.

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: National Council of Teachers of Mathematics.

National Research Council. (2001). *Adding it up: Helping children learn mathematics*. J. Kilpatrick, J. Swafford, and B. Findell (Eds.). Mathematics learning study committee, center for education, division of behavioral and social sciences and education. Washington, DC: National Academy Press.

Owen, R.L. & Fuchs, L.S. (2002). Mathematical problem-solving strategy instruction for third-grade students with learning disabilities. *Remedial and Special Education, 23(5)*, 268–278.

Pasnak, R., McCutcheon, L., Holt, R.W., & Campbell, J.W. (1991). Cognitive and achievement gains for kindergartners instructed in Piagetian operations. *Journal of Educational Research, 85(1)*, 5–13.

Riley, M.S., Greeno, J.G., & Heller, J.J. (1983). Development of children's problem-solving ability in arithmetic. In H.P. Ginsburg (Ed.), *The development of mathematical thinking* (pp. 153–196). New York: Academic Press.

Rittle-Johnson, B., Siegler, R.S., & Alibali, M. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of Educational Psychology, 93(2)*, 346–362.

Ross, P.A, & Braden, J.P. (1991). The effects of token reinforcement versus cognitive behavior modification on learning-disabled students' math skills. *Psychology in the Schools, 28(3)*, 247–256.

Schopman, E.A.M., & Van Luit, J.E.H. (1996). Learning and transfer of preparatory arithmetic strategies among young children with a developmental lag. *Journal of Cognitive Education, 5*, 117–131.

Schunk, D.H., & Cox, P.D. (1986). Strategy training and attributional feedback with learning disabled students. *Journal of Educational Psychology, 78(3)*, 201–209.

Siegler, R.S. (1987). The perils of averaging data over strategies: An example from children's addition. *Journal of Experimental Psychology: General, 116(3)*, 250–264.

Siegler, R.S., & Shrager, J. (1984). Strategy choice in addition and subtraction: How do children know what to do? In C. Sophian (Ed.), *Origins of cognitive skills* (pp. 229–293). Hillsdale, NJ: Erlbaum.

Silbert, J., Carnine, D., & Stein, M. (1989). *Direct instruction mathematics* (2nd ed.). Columbus, OH: Merrill.

Sowder, J.T. Philipp, R.A., Armstrong, Barbara E., & Schappelle, B.P. (1998). *Middle-grade teachers' mathematical knowledge and its relationship to instruction: A research monograph.* Albany, NY: State University of New York Press.

Stebbins, L., St. Pierre, R.G., Proper, E.L., Anderson, R.B., & Cerva, T.R. (1977). *Education as experimentation: A planned variation model* (Vol. IV A–D). Cambridge, MA: Abt Associates.

Swanson, H.L., & Hoskyn, M. (1998). Experimental intervention research on students with learning disabilities: A meta-analysis of treatment outcomes. *Review of Educational Research, 68(3)*, 277–321.

Swanson, H.L., & Sachse-Lee, C. (2001). Mathematical problem solving and working memory in children with learning disabilities: Both executive and phonological processes are important. *Journal of Experimental Child Psychology, 79(3)*, 294–321.

Thackwray, D., Meyers, A., Schleser, R., & Cohen, R. (1985). Achieving generalization with general versus specific self-instructions: Effects on academically deficient children. *Cognitive Therapy and Research, 9(3)*, 297–308.

Torgeson, J. (1994). Issues in the assessment of executive function: An information processing perspective. In G. R. Lyon (Ed.), *Frames of reference in the assessment of learning disabilities: New views on measurement issues*. (pp. 143–162). Toronto: Brooks.

Tournaki, N. (2003). The differential effects of teaching addition through strategy instruction versus drill and practice to students with and without learning disabilities. *Journal of Learning Disabilities, 36(5)*, 449–458.

Van Luit, J.E.H., & Naglieri, J.A. (1999). Effectiveness of the MASTER program for teaching special children multiplication and division. *Journal of Learning Disabilities, 32(2)*, 98–107.

Walker, D.W., & Poteet, J.A. (1989–1990). A comparison of two methods of teaching mathematics story problem-solving with learning disabled students. *National Forum of Special Education Journal, 1,* 44–51.

Wilson, C.L., & Sindelar, P.T. (1991). Direct instruction in math word problems: Students with learning disabilities. *Exceptional Children, 57(6),* 512–519.

Witzel, B.S., Mercer, C.D., & Miller, M.D. (2003). Teaching algebra to students with learning difficulties: An investigation of an explicit instruction model. *Learning Disabilities Research and Practice, 18(2),* 121–131.

Woodward, J. (2006). Developing automaticity in multiplication facts: Integrating strategy instruction with timed practice drills. *Learning Disability Quarterly, 29(4),* 269–289.

Woodward, J., & Baxter, J. (1997). The effects of an innovative approach to mathematics on academically low achieving students in inclusive settings. *Exceptional Children, 63(3),* 373–388.

Woodward, J., & Brown, C. (2006). Meeting the curricular needs of academically low-achieving students in the middle grade mathematics. *Journal of Special Education, 40(3),* 151–159.

Woodward, J., & Howard, L. (1994). Misconceptions of youth. *Exceptional Children, 61(2),* 126–136.

Woodward, J., & Montague, M. (2002). Meeting the challenge of mathematics reform for students with learning disabilities. *Journal of Special Education, 36(2),* 89–101.

Xin, Y.P., Jitendra, A.K., & Deatline-Buchman, A. (2005). Effects of mathematical word problem-solving instruction on middle school students with learning problems. *The Journal of Special Education, 39(3),* 181–192.

IV. 'Real-World' Problem Solving

A. Introduction and Background

Discussion and debate about the place of "real-world" problems in mathematics instruction—both as a site for learning mathematics, and as an outcome—has been a central theme in U.S. mathematics education for more than a century. The earliest goals of mathematics instruction in this country related to practical uses of mathematics, in such areas as shopkeeping, commerce, surveying, banking, and carpentry (see Michalowicz & Howard, 2003). Early U.S. mathematics textbooks are filled with practical problems intended to prepare students to enter the workplace and to address the particular application needs of the growing nation and society. Over the decades this emphasis on practical applications has ebbed and flowed. In the "new math" years (1960s and 70s), the predominant curricular emphasis was on mathematical precision and the structure of mathematics, but, even then, some critics, such as the applied mathematician Morris Kline (1973), continued to call for applications in the school curriculum. Indeed, even amidst the abstract and logic-focused new mathematics materials developed during the post-Sputnik era, there was at least one applications-oriented curriculum, the Unified Science and Mathematics for Elementary Schools (USMES) project.[3]

A resurgence of calls for emphasis on "real-world"[4] problems came in the 1989 Curriculum and Evaluation Standards for School Mathematics, of the National Council of Teachers of Mathematics (NCTM), which argued that "instruction should be developed from problem situations" (NCTM, 1989, p. 11). The document recommends that in the early grades (K–4), most problems used in instruction should arise from "school and other everyday experiences" (p. 23). Progressing through the grades, there should be a balance between "problems that apply mathematics to the "real-world" and problems that arise from the investigation of mathematical ideas" (p. 75), and by high school, even more of the problems can arise from mathematics itself.

The Instructional Practices Task Group begins by summarizing the definitions and operational meanings that researchers and developers have given to the term "real-world" problems. In the next section the Task Group will highlight some of the rationale and justifications that researchers and developers have used when arguing for and against particular characteristics of "real-world" problems and their uses as a part of the school mathematics curriculum. These viewpoints are sometimes based in research, and sometimes are more directly tied to experience and expert judgments. They sometimes relate to the question of how important it is for students to be able to apply their mathematical knowledge to particular types of problems as an outcome of schooling is. Finally, a synthesis of the

[3] Launched in 1970 through the Education Development Center, with funding from the National Science Foundation; described as a project designed so "students could carry out long-range investigations of real and practical problems based in their local environment" (http://www.coe.ufl.edu/esh/Projects/usmes.htm).
[4] We will include "real world" in quotation marks throughout because of the ambiguity of definition of this phrase in the current literature.

findings from 21 studies examining the impact of "real-world" problem-based instruction on mathematics learning outcomes is provided, followed by commentary about issues in this body of research, and recommendations.

B. What Do Researchers and Developers Mean by 'Real-World' Problems?

Here the Task Group draws on the conceptualizations of "real-world" problems assembled from both developers of instructional materials and researchers who study the impact of such problems. A serious problem in synthesizing the research in this area is that there is no clear, agreed-upon meaning for "real-world" problems. One characteristic mentioned frequently is the meaningfulness and relevance of the problem to the student audience. The Realistic Mathematics Education (RME) movement, which originated in the Netherlands through the work of the mathematician Hans Freudenthal (1973, 1991), has been influential in school mathematics in some countries. RME emphasizes relevance and the activity of doing mathematics; there are several lines of research that have a basis in the RME movement. For instance, researchers De Bock, Verschaffel, Janssens, Van Dooren, and Claes (2003) discuss "organizing mathematical activities around rich, attractive, and realistic contexts … [not only] aspects of the 'real' social or physical world; they can also refer to imaginary, fairy-like worlds as long as they are meaningful, familiar, and appealing to the students. It is not the amount of realism in the literal sense … but rather the extent to which it succeeds in getting students involved in the problem and engage them in situationally meaningful thinking and interaction" (p. 445). In the same tradition, van Dijk and others (2003) describe problems that "bring pupils into situations that make sense to them and provide them with opportunities to experience mathematics as it was developed in cultural history" (p. 164). Other scholars feel that "real-world" problems should be similar to problems that are encountered in applications beyond school, and that are authentic, for instance problems that are "…embedded in a rich narrative structure" and that may require students to make both mathematical and nonmathematical (e.g., ethical) decisions.

In contrast, the problems typically found in algebra textbooks that are sometimes called "story problems" or "word problems" also are sometimes studied in efforts to look at "real-world" problem solving. Jonassen (2003), for example, defines story problems as those that "typically present a quantitative solution problem embedded within a shallow story context" (p. 267).

A "real-world" oriented curriculum that has been studied by a number of the researchers cited in this paper, and whose authors provide detail about their conceptualization of "real-world" problems, is the Adventures of Jasper Woodbury video series (http://peabody.vanderbilt.edu/projects/funded/jasper/Jasperhome.html). This technology-based series is designed to motivate students by engaging them in the solution of complex, multistep problems. The goals of the materials are to promote problem-finding and to develop problem-solving skills. Each 15–20 minute video segment presents an adventure story which involves solving a challenge. The Cognition and Technology Group at Vanderbilt (CTGV) identifies "real-world" problems as being complex, which means having multiple steps, requiring integration of mathematical concepts, involving identification of relevant data, and demanding generation of appropriate questions (see CTGV, 1992; Hickey et al., 2001, pp. 613–14). The goal of this type of problem is to allow students to experience some of the ambiguity and complexity, as well as the intellectual excitement, that adults experience when solving actual problems involving mathematics, be it in engineering, business, accounting, architecture, transportation planning, etc. The hope is that, by anchoring mathematical procedures and concepts in an array of actual situations, students will see the value of knowing the procedures and will more likely be able to transfer what they learn in mathematics to actual problems. The terms "anchored instruction," "situated cognition," and "teaching for transfer" often recur in this literature.

In summary, note how diverse these meanings of "real-world" problem solving are in the literature. This creates challenges and opportunities for researchers, who in general could make progress on some of the fundamental "real-world" problem solving questions with more clarity and focus in the operationalization of the terminology.

What are the purported advantages and disadvantages of using various types of "real-world" problems in school mathematics instruction?

There are several related but distinct reasons advanced in both research and other educational rhetoric for including "real-world" problems in the school mathematics curriculum. Those who believe that students' ability to solve "real-world" problems should be an important outcome of school mathematics argue for the inclusion of such problems in the curriculum as preparation. Others contend that "real-world" problems should be in the curriculum because of their potential to engage and motivate students by engaging them in something they see as meaningful and important (see Bransford, Sherwood, Hasselbring, Kinser, & Williams, 1990; Bransford, Vye, Kinser, & Risko, 1990; CTGV, 1991; DeBock et al., 2003). Hiebert et al. (1996) comment that the "mathematics acquired in these realistic situations, proponents argue, will be perceived by students as being useful" (p. 14).

Another reason sometimes given is that such problems, especially when assigned to be done in groups, provide students with opportunities to learn some of the social problem-solving skills they will need to use later in the workplace (see Resnick, 1987b). The Vanderbilt group (CTVG) contends that "'anchor[ing]' or 'situat[ing]' instruction in the context of meaningful problem solving environments … allow[s] teachers to simulate in the classroom some of the advantages of 'in-context' apprenticeship training" (CTGV, 1992, p. 294; also citing Brown et al., 1989).

Finally, there is a view that students' learning and ability to make mathematical connections in the process of applying their knowledge to a wider range of "real-world" problems might be enhanced (see Hiebert et al., 1996). The notions of both near and far transfer in problem solving appear later in the literature synthesis.

Common to these views seems to be the assumption that, by teaching students mathematics through "real-world" problems, and by teaching students to solve such problems in school, students will become better solvers of the types of problems that they might encounter in everyday life or the workplace (see Verschaffel & De Corte, 1997), and that they will develop a genuine disposition to, and interest in, solving such problems. Some research has looked specifically at whether it is indeed the case that the use of “real-world” problems in instruction promotes such outcomes.

Anderson, Reder, and Simon (1996) have countered some of the claims attributed to the proponents of the use of “real-world” problems.[5] They claim that research has indicated that transfer can happen even if students learn in a situation that is not specific to the site of application. They argue further that using such problems can be inefficient: “Often real-world problems involve a great deal of busy work and offer little opportunity to learn the target competencies” (Anderson et al., 1996, p. 9). They also note that research indicates that workplace skills can be learned separately from the social context, and that in some cases they should be.

There is also the question of whether the contexts that developers imagine as being motivational and engaging for students actually are. Some researchers (e.g. Geary, 1995) have suggested that the contexts in which problems are offered may not be that intrinsically motivating to students. Geary emphasizes that making the mathematics interesting and also ensuring that adequate mathematics is learned may require “degrading” the mathematical content in ways that are not satisfactory. Hiebert et al. (1996) advocate the importance of students’ “problematizing” mathematics and suggest that the particular context chosen for a problem is not necessarily as important as the way the teacher engages the students: “Given a different culture [valuing reflective inquiry and problematizing], even large-scale real-life situations can be drained of their problematic possibilities. Outside-of-school problems can provide contexts for important mathematical work, but the packaging of the task is not the primary determinant of the engagement” (pp. 16–18).

“Real-world” problems are often considered to be “open-ended,” a term that is equally ill-specified in its meaning. Pehkonen (1997) provides some history of the idea of “openness” in mathematics education, citing work initiated in Japan in the 1970s that helped to launch international focus. He defines open problems in contrast to “closed problems” in mathematics, in which in “open problems,” the starting situation or the goal is not explained exactly. Such problems then encompass problem situations in which the student must “find” the problem, “what if” problems, and problems in which multiple solution processes are possible. In a 1994 essay, Hung-Hsi Wu uses examples of problems from K–12 mathematics curricula to highlight that in the case of some open-ended problems, teachers are unlikely to know the required mathematics deeply and at the same time provide a suitably simplified explanation to students. Others have raised concerns about the adequacy of teachers’ knowledge of the nonmathematical contexts—in which some of these problems are embedded—to assess the reasonableness of the problem’s assumptions, and about the efficiency of using elaborate “real-world” problems in covering mathematics content.

In summary, even without a consistent definition of the notion of “real-world” problems, there are strongly held and argued positions, founded on a variety of bases, in support of, and critical of, the use of various types of “real-world” problems in school mathematics instruction.

[5] For replies and additional commentary concerning the Anderson et al. (1996) paper, see Greeno (1997) and Cobb and Bower (1999).

C. Research Studies Examining the Impact of 'Real-World' Problems in Mathematics Instruction

Despite the variety of reasons that have been advanced supporting the inclusion of "real-world" problems in mathematics instruction, the available research on the topic that met the Instructional Practices Task Group standards for inclusion addresses only two rather focused types of questions:

- Does the use of "real-world" problems in mathematics instruction, in comparison to typical instructional practice, lead to improved understanding of mathematical ideas, or improved computational performance, or improved mathematics performance? (Using "Real-World" Problems to Teach Mathematical Ideas).

- Does the use of particular instructional strategies to help students learn to solve "real-world" problems, in comparison to other strategies and to typical instructional practice, lead to improved performance on assessments that involve solving "real-world" problems; i.e., can near and far transfer be achieved? (Using Specific Strategies to Improve "Real-World" Problem Solving).[6]

Researchers from cognitive science, psychology, and mathematics education have undertaken a range of studies that examine phenomena related to "real-world" problems in mathematics teaching and learning. Most of this work has been descriptive and is not included in the meta-analytic discussion to follow. However, it could serve as an important basis for clarifying and disentangling the meanings of "real-world" problems as an instructional approach and as an outcome of schooling, and provide insights into the design of interventions and assessments that are focused on "real-world" problems. Ethnographic studies have looked, for instance, at the problem-solving strategies used in practices such as candy-selling, tailoring, carpentry, gardening, etc., and the relationship of such craft knowledge to performance on school-based problem tasks (see Presmeg, 2007). International studies such as the Programme for International Student Assessment (PISA) study provide a snapshot of U.S. students' performance on problem solving. This Organisation for Economic Co-operation and Development (OECD) initiative is a collaborative effort of the OECD member countries to "measure how well students at age 15, and therefore approaching the end of compulsory schooling, are prepared to meet the challenges of today's societies… moving beyond the school based approach towards the use of knowledge in everyday tasks and challenges" (Programme for International Student Assessment, 2003, p. 9). PISA is unique as an international assessment in its explicit effort to assess students' ability to "apply their knowledge and experience to real-world issues" (Programme for International Student Assessment, 2003, p. 9). The 2003 administration of PISA examined mathematical literacy and problem solving, and the performance of U.S. students was lower than the average performance for students from OECD countries (see Lemke et al., 2004). Thus for those

[6] The initial search for and screening of literature was initially only for the use of "real world" problems in instruction (question 1), not as the outcome (question 2). The literature identified for the second question resulted from the first, but we did not go back and systematically search for other studies that might fit this second question.

concerned that ability of U.S. students to solve "real-world" problems is an important outcome of schooling, studies such as PISA indicate that U.S. performance has substantial room for improvement.

Here the Task Group draws on the 21 studies that qualified as Category 1 or Category 2 studies and that examined, with random assignment or quasi-experimental methods, the impact of some type of "real-world" problem instructional intervention on student mathematics learning outcomes. There were 13 studies (five Category 1 and eight Category 2) that examined the effect of using "real-world" problems as the means of instruction on mathematics achievement. An additional eight studies (five Category 1 and three Category 2) examined specific strategies to solve "real-world" problems. The second group did not examine a "real-world" problem instructional intervention. Although many have argued that a major reason for using "real-world" problems in mathematics instruction is to increase interest and motivation, the Task Group did not search for studies that looked at motivation only as an outcome. For those studies of mathematics achievement that did include a motivation outcome, those outcomes are not discussed. The studies are presented according to the two categories mentioned earlier. See Tables 15 and 16 for a summary of the Category 1 studies, including effect size calculations.

D. Using 'Real-World' Problems to Teach Mathematical Ideas

The Task Group located 13 studies that introduced some version of a "real-world" problem instructional treatment, and that compared outcomes on student performance in mathematics. Five of these can be considered Category 1 studies for which effect sizes could be computed. Four of these studies contrast some type of "real-world" problem-based instruction with more typical mathematics instruction (although even this varies to some degree). The fifth is concerned with contrasting two different approaches to using "real-world" problems as an instructional approach. All employ outcome measures that assess mathematics performance on what might be considered "typical" types of school mathematics outcomes. In addition, some include outcome measures for transfer, or involving contextualized problems.

Three of the Category 1 studies in this area focus on computations with fractions. Anand and Ross (1987) developed three versions of an intervention aimed at teaching fifth- and sixth-graders how to divide fractions. The operationalization of "real-world" problem solving involved two ways of contextualizing problems: by adding such student-specific information as name, favorite candy bar, etc. into problems, or by simply providing a concrete context for a computational problem. The intervention was a CAI unit that included a "review of prerequisite mathematics facts… introduced the rule for dividing fractions and demonstrated its application to an example problem by using the following four-step solution.… This rule application was repeated for four additional problems" (p. 73). The treatments varied by changing the contexts for the learning material; there were abstract contexts provided, concrete contexts, and personalized contexts based on a biographical questionnaire for the students. The posttest involved context problems similar to those presented in the practice examples, transfer problems, and recognition memory of the rule definition and steps problems. Ninety-six students were randomly assigned to the four treatment conditions (control; concrete; personalized; abstract).

Table 15: Studies That Examine Use of "Real-World" Problems in Mathematics Instruction

Study	Design	Sample	Duration/Content	Contrast	Measure	Hedge's g	Standard Error	
Contextualized Mathematics Outcomes								
Anand & Ross, 1987	RCT	96 students in fifth or sixth grade attending a university affiliated elementary school that emphasized individual learning and progression	One lesson/ Division of fractions	Concrete and Personalized vs. Abstract	Transfer subtest	0.379	(ns)	0.249
Bottge & Hasselbring, 1993	RCT	36 students in two ninth-grade remedial mathematics classes in one Midwest high school	5 days/ Adding and subtracting fractions in relation to money and linear measurement	Contextualized problems vs. Word problems	Contextualized problem test	1.009	**	0.385
Bottge, 1999[a]	RCT	49 middle school average-achieving students in two intact pre-algebra classes	10 school days/ Story problems and transfer problems involving fraction computation	Contextualized problems vs. Word problems	Contextualized problem test	1.131	(ns)	0.693
Brenner et al., 1997[a]	RCT	128 seventh- and eighth-grade students in six intact pre-algebra classes at three junior high schools in a small urban area in Southern California	1 month/ Meaningful thematic contexts and other features	Anchored instruction vs. Traditional textbook	Pooled word problem representation (ES = -0.281), function word problem representation (ES = 0.877), and function word problem solution (ES = 0.393) tests	0.631	(ns)	0.402
Heterogeneity								
Q-value	*df (Q)*	*P-value*	*I-squared*	*Pooled ES (4 studies, 4 effect sizes)*		*Hedge's g*	*Standard Error*	
2.499	*3*	*0.475*	*0.000*			*0.616*	*****	*0.179*
Standard Mathematics Outcomes								
Anand & Ross, 1987	RCT	96 students in fifth or sixth grade attending a university affiliated elementary school that emphasized individual learning and progression	One lesson/ Division of fractions	Concrete and Personalized vs. Abstract	Pooled context (ES = 0.931***) and recognition (ES = 0.727**) subtests	0.828	**	0.257
Bottge & Hasselbring, 1993	RCT	36 students in two ninth-grade remedial mathematics classes in one Midwest high school	5 days/ Adding and subtracting fractions in relation to money and linear measurement	Contextualized problems vs. Word problems	Word problem test	-0.553	(ns)	0.368
Bottge, 1999[a]	RCT	49 middle school average-achieving students in two intact pre-algebra classes	10 school days/ Story problems and transfer problems	Contextualized problems vs. Word problems	Pooled computation (ES = 0.049) and word problem tests (ES = -0.198)	-0.124	(ns)	0.683
Brenner et al., 1997[a]	RCT	128 seventh- and eighth-grade students in six intact pre-algebra classes at three junior high schools in a small urban area in Southern California	1 month/ Meaningful thematic contexts and other features	Anchored instruction vs. Traditional textbook	Pooled equation solving (ES = -0.281) and word problem solving (ES = 0.110) tests	-0.086	(ns)	0.399
Heterogeneity								
Q-value	*df (Q)*	*P-value*	*I-squared*	*Pooled ES (four studies, four effect sizes)*		*Hedge's g*	*Standard Error*	
10.835	*3*	*0.013*	*72.312*			*0.066*	*(ns)*	*0.374*
A Study that Examined Two Different Approaches to "Real-world" Problem-based Instruction								
van Dijk et al., 2003[a]	RCT	238 fifth-grade students in 10 classes in the Netherlands	13 lessons in 3 weeks/ "Real-world" problems that entail division with a remainder	Student vs. Teacher constructed models	Curriculum specific posttest	0.402	(ns)	0.307

~ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

[a] Data were adjusted for clustering that occurred within classrooms.

Several contrasts (comparisons of results for two different treatment conditions) resulted in statistically significant and meaningful effect sizes: most striking was the significantly better performance of the personalized group than the abstract group[7] on the context and recognition measures (effect sizes = 1.434 and 1.116, respectively). The concrete and personalized treatment group also performed significantly better than the abstract treatment group on the context and recognition measures (effect sizes = 0.931 and 0.727, respectively). Of the nine effects computed for this study,[8] five produced effect sizes significant at the .05 level or better. All favored the personalized or personalized and concrete treatments over the abstract, with the strongest differences on the context and recognition outcome measures. Effect sizes on the transfer measure for the personalized group in comparison with the abstract group also was significant, with an effect size of 0.630, and combining the concrete and personalized treatments and comparing to the abstract also yielded an appreciable though non-significant effect size (0.379) on the transfer measure. For all of these measures the combined effect size is 0.679, which is significant.

The Task Group notes that the use of "contextualized" in the Anand and Ross study is a narrowly focused operationalization of "real world." "Contextualization" occurred by personalizing the problems through such means as using the students' names or interests within the problems. This would not fit most of the operational meanings for "real-world" problems discussed earlier. Nonetheless, the strong effects on context and recognition problems is interesting, and suggest that a very specifically focused type of contextualization can be more effective on context and recognition outcomes than abstract presentation of problems, and can have some effect on transfer.

Bottge and his colleagues have published two studies that met the Category 1 criteria (Bottge & Hasselbring, 1993; Bottge, 1999). Both studies pursue questions about the effect of "contextualized mathematics instruction" on the problem-solving performance of middle school and ninth grade students. The instructional interventions are video-based problem solving materials based on the principles that guided the Cognition and Technology Group at Vanderbilt (CTGV) in the design of the *Adventures of Jasper Woodbury series*. These include commitment to "guidance by an effective teacher; a rich, realistic source of information; and a meaningful problem-solving context" (Bottge & Hasselbring, 1993, p. 5).

In the 1993 study, 36 students in two ninth-grade remedial mathematics classes were assigned to treatment and control conditions, where the instruction was focused on problem solving in the area of fraction addition and subtraction. The students had experienced behavioral or academic difficulties. Students were compared on their ability to solve a contextualized problem following instruction. All students received review in fraction computation skills for five days prior to the intervention. The intervention was then an additional five days of problem solving that employed, for the "contextualized problem" (CP) group, an 8-minute contextualized problem presented via videodisc called *Bart's Pet Project*. The "word problems" (WP) condition received a series of standard word problems in instruction. In both conditions the students were guided to solve the problems by their teachers.

[7] Contrasts not included in table because of the need to use only independent contrasts in pooling effect sizes from multiple contrasts.
[8] There were three contrasts, each with three measures. Only one is included in the table. See above footnote.

Effect sizes were computed on two outcome measures; a word problem test and a contextualized problem-solving test administered via video. The effect of the CP condition in contrast with the WP condition on the contextualized problems was significant (ES = 1.009). The effect size on the word problem measure, perhaps a more typical school mathematics outcome measure, was not significant and favored the control group (ES = -0.553). It is possible that the video-based contextual outcome measure was overly aligned with the treatment and the word problem measure was overly aligned with the control, which makes the results unsurprising. Nonetheless, the study seems to demonstrate that with this particular group of at-risk students, the “real-world” treatment can make a difference on the transfer task.

In 1999, Bottge again looked at the effect of contextualized mathematics instruction on the problem-solving performance of middle school students. The topic was fraction computation, and the interventions involved two video-based contextualized mathematics problems, both in the spirit of the CTGV-designed *Adventures of Jasper Woodbury*. The control treatment was a more standard presentation of word problem instruction, using problems parallel to those in the video materials. Outcome measures included computation, word problems, a contextualized problem, and an applied transfer task. There was a noteworthy but nonsignificant effect size (1.131), favoring the treatment groups, on the contextualized problem. It is worth noting that on the computational outcome there was a slight but not statistically significant advantage for the control group (ES = -0.049) and similarly, on the word problem outcome measure, the nonsignificant effect size (-0.198) slightly favored the control group. Note also that the mathematical content of this instruction is not standard ninth-grade content and that students were provided with review on the procedural aspects of fraction computation prior to the intervention.

The fourth study in this group (Brenner et al., 1997), focused on student understanding of key pre-algebra ideas such as the functional relationship between two variables, and contextual translation and application. A unit emphasizing meaningful thematic contexts and other features (thereby possibly confounding the “real-world” emphasis with other characteristics) was developed and used in three pre-algebra classes, and the control condition was three pre-algebra classes using a traditional algebra textbook. The effect size for the anchored instruction treatment in contrast with the traditional condition on solution to the function word-problem test was appreciable, though not significant (ES = 0.393). The effects on the word problem and equation solving measures were notably nonsignificant and slightly favored the control group. So, in this case, the influence of the treatment on the mathematical content that was especially aligned with the treatment was the strongest.

The final study in this group is of a different type. A group of researchers in the Netherlands (van Dijk, van Oers, Terwel, & van den Eeden, 2003) undertook a study with fifth-grade students to compare two different approaches to “real-world” problem-based instruction, in the spirit of the Dutch Realistic Mathematics Education (RME) movement of teaching through problems. Two different approaches to mathematization, or modeling (a type of “real-world” problem instruction) were used. The experimental treatment was called “guided co-construction,” where during instruction on the topic of percentages and graphs, students were guided by teachers to create their own models of the problems that were serving as the foundation of instruction. This was compared to a more traditional (within RME) expository approach, in which teachers provide students with models for the instructional problems. The posttest "measured the pupils' achievement regarding percentages and graphs in a quantitative way" (p. 177) which the Task Group takes to be a measure of what are typical school mathematics outcomes for the Dutch context, rather than a transfer task. The effect size favoring the guided co-construction group was not significant but encouraging (ES = 0.402).

In some of these studies the medium for introducing contextualized problems is video-based material, and the outcome measure is also video-based, causing over-alignment of the treatment with the outcome measure. Although the novelty of using video is not mentioned as a possible confound for studies of this type, it is worth considering how this particular instructional approach may affect students' interest and engagement.

The Task Group also calculated a pooled effect size across the four studies that are most similar (Anand & Ross, 1987; Bottge, 1999; Bottge, & Hasselbring, 1993; and Brenner et al., 1997) on the contextualized mathematics outcomes; see Table 15. Using the pooled measures in each of these studies, the pooled effect size was 0.616 and statistically significant. Thus the meta-analysis suggests that the impact of using "real-world" contexts in mathematics instruction on mathematics performance on similar "real-world" problems is significant. And, the impact on performance on other areas of mathematics, including computation, simple word problems, and equation solving, is not, at least when using this small set of studies as evidence. There are a number of caveats to be considered here; only four studies, all of them somewhat different, were included. And, the outcome measures are a mix of what might be thought of as "typical" mathematics measures, as well as more specialized transfer measures of contextualized or "real-world" problem solving.

In summary, the findings from these five studies, taken together, indicate that under certain conditions, the effect of treatments that employ contextualized problems in instruction on performance on contextual problems involving particular areas of mathematics can be significant. The results of these studies cannot be considered conclusive in providing direction on the general question of the use of "real-world" problem solving as a strategy for improving mathematics learning. However, they do suggest that certain well-defined "real-world" problem solving approaches can lead to improved performance on specific outcome measures, both for typical school mathematics performance, and more strongly, for transfer to "real-world" problem solving.

There were eight additional studies[9] that were classified as Category 2 but which will be discussed here because they provide additional insight into what has been learned from the Category 1 studies, or because they raise other interesting research issues. There were various, distinct flaws in these studies. For instance, in some, the use of "real-world" problems in instruction is confounded by concomitant instructional interventions, such as use of small groups, or emphasis on exploration in the curriculum, or inclusion of student writing as an instructional strategy. Other flaws also occurred, including the use of volunteer teachers in the treatment conditions, lack of matched control groups, lack of evidence of testing of

[9] Ben-Chaim, Fey, Fitzgerald, Benedetto, & Miller, 1998; Bottge Rueda, Serlin, Hung, & Kwon, 2007; CTGV, 1992; DeBock, Verschaffel, Janssens, Van Dooren, & Claes, 2003; Hickey et al., 2001; Henderson & Landesman, 1995; Irwin, 2001; Klein, Beishuizen, & Treffers, 1998.

group equivalence, outcome measures overly aligned with the treatment or control conditions, or only one unit assigned to each experimental condition. Thus no conclusions about impact of "real-world" problem solving as an instructional approach can be drawn from this set of studies.

Seven of the studies (all but Bottge et al., 2007) are designed to compare some type of "real-world" problem solving treatment to a control condition of typical mathematics instruction. Several use specific curricular interventions as the "real-world" problem intervention. In the case of Ben-Chaim et al. (1998) the treatment involved a full curriculum at the middle grades, the Connected Mathematics Program. A problem in using such a study to examine the specific impact of "real-world" problems is that full curricula such as this employ a range of principles and instructional approaches, and so findings cannot be clearly attributed to any particular component of the intervention. Henderson and Landesman (1995) used a similarly broad intervention (thematically integrated instruction) in their study, thereby making it difficult to interpret the impact of "real-world" problems. Given this, finding appropriate designs and measures that would allow a more focused look at the place of "real-world" problems in curricula that also include other interventions seems a worthwhile direction to pursue.

Klein et al. (1998) is another study in the Dutch context, comparing effects of two different approaches to teaching with realistic problems, where the introduction of flexible solution is handled differently in the treatment and control. This study is noteworthy because of the fine-grained detail in explaining the difference in these two approaches to working with "real-world" problems. The Task Group mentions the CTGV 1992 study because there is a creative outcome measure that has to do with problem solving planning. In the DeBock et al. (2003) research study, there is a very strong initial focus on the mathematical topic (applying linear models), raising the possibility that "real-world" instructional approaches may be better used for the teaching of some specific mathematical ideas rather than others. This study is also interesting because the assessments are varied according to the treatments, in an attempt to compare the impact of different treatments on assessments.

The Task Group also notes a final study that does raise some ideas that are worthy of consideration. In Bottge et al. (2007), the performance of different groups of students who were instructed using Enhanced Anchored Instruction (EAI) is compared. EAI is an instructional approach based on the concept of anchored instruction as advanced by the CTGV, which involves having students solve a problem in a multimedia format and then apply what they have learned in hands-on problem settings, such as building skateboard ramps (p. 32). The mathematical topics in this case involved rates, construction of graphs, lines of best fit, and fraction calculation. In this study, the emphasis is on the possibly differential impact of the "real-world" oriented curriculum on different groups of students (students with learning disabilities, and high- and average-achieving students). This is classified as a Category 2 study because of design issues, but it is an interesting example for consideration. The authors report that, following treatment, students in the inclusive classes (which include learning disabled students) outscore the students in the typical classes. This, together with other studies by Bottge, as well as the study by Henderson and Landesman (1995) that is concerned with bilingual instruction as well as thematic integration, suggests that more systematic research on the impact of "real-world" problem based instruction on particular subgroups of students who have been traditionally underserved in mathematics, may be worthwhile.

Three of the studies examine the impact of video-based instruction that involves the presentation of mathematical problems through "real-world," contextual settings (Cognition and Technology Group at Vanderbilt [CTGV], 1992; DeBock et al., 2003; and Hickey, Moore, & Pellegrino, 2001). Two (CTGV, 1992; Hickey, Moore, & Pellegrino, 2001) report on the impact of the implementation of the *Jasper Woodbury* series. DeBock et al. (2003) use video material based on Gulliver's Travels. There are other innovations in the use of video-based instruction, including involvement of students in cooperative groups, for instance, which can cause confounding; in addition, the types of outcome measures used in these studies vary in terms of their closeness to the focus of the intervention.

Remaining mindful that all of these studies have flaws that prevent their inclusion in Category 1, five of them report significantly better performance of treatment groups (some kind of "real-world" instruction) than of control groups, on particular measures that tend to emphasize "real-world" problems in one way or another (Ben–Chaim et al. 1998; CTGV, 1992; Hickey et al., 2001; Henderson & Landesman, 1995; and Irwin, 2001). In contrast, Klein et al. (1998) report no difference on procedural competence between the control and treatment groups, and DeBock et al. (2003) report a negative result, where students using the video instructional treatment performed worse than those who solved non-embedded problems on the outcome measure.

No conclusions about impact can be drawn from these studies. Instead, they highlight the complexities of these research issues, and point toward interesting questions, designs, and measures that could help form a foundation for subsequent research.

E. Using Specific Strategies to Improve 'Real-World' Problem Solving

The search for studies that examined the use of "real-world" problems as an instructional strategy led to a small number of Category 1 studies that concerned the impact of different instructional strategies for teaching students to solve "real-world" problems. Note that this is not necessarily all of the studies that have examined strategies for improving "real-world" problem solving. This work is distinguished from what is included in the prior section in part by a particularly strong focus on the primary goals of both near and far transfer outcomes. Lynn Fuchs and her research group have undertaken a series of studies in this vein, several of which met our criteria. A study by Fuchs, Fuchs, Hamlett, and Appleton (2002) was aimed at enhancing mathematical problem-solving performance of fourth-graders with mathematical disabilities. All students participated in their regular classroom mathematics instruction using a basal text, and a six-lesson base treatment on approaching mathematical problem solving. One experimental group received 24 sessions of problem-solving tutoring; one received 24 sessions of computer-assisted practice; a third received both the tutoring and computer-assisted practice sessions.[10] Small group tutoring was provided on problem-solving rules and on transfer. The computer-assisted practice emphasized tasks intended to lead to far-transfer. There were three types of outcome measures: ability to solve story problems, transfer story problems, and "real-world" problems. Significant effects favoring the problem-solving tutorial group were found on the story problem and transfer story problem measure (effect sizes of 1.340 and 0.982, respectively). The effect size on the "real-world" problem solving measure was not significant (-0.041) and slightly favored the computer-assisted practice group, indicating no significant differences between treatments on the primary outcome measure of far transfer.

Table 16: Studies That Examine Strategies to Improve "Real-World" Problem Solving

Study	Design	Sample	Duration/ Content	Contrast	Measure	Hedge's g	Standard Error	
Barron, 2000	RCT	96 sixth-grade students in a public magnet school for academically talented children	Four 1-hour sessions/ Contextual problem solving	Problem solving collaboratively in triads vs. Problem solving individually	Pooled transfer measures	0.287	(ns)	0.204
Fuchs et al., 2002	RCT	40 fourth-grade students with mathematical disabilities in six classrooms in three schools	24 sessions/ Mathematical problem solving	Problem-solving tutoring vs. Computer-assisted practice	Pooled story problems (ES = 1.340**), transfer story problems (ES = 0.982**) and "real-world" problem-solving measures (ES = -0.041)	0.760	~	0.454
Fuchs et al., 2004[a]	RCT	351 third-grade students in 24 classrooms in seven schools in an urban district	34 lessons over 16 weeks/ Mathematical problem solving	SBTI vs. Control	Transfer-4 measure (a measure of far transfer that approximated real life problem solving)	1.123	*	0.513
				SBTI expanded vs. Control		2.087	***	0.600
Fuchs et al., 2006[a]	RCT	445 third-grade students in 30 classrooms in seven schools in an urban district	16 weeks/ Mathematical problem solving strategies	SBI vs. Control	Pooled transfer measures	0.545	(ns)	0.439
				SBI-RL vs. Control		1.077	*	0.464
Rudnitsky et al., 1995	RCT	401 third- and fourth-grade students in 21 classrooms in six schools	18 days/ Addition and subtraction word problems	Writing and discussion vs. Practice and explicit heuristics	Near transfer posttest	0.190	~	0.115

~ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$
[a] These studies use classroom-level analyses.

[10] Contrasts with this condition not included in table.

A second study (Fuchs, Fuchs, Finelli, Courey, & Hamlett, 2004) built on these findings and implemented schema-based transfer instruction (SBTI), which explicitly teaches students about transfer features of problems in an effort to improve their near- and far-transfer performance. Twenty-four third-grade teacher volunteers in seven urban schools were randomly assigned to one of three conditions: control, SBTI, and SBTI expanded (this included focus on additional and challenging superficial problem features such as irrelevant information, and the concept of “real-life” situations that introduce more information than problems typically used in school). The 16-week treatments were compared using four outcome measures: Transfer-1 (novel problems structured in the same was as those in the instruction); Transfer-2 (novel problems that varied in the three transfer features taught in SBTI); Transfer-3 (novel problems that varied in transfer features taught in both SBTI and expanded SBTI); and Transfer-4 (measure of far transfer that varied from the problems used in instruction in six major ways). Calculation of effect sizes for Transfer-4 (measure of far transfer that approximated “real-life” problem solving) yielded significant effects for the SBTI expanded vs. the control condition (ES = 2.087), and for the SBTI vs. the control group (ES = 1.123). The Task Group can conclude that this particular, highly specific instructional approach can result in stronger performance on a “real-world” problem outcome.

This group of studies led to a randomized controlled study published in 2006 (Fuchs, Fuchs, Finelli, Courey, Hamlett, Sones et al., 2006). Three treatment conditions were implemented: the “teacher-designed” condition, which was the control, with teachers using the district curriculum; and two schema-broadening instruction (SBI) conditions. One SBI condition was the problem-solving instruction used in earlier studies, emphasizing superficial problem features, problem structures, and problem types. The second SBI condition is expanded to include “explicit instruction in strategies for tackling the complexities involved in real-life problems” (p. 296). The Task Group found a significant effect size (ES = 1.077) for the enhanced schema-broadening instruction aimed at preparation for solving real-life problems in comparison to the control group members on pooled far transfer measures. In addition, the effect of the SBI treatment in comparison to the control on the far transfer outcome measures was encouraging, though not significant (ES = 0.545). The results of these two studies (Fuchs et al. 2004, Fuchs et al. 2006), in contrast to the Fuchs, Hamlett, & Appleton et al. (2002) suggest that the enhanced schema-broadening instruction, which explicitly helps students to recognize and attend to irrelevant and extraneous features in real-life problems, is effective in enabling students to successfully solve real-life transfer problems.

Two additional studies met the Category 1 criteria and have been classified as being about promoting student performance on “real-world” problems through a particular instructional strategy. Barron (2000) used *Jasper Woodbury*-style video-based microworlds as the instructional treatment being tested, in comparison to control conditions, for its effect on a student problem solving performance measure. Two different grouping strategies for students using the *Jasper Woodbury* materials were compared. In one condition, the sixth-grade students worked in triads. In the other, they worked individually. The effect size calculation yielded an encouraging though not significant effect of the triad arrangement (ES = 0.287). Rudnitsky, Etheridge, Freeman, and Gilbert (1995) focused on helping third- and fourth-grade students solve arithmetic word problems through two different treatments: a "writing-to-learn" approach, in which students created their own mathematical stories and problems, and a control condition involving practice and explicit heuristics. On a near-transfer problem-solving posttest, there was no significant effect size (ES = 0.190).

In summary, these five studies examine several very different types of instruction intended to improve near or far transfer performance on "real-world" problems. The strategies were: student grouping, computer-assisted instruction, problem-solving tutoring, schema-based instruction, enhanced schema-based instruction, and problem writing. It is not reasonable to calculate pooled effect size for these five studies, given the differences in the instructional interventions. It is important to note that, of all of these strategies, the only one that shows promise on an empirical basis is the enhanced schema-based instruction in both Fuchs et al. (2004) and Fuchs et al. (2006). Note too that the mathematical domain is narrow (whole number arithmetic) and this was undertaken only at the third grade. At the same time, the heart of the intervention—a focus on extraneous and irrelevant information—is a feature that some would surely say is a defining feature of "real-world" problems; these are messy problems. Fuchs and her colleagues seem to have demonstrated that, under very specific conditions, in a very narrow area of mathematics, it is possible to teach students how to address these issues and be effective problem solvers.

There were three studies classified as Category 2 that also examine instructional strategies for improving performance on "real-world" problems. All of them have design flaws which exclude them as studies from which the Task Group can draw conclusions about impact. However, these studies are instructive because they provide ideas about various kinds of instructional interventions that have been attempted, and about interesting outcome measures. Serafino and Cicchelli (2003) contrast two instructional approaches within the anchored instruction model on which the Jasper materials are based. The Structured Problem Solving model includes a more teacher-dominated, structured approach, with more focused teacher questions and summaries. The Guided Generation model casts the teacher in more of a facilitator role. Shyu (1999) investigated the effects of a video-based anchored instruction program based on the Jasper Series. There were three instructional treatments—the video-based instruction, the printed, story-book version, and regular instruction. Both of these studies designed alternative instructional approaches for use with problem-based curricula.

Verschaffel and DeCorte (1997) conducted an experiment with 10–12 year olds in which the treatment involved a sequenced introduction of "real-world" problems and discussion of the information available and the approach to the problem. What is of particular interest in this study is the outcome measure, "disposition toward realistic modeling," which is intended to assess students' tendency to use "real-world" knowledge and realistic considerations in their problem solving. The authors report a significant difference favoring the treatment group, but these results are not robust given that the treatment condition had only one unit. Nonetheless, it is worth noting that the instructional interventions in this study seem to share the principles in the Fuchs et al. approaches, in which there is explicit focus on features characteristic of "real-world" problems. Further development of such approaches for use in wider contexts might be fruitful.

Based on an examination of studies primarily concerned with testing different approaches to teaching that enable students to solve “real-world” problems, the Task Group concludes that some instructional approaches will lead to better student performance on “real-world” problems than others. However, these instructional practices are highly specified, and the studies only demonstrate their effectiveness for relatively narrow classes of problems. For those who view performance on “real-world” problems as an important outcome of K–12 mathematics education, there are still far more open questions about what will lead to far transfer and which instructional methods are best than there are conclusions.

F. Conclusion

It is difficult to draw conclusions from the set of studies that examine the impact of the use of “real-world” problems and related instructional strategies in instruction on student mathematics performance, including performance on “real-world” problems. The body of studies is small; the outcome measures are often designed by the researchers and information is not available on psychometric characteristics of these measures; and, confounding variables that are difficult to measure reliably, such as fidelity of implementation and other contextual features, are not always included in the study reports.

The set of studies also has a certain homogeneity. Of the 21 studies discussed here, 10 of them are focused on instructional materials that introduce “real-world” problems through the Jasper Woodbury series, or similar video materials. Researchers have not undertaken the necessary rigorous examination of print instructional materials that have as their primary goal the introduction of mathematical ideas through “real-world” problems. Nor has there been adequate attention to the possibility that different mathematical ideas, topics, and procedures might best be learned through particular instructional approaches; perhaps using “real-world” problems is good for some mathematical topics and not for others. The Task Group found very few studies that started from any clear hypotheses about why a particular intervention would be likely to help with a particular area of mathematics.

Debates about the place of “real-world” problems in the mathematics classroom are complicated by a number of issues; the operationalization of the term “real-world” problems varies by mathematician, researcher, developer, and teacher; fidelity of the teachers’ implementation of the instructional materials or instructional strategy is difficult to assess; contextual features, such as SES, or the school’s orientation toward reform matter; and most likely, although not addressed in the studies the Task Group examined, teachers’ knowledge and capacity to use such problems effectively varies greatly.

A particularly relevant issue to focus on in this domain may be the degree to which students’ ability to apply mathematical knowledge to “real-world” or “authentic” problem situations is a valued and agreed-upon outcome of school mathematics. If “real-world” problem solving is not seen as an essential outcome of K–12 mathematics education, then the modest accumulation of research available (meeting our screening criteria) on the topic would suggest that there is no great value in using “real-world” problems as a main element of mathematics instruction nor is there great value investing significant time to design effective instructional strategies that rely on “real world” contexts. However, if ensuring that students know the needed mathematics, and can apply that mathematics to the more complex and open kinds of problems that can be encountered in the "real world" is important, then the studies reviewed here offer some promise.

The Task Group concludes that, under certain conditions, for specific domains of mathematics, instruction that features the use of "real-world" contexts shows potential promise for having a positive impact on student achievement. However, these results are not yet sufficient as a basis for widespread policy recommendations.

If the goal of application of the mathematical knowledge in contexts is considered important, then these studies would suggest that continued investment in research and development that is coordinated with state standards may be worthwhile, with several caveats. More studies should use standardized outcome measures in place of the researcher or developer-designed instruments, so that the results can accumulate in a more useful way. If such measures are not used, then the design of outcome goals and measures needs more integrated involvement of psychometricians and mathematicians, who can watch for the difficulties of overly confounding the outcome assessment with the intervention, or of assessing mathematics too narrowly. Studies that look beyond special populations of students (e.g., remedial students, special education students) are needed. Randomized control experiments are necessary for generalization and clarity about the scale-up potential and outcomes of specific interventions. And, more attention is needed to the specific kinds of mathematical outcomes that are obtained by specific types of "real-world" problem interventions. For instance, "real-world" approaches may be especially useful for introducing particular mathematical concepts and processes, and less useful or inefficient for the introduction of other topics. Thus far the research has made little systematic progress on this matter.

BIBLIOGRAPHY

Anand, P.G., & Ross, S.M. (1987). Using computer-assisted instruction to personalize arithmetic materials for elementary school children. *Journal of Educational Psychology, 79(1),* 72–78.

Anderson, J., Reder, L.M., & Simon, H.A. (1996). Situated learning and education. *Educational Researcher, 25,* 5–11.

Barron, B. (2000). Problem solving in video-based microworlds: Collaborative and individual outcomes of high-achieving sixth-grade students. *Journal of Educational Psychology, 92(2),* 391–398.

Ben-Chaim, D., Fey, J.T., Fitzgerald, W.M., Benedetto, C., & Miller, J. (1998). Proportional reasoning among 7th grade students with different curricular experiences. *Educational Studies in Mathematics, 36(3),* 247–273.

Bottge, B.A (1999). Effects of contextualized math instruction on problem solving of average and below average achieving students. *The Journal of Special Education, 33*(2), 81–92.

Bottge, B. & Hasselbring, T. (1993). A comparison of two approaches for teaching complex, authentic mathematics problems to adolescents in remedial math classes. *Exceptional Children, 59(6),* 556–566.

Bottge, B.A., Rueda, E., Serlin, R.C., Hung, Y., & Kwon, J.M. (2007). Shrinking achievement differences with anchored math problems: Challenges and possibilities. *Journal of Special Education, 41(1),* 31–49.

Bransford, J.D., Sherwood, R., Hasselbring, T., Kinzer, C., & Williams, S. (1990). Anchored instruction: Why we need it and how technology can help. In D. Nix & R. Spiro (Eds.), *Cognition, education, and multimedia: Exploring ideas in high technology* (pp. 115–141). Hillsdale, NJ: Lawrence Erlbaum Associates.

Bransford, J.D., Vye, N., Kinzer, C., & Risko, V. (1990). Teaching thinking and content knowledge: Toward an integrated approach. In B. Jones, & L. Idol (Eds.), *Dimensions of Thinking and Cognitive Instruction* (pp. 381–413). Hillsdale, NJ: Lawrence Erlbaum Associates.

Brenner, M., Mayer, R., Moseley, B., Brar, T., Duran, R., Reed, B., & Webb, D. (1997). Learning by understanding: The role of multiple representations in learning algebra. *American Educational Research Journal, 34(4),* 663–689.

Cobb, P., & Bower, J. (1999). Cognitive and situated learning perspectives in theory and practice. *Educational Researcher, 28(2),* 4–15.

Cognition and Technology Group at Vanderbilt. (1991). Technology and the design of generative learning environments. *Educational Technology, 31(5),* 34–40.

Cognition and Technology Group at Vanderbilt. (1992). The Jasper Series as an example of anchored instruction: Theory, program description, and assessment data. *Educational Psychologist, 27(3),* 291–315.

De Bock, D., Verschaffel, L., Janssens, D., Van Dooren, W., & Claes, K. (2003). Do realistic contexts and graphical representations always have a beneficial impact on students' performance? Negative evidence from a study on modeling non-linear geometry problems. *Learning and Instruction, 13(4),* 441–463.

Freudenthal, H. (1973). *Mathematics as an educational task.* Dordrecht, The Netherlands: Reidel.

Freudenthal, H. (1991). *Revisiting mathematics education*. Dordrecht, The Netherlands: Reidel.

Fuchs, L. S., Fuchs, D., Hamlett, C.L., & Appleton, A.C. (2002). Explicitly teaching for transfer: Effects on the mathematical problem-solving performance of students with mathematics disabilities. *Learning Disabilities Research & Practice, 17*(2), 90–106.

Fuchs, L.S., Fuchs, D., Finelli, R., Courey, S.J., & Hamlett, C.L. (2004). Expanding schema-based transfer instruction to help third graders solve real-life mathematical problems. *American Educational Research Journal, 41*(2), 419–445.

Fuchs, L.S., Fuchs, D., Finelli, R., Courey, S.J., Hamlett, C., Sones, E.M., et al. (2006). Teaching third graders about real-life mathematical problem solving: A randomized controlled study. *The Elementary School Journal, 106*(4), 293–311.

Geary, D.C. (2005). Role of cognitive theory in study of learning disability in mathematics. *Journal of Learning Disabilities, 38,* 305–307.

Greeno, J.G. (1997). On claims that answer the wrong questions. *Educational Researcher, 26(1),* 5–17.

Henderson, R. & Landesman, E. (1995). Effects of thematically integrated mathematics instruction on students of Mexican descent. *Journal of Educational Research, 88(5),* 290–300.

Hickey, D., Moore, A., & Pellegrino, J. (2001). The motivational and academic consequences of elementary mathematics environments: Do constructivist innovations and reforms make a difference? *American Educational Research Journal, 38(3),* 611–652.

Hiebert, J., Carpenter, T., Fennema, E., Fuson, K., Human, P., Murray, H., et al. (1996). Problem solving as a basis for reform in curriculum and instruction: The case of mathematics. *Educational Researcher, 25(4),* 12–21.

Irwin, K. (2001). Using everyday knowledge of decimals to enhance understanding. *Journal for Research in Mathematics Education, 32(4),* 399–420.

Jonassen, D.H. (2003). Designing instruction for story problems. *Educational Psychology Review,* 15*(3),* 267–296.

Klein, A.S., Beishuizen, M., & Treffers, A. (1998). The empty number line in Dutch second grades: Realistic versus graduate program design. *Journal for Research in Mathematics Education, 29(4),* 443–464.

Kline, M. (1973). *Why Johnny can't add: The failure of the new math*. New York, NY: Random House, Inc.

Lemke, M., Sen, A., Pahlke, E., Partelow, L., Miller, D., Williams, T., et al. (2004). *International outcomes of learning in mathematics literacy and problem solving: PISA 2003 results from the U.S. perspective* (No. NCES 2005-003). Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Michalowicz, K.D., & Howard, A.C. (2003). Pedagogy in text: An analysis of mathematics texts from the nineteenth century. In G. Stanic and J. Kilpatrick (Ed.), *A history of school mathematics* (pp. 77–109). Reston, VA: National Council of Teachers of Mathematics.

National Council of Teachers of Mathematics (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.

Pehkonen, E. (1997). Introduction to the concept of "open-ended problem." In E. Pehkonen (Ed.), *Use of open-ended problems in the mathematics classroom*. Research Report 176. Helsinki: ERIC. ED 419714.

Presmeg, N. (2007). The role of culture in teaching and learning mathematics. In F. Lester (Ed.), *Second handbook of research on mathematics teaching and learning*. Charlotte, NC: Information Age Publishing.

Programme for International Student Assessment. (2003). *PISA 2003 assessment framework–Mathematics, reading, science and problem-solving knowledge and skills*. Paris: Organisation for Economic Co-operation and Development.

Resnick, L.B. (1987a). *Education and learning to think.* Washington DC: National Academy Press.

Resnick, L.B. (1987b). Learning in school and out. *Educational Researcher, 16*, 13–20.

Rudnitsky, A., Etheredge, S., Freeman, S., & Gilbert, T. (1995). Learning to solve addition and subtraction word problems through a structure-plus-writing approach. *Journal for Research in Mathematics Education, 26(5),* 467–486.

Serafino, K., & Ciccelli, T. (2003). Cognitive theories, prior knowledge, and anchored instruction on mathematical problem solving and transfer. *Education and Urban Society, 36(1),* 79–93.

Shyu, H. (1999). Effects of media attributes in anchored instruction. *Journal for Research in Mathematics Education,* 32(4), 399–420.

The Adventures of Jasper Woodbury. Retrieved November 1, 2007, from http://peabody.vanderbilt.edu/projects/funded/jasper/Jasperhome.html.

van Dijk, I.M.A.W., van Oers, B., Terwel, J., & van den Eeden, P. (2003). Strategic learning in primary mathematics education: Effects of an experimental program in modeling. *Educational Research and Evaluation, 9(2),* 161–187.

Verschaffel, L., & De Corte, E. (1997). Teaching realistic mathematical modeling in the elementary school: A teaching experiment with fifth graders. *Journal for Research in Mathematics Education, 28(5),* 577–601.

Wu, H.S. (1994). The role of open-ended problems in mathematics education. *Journal of Mathematical Behavior, 13(1),* 115–128.

V. The Role of Technology in Mathematics Education

Although young in historic terms, computer technology has a strong presence in our lives and in the research literature. This report synthesizes what is known from high-quality research about the effectiveness of a variety of approaches to applying computer technology to the solution of educational problems in mathematics instruction. The report begins with a brief overview of the categories of computer applications that mathematics educators have used. Next, using the prior reviews, syntheses, and meta-analyses as context and background, the Instructional Practices Task Group's own original meta-analyses of rigorous studies for those categories that included an adequate body of studies that fit the Task Group's criteria are presented. These included drill and practice, tutorials, calculators, and computer programming. The Task Group's basic question is: What is the role of technology including computer software, calculators and graphing calculators in mathematics instruction and learning? The last section summarizes answers to this question on the basis of the Task Group's review of high-quality research.

A. Categories of Instructional Software

As an all-purpose device, a computer can take a variety of forms and play a variety of instructional roles. The term *computer-based instruction* (CBI) will refer to all these applications of computer technology to education. As an interactive device, the computer can be programmed to provide opportunities for active learning and reflective thinking on the one hand, or to provide drills on the other. It might manage and individualize instruction. It can perform tedious calculations, potentially having positive effects (if it thus allowed engagement in topics otherwise impossible or difficult for students to approach) or unintended consequences (students become overly dependent on calculators). This paper uses the following categories to classify instructional software (with the caveat that software programs can combine pedagogical categories):

- Drill and practice;
- Tutorials;
- Tools (including calculators) and problem solving;
- Computer programming;
- Simulations;
- Games;
- Internet;
- Tools for teachers.

A brief description of each of these categories is provided below, and Table 17 provides complementary descriptions. Table 17 lists several features that may distinguish more effective from less effective computer-based practice (including unique features—those that can not easily be duplicated in noncomputer environments).

Drill and practice software provides practice on skills and knowledge to help students remember and use that which they have been taught. A main goal is to achieve automaticity, or fast, accurate, and effortless performance, freeing working memory so that attention can be directed to the more complicated aspects of complex tasks. As with most drill in any medium, drill and practice computer programs present tasks or exercises and give feedback to students.

Tutorials attempt to introduce and teach new subject-matter content, by presenting information and often by attempting to engage students in one-to-one Socratic dialog (e.g., tutorials using artificial intelligence to engage in dialogues). These are usually developed in situations in which a well-defined set of information must be acquired.

The term *computer-assisted instruction* (CAI) is commonly used to refer to drill and practice programs, tutorials, or their combination. A specific type of CAI is the integrated learning system (ILS), a large suite of programs, mainly tutorial, but with drill and practice included, that provides sequenced instruction across several grade levels, tracking students' progress and branching as necessary, and maintaining extensive records of student progress (using *computer-managed instruction*, or CMI, which is discussed in a following section).

Simulations are models of some part of the world (such as the noncomputer board games "Life" or "Monopoly"), and computer simulations are often more complex mathematical models that respond in relatively realistic ways to input based on "real-world" data. Most simulations present situations with components and interactions among those components and generate data about them in response to student input that mirrors relationships in those physical-world or mathematical situations. Thus, students play a role of an active member of a system, making decisions and analyzing the results of those decisions. Goals often are to motivate engagement, develop intuition about a problematic situation, facilitate acquisition of skills and knowledge, and enhance transfer of mathematical skills as students perform activities reflecting those in the "real world."

Games may share characteristics with simulations, as the term "simulation games" suggests. This category is broader, encompassing games that are no more than drill and practice with game-like elements used as rewards, to those in which mathematics is intrinsic to the goals, rules, and tasks of the game. Games of the former type, in which mathematics is extrinsic, often have goals similar to those of drill and practice software. The latter are often designed to promote acquisition of mathematical knowledge and skill, as well as problem solving.

Tools include a wide variety of software programs that perform specific sets of functions, such as calculation, statistics and graphing, computer-based laboratories (CBL, e.g., sensors, including statistical analysis and display of the resulting data), or manipulation of mathematical expressions in symbolic form. Pedagogical goals may include allowing more complex problem solving by transferring routine aspects of tasks to the technological tool and encouraging students to solve problems in practical, applied settings. *Calculators*, including graphing calculators, are widely available tools that have generated a large amount of interest and research. They have been used for many purposes, from facilitating problem solving by allowing students to perform far more, and more complex, arithmetic operations than would have been possible without technology, to serving as simple fact checkers.

Problem solving applications may be one or more tools as above, but may also include the presentation of problems and feedback (similar to the feedback of CAI).

Computer programming involves the provision of computer languages or environments to facilitate students' creation of procedures that solve mathematical problems. Goals may include students' learning and reflecting on algorithms (arithmetic or algebraic), as expressed in the computer language, gaining specific knowledge and skills (e.g., in geometry), and learning certain problem-solving strategies, such as problem determination and explication, problem decomposition, and construction and evaluation of procedures. Some environments are tuned for special purposes, such as the development of mathematical models for simulations, or providing a scripting language within a geometric construction program.

The *Internet* provides general information searching and retrieval functions. Educational applications include specifically organized inquiries (e.g., "WebQuests"). The Internet also offers myriad applications and features ("blogs," groups, etc.) that may be harnessed for the purposes of mathematics education. The Internet can also be the delivery medium for any of the other categories of software; those are considered within their specific category.

Tools for teachers include a variety of software programs designed to aid pedagogical tasks. For example, electronic blackboards ("smart boards") facilitate the display of information or demonstration of any type of software, and with "clickers," can aggregate students' responses; management systems help store, organize and analyze information, such as achievement data, and may include item, test, or practice generators; and hand-held devices facilitate classroom interaction (e.g., each student has a device, and responses or data entry are easily and quickly inputted and evaluated or aggregated).

Computer management systems include *computer-managed instruction*, or CMI, in which the computer analyzes assessments of students, directs their course through a curriculum, and provides reports at individual and aggregate levels. These can be stand-alone systems or can form the foundation for other categories of CBI, such as CAI.

Table 17: Categories of Educational Software

Category	Typical Pedagogy	Possible Features
Drill and Practice	Linear Repetitious Presentation of task, student response, feedback	Sequence[a] Management[b] Feedback[c] Controlled introduction of items Distributed practice Reinforcement schedules
Tutorials	Linear progression with various amounts of branching Didactic presentation and, sometimes, Socratic dialog, presentation of information, questioning, and feedback depending on the response; branching to explanations or review	Sequence Management Feedback Instructional events[d]
Tools and Problem Solving	Specific functions (calculator, graphing, computer-based laboratories, geometric construction, CAS) Problem Solving may include presentation of problems and feedback	Integration/data communication across tools (or with other software categories) Specific feature sets
Computer Programming	Specific language Specific educational environment, specific tasks	Mathematics emphasis Integration into curriculum
Simulations	Nonlinear; exploratory/inquiry-oriented Provides a model of "real-world" or mathematical situation in which students act; then responds to students input following that model	Integrated with tutorials or teaching tools Appropriate simulation[e]
Games	Provides a set of tools and/or miniature "world" as setting for attempting to achieve a goal within a framework of rules Provides clear goals, a set of artificial rules, and elements of competition	Mathematics emphasis Intrinsic mathematics Manipulation of concepts Motivational elements
Internet	Type: General information search/retrieval, "WebQuest," other	
Tools for Teachers	Type: electronic blackboard, demonstration/display, management system (CMI; may include practice generator), item/test/practice generator, classroom interaction (each student has device)	Integration/data communication across tools (or with other software categories)

[a] Sequence: Consists of building a sequence of mathematical strategies/skills/concepts.
[b] Management: Computer management may consist of record keeping only (includes "picks up where left off"), more sophisticated formative assessment, or formative-assessment-with-branching [e.g., remediation].
[c] Feedback: Corrective feedback may be knowledge of correctness only, or also provide answer, or also provide remediation or explanation. May attend to speed of response.
[d] Features most of the events of instruction identified by cognitive psychology to correspond to learning processes (e.g., gaining attention, informing learner of objectives, stimulating recall of prior learning, presenting stimuli with distinctive features, guiding learning, eliciting performance, providing informative feedback, assign performance, enhancing retention and transfer).
[e] Appropriate abstraction or simplification of the problem situation vs. oversimplification or misrepresentation of the "real-world" situation or the mathematics.

B. Methods

1. Syntheses of Existing Reviews

Prior syntheses and meta-analyses related to the effects of different forms of instructional technology on student mathematics achievement were identified through keyword searches in PsycInfo and Web of Social Sciences Citation Index. Experts in the field of technology instruction and meta-analysis provided additional references. Finally, the reference lists from the identified syntheses and related original studies were reviewed to identify additional syntheses.

From among the group of reviews that were identified, 26 quantitative syntheses and meta-analyses were included in the Task Group's synthesis of existing reviews.[11] These were reviewed to ascertain the number of included studies that focused on the primary population of interest (elementary and junior high school students taking part in mathematics-related technology interventions), the nature of the technology, and the syntheses procedures. Results from these quantitative syntheses and meta-analyses that addressed the primary population of interest of the Panel and the technologies considered were then summarized. The pooled effect sizes from these meta-analyses are presented in Tables C-1 through C-7 of Appendix C.

2. The Task Group's Meta-Analyses

For the Task Group's original meta-analyses, studies were located using the Group's search procedures and the keywords listed in Appendix A. Original empirical studies on technology were categorized based on the category of software on which the intervention focused. Effect sizes were calculated for the Category 1 studies, and effect sizes were pooled when appropriate. All effect sizes have been adjusted for clustering, when appropriate. Study characteristics are provided for each of the Category 1 studies that were included in the meta-analyses.

A number of methodological decisions in preparing the data for analysis and in choosing which effect sizes to include in the pooled analyses were made. In particular, four key issues were confronted, as follows.

First, a number of studies evaluated the effects of more than one technology intervention and/or more than one comparison group. Specifically, three studies (Battista & Clements, 1986; Clements, 1986; and Emihovich & Miller, 1988) evaluated the effects of a programming

[11] The Task Group examined literature reviews, syntheses, and meta-analyses, and conducted syntheses reported here of prior *quantitative* syntheses and meta-analyses, as follows: Becker (1992); Burns & Bozeman (1981); Chambers (2002); Christmann, Badgett, & Lucking (1997); Ellington (2003); Ellington (2006); Fletcher-Flinn & Gravatt (1995); Gordon (1992); Hamilton (1995); Hartley (1978); Hembree (1984); Hembree & Dessart (1986); Hembree (1992); Khalili & Shashaani (1994); Khoju, Jaciw, & Miller (2005); Kuchler (1999); Kulik & Kulik (1991); Kulik (1994); Kulik (2003); Lee (1990); Lou, Abrimi, & d'Apollonia (2001); Niemiec & Walberg (1984); Ryan (1991); Slavin, Lake, & Groff (2007); Slavin & Lake (2007); Smith (1997).

treatment (using Logo) and a tutorial or drill and practice (i.e., CAI) treatment, compared to a no-treatment control. In two of these cases (Clements, 1986; and Emihovich & Miller, 1988), the CAI group was focused, at least in part, on mathematical content. In these cases, the programming versus control group contrast was included in the meta-analysis exploring the effects of programming interventions, and the CAI versus control group contrasts were included in the meta-analyses exploring the effects of drill and practice programs (Emihovich & Miller, 1988) or tutorial programs (Clements, 1986). The programming vs. CAI comparisons are noted in the programming section and presented in Appendix C. In other studies, two similar treatments were compared with a no-treatment comparison group. In these cases (for example, in Oprea, 1988), the treatment that was more similar to a typical intervention that schools would be likely to implement was included. Still, in other studies, a specific intervention was compared to multiple comparison groups. In this situation, the most relevant intervention versus control contrasts was chosen on a case-by-case basis.[12]

Second, studies often explored the effects of interventions on a range of outcomes. For the purposes of this meta-analysis, the focus was only on mathematics-related or problem-solving outcomes. In cases in which multiple outcomes within these domains were available, an average effect size across the multiple outcomes was calculated.

Third, studies often reported effects on a variety of independent samples of students. For example, studies sometimes reported results by race, gender, grade level, or disability status. In cases in which it is likely that the intervention experience was different for these subgroups multiple effect sizes for a study are presented; for example, separate effects by grade level and disability status. In addition, multiple effect sizes are reported for studies that present results from multiple trials exploring the same intervention and outcome (for example, across sites or across samples or cohorts).

Fourth, a number of studies met the criteria for being Category 1 studies but did not compare a technology intervention to a no-treatment control group. Instead, these studies compared two different versions of technology interventions. Although these studies are not appropriate to pool in a meta-analysis, some suggested findings from these comparison studies are presented.

Finally, for studies about calculators only, there were three additional methodological decisions that were made in preparing the data for analysis and in choosing which effect sizes to include in the pooled analyses. First, three studies (Szetela, 1980; Szetela, 1982; and Wheatley, 1980) presented outcomes based on assessments where the calculator treatment group was allowed to use calculators, while the no-calculator comparison group was not. In two of these studies, (Szetela, 1980; Wheatley, 1980) this was the only information available. In the one case in which both an assessment allowing calculator use and a standard paper and

[12] For example, in Johnson-Gentile et al., 1994, one comparison group received an intervention that used manipulatives that was almost identical to the curriculum of the programming intervention. In this case, the programming versus the no-treatment comparison group contrast was included. However, in Ortiz and MacGregor (1991), we chose to include the programming intervention versus a textbook-based intervention contrast, because the no-treatment comparison group did not receive any instruction on the outcome that was being evaluated ("the concept of variable").

pencil assessment that did not allow calculator use were available (Szetela, 1982), the effect size for the latter was included Table C-9 in Appendix C summarizes the effects for any additional sets of effect sizes in which students were allowed to use calculators.

Second, three studies (Duffy & Thompson, 1980; Standifer & Maples, 1981; and Standifer & Maples, 1982) evaluated the effects of two calculator interventions and a no-calculator control group. For the purposes of the meta-analysis, contrasts that are most similar to contrasts in other studies are included, thus attempting to compare the basic treatment of using a calculator during instruction versus not using a calculator.[13] The two Standifer and Maples studies compared the effects of using a standard hand-held calculator and a "programmed feedback" calculator. Focus was on the hand-held versus control group contrast, with the additional effect sizes presented in Table C-9 in Appendix C. The Duffy and Thompson study includes one condition that simply provides students with calculators in the classroom and does not provide guidance to teachers, a second condition that provides calculators plus instructional packages for teachers, and a no-calculator control condition. Again, focus was on the basic calculator versus no calculator contrast, with the effect sizes for the more enhanced treatment documented in Table C-9.

Third, there were three studies (the same three as noted in the previous paragraph) that also provided effect size information for Total Achievement scores. This information is also presented in Table C-9.

C. Categories of Instructional Software: Findings

This section summarizes findings from studies that examined specific categories of instructional software. For each category, results from prior syntheses and meta-analyses provide background information and, for each category for which the Task Group conducted its own meta-analyses, those results are presented.

1. Drill and Practice

a. Prior Syntheses and Meta-Analyses

Many of the studies in the prior CBI reviews probably included drill and practice software, so there is reason to expect similar results for reviews that delineated this category. The detailed effect size information from prior quantitative syntheses and meta-analyses are presented in Table C-1 in Appendix C. Prior syntheses and meta-analyses (see Table 18) suggest that CAI drill and practice generally improves students' performance compared to conventional instruction, with the greatest effects on computation, and less effect on concepts and applications (recall the caveats expressed previously, and note the discussion in the following section on tutorials). Prior reviews have found that drill and practice positively affects attitudes toward mathematics and instruction in mathematics. They suggest that drill and practice is equally effective at all grade levels and may be more effective for males. Drill

[13] In some studies, however, the only treatment or comparison contrasts were a calculator plus additional materials treatment versus a no-calculator control group (for example, Szetela, 1982). In these cases, these contrasts were included.

and practice is the only category of instructional software that shows stronger effects for serving as a substitute for conventional instruction, rather than as a supplement to it. It may be that such programs address students' instructional needs for practice adequately and efficiently, making substantial teacher intervention less important.

Variance in the findings of these reviews, and even wider variance in the individual studies, suggests that general conclusions should be made cautiously. Probably at work here are critical variables, including the quality of the particular software, but also contextual variables (e.g., settings, such as urban, suburban, or rural and student or family characteristics) and implementation variables (e.g., duration, use of the intervention as a supplement or substitution for conventional instruction, and fidelity of implementation; support and availability of resources, funds, and time; setting within the school) (Clements, 2007). One implication is that we should examine the influence of these variables when possible. The scarcity of information in this regard suggests a second implication, to which we will return: The field needs more comprehensive and nuanced reporting and analysis.

Table 18: What Prior Reviews Say About Drill and Practice

- General findings
  - — Generally improves students' performance when compared to conventional classroom instruction (median ES[a] = 0 .345)
  - — Greatest effects on computation, less on concepts and applications
  - — Positive effects on attitudes toward mathematics and instruction in mathematics
- Contextual variables
  - — No consistent differences by grade level
  - — No consistent differences by ability level
  - — Differences favored males
- Implementation variables
  - — Differences favored programs that substitute for other mathematics instruction
  - — Differences favored experimenter or teacher developer-designed (vs. commercially designed) software
  - — No consistent differences by program duration

[a] The median effect size is the median across meta-analyses that reported a pooled effect size. Pooled effect sizes for individual meta-analyses are provided in Appendix C; Drill and Practice is Table C-1.

b. The Task Group's Meta-Analysis of Drill and Practice Software

Table 19 presents the studies in the Task Group's meta-analysis of high-quality experimental and quasi-experimental studies on the effects of drill and practice software on achievement. From all the studies reviewed, only 12 met the criteria for inclusion. These 12 studies yielded a total of 18 effect sizes. Of these, 16 were positive (4 of which were statistically significant) and 2 negative (neither statistically significant), with a mean pooled effect size of 0.320, which was statistically significant. Although this is conjectural, there seems to be a trend for greater effects of interventions that were shorter and focused on developing the automaticity of specific skills. If this is indeed the case, this would be consistent with reports from other reviews.

The Task Group extended each meta-analysis to ascertain whether effect sizes were mediated by particular contextual and implementation variables. Results are presented in Table 20. A between-group p-value was calculated using CMA software to determine if the effect of a particular contextual or implementation variable was significant, and results of these analyses are shown in the same row as the name of the variable.

Table 19: Studies That Examine Effects of Drill and Practice Technology on Mathematics Achievement

Study	Design	Sample	Duration/Content	Contrast	Measure	Hedge's g	Standard Error	
Drill and Practice								
Ball, 1988[a]	QED	91 fourth-grade students in five classes in two schools	12 computer lessons over 6–8 weeks/ Fractions	Computer lessons on fractions vs. traditional fraction instruction	Posttest—Fractions	0.815	~	0.457
Campbell et al., 1987	RCT	48 third-grade students in a middle-class suburban school in the Southeast	20 minutes daily of D&P for 5 weeks (+30 min instruction for both T &C)/ Division of whole numbers/Milliken Mathematics Sequences program	Milliken Mathematics D&P vs. worksheets	Posttest—Division	0.445	(ns)	0.288
Carrier et al., 1985	RCT	144 fourth-grade students in six classrooms in a metropolitan school district	10–15 minutes per lesson over 14 weeks/ Multiplica-tion and division	Three different D&P vs. Worksheets	Post: Symbolic algorithms, mult. & division	0.228	(ns)	0.167
Emihovich & Miller, 1988	RCT	24 first-grade students in five classrooms in an elementary school in the Southeast	20, 30-min sessions (3 months)/ Addition/ subtraction, basic mathematics skills	Series of CAI software vs. regular reading and mathematics instruction	CTBS—Mathematics	0.407	(ns)	0.399
Fletcher et al., 1990[a]	RCT	41 third-grade students in rural Saskatchewan, Canada	Spring semester/ 3rd-grade mathematics	Milliken Mathematics Sequence vs. Control (traditional instruction + worksheets)	Canadian Tests of Basic Skills (CTBS)	0.412	(ns)	0.693
		38 fifth-grade students in rural Saskatchewan, Canada	Spring semester/ 5th-grade mathematics			0.338	(ns)	0.697
Fuchs et al., 2006	RCT	33 first-grade learning disabled students in nine classrooms in three Title 1 schools in a metropolitan school system	50, ten-minute sessions over 18 weeks/ Addition and subtraction	mathematics FLASH vs. spelling FLASH	Post: addition, subtraction, and story problems	0.177	(ns)	0.349
Kraus, 1981	RCT	19 second-grade students in one school in a southwestern Ohio city	5 sessions over a two week period (average 64 minutes)/ Fish Chase game: addition	Fish Chase vs. Hangman	Addition speed test	1.454	**	0.523

Continued on p. 6-118

Table 19, continued

Study	Design	Sample	Duration/Content	Contrast	Measure	Hedge's g		Standard Error
Drill and Practice								
McCollister et al., 1986	RCT	15 Kindergarten students in achievement level 1 in one public school in a large Southern city	6 sessions/ Numeral recognition and cardinal counting	How Many Squares computer program vs. Milton Bradley Flannel board	Pine & Burts (1984) numeral recognition and cardinal counting	-0.548	(ns)	0.529
		13 Kindergarten students in achievement level 2 in one public school in a large Southern city				0.344	(ns)	0.575
		25 Kindergarten students in achievement level 3 in one public school in a large Southern city				0.429	(ns)	0.405
Podell et al., 1992: Study 1	RCT	24 second-grade students in New York City public schools	Up to 10 15-minute sessions, three times per week/ Addition	Mathematics Blaster - Addition vs. Worksheets	Accuracy rate: mean trials to criterion	0.150	(ns)	0.408
		28 learning disabled students in Grades 2–4 in New York City public schools				0.783	*	0.397
Podell et al., 1992: Study 2	RCT	20 students in New York City public schools, ages 6–9	Up to 10 15-minute sessions, three times per week/ Subtraction	Mathematics Blaster - Subtraction vs. Worksheets	Accuracy rate: mean trials to criterion	0.627	(ns)	0.478
		22 learning disabled students in New York City public schools, ages 6–11				0.568	(ns)	0.435
Saracho, 1982[a]	QED	256 Spanish speaking migrant children attending third through sixth grade	3 hours a week, 60 hours for the academic year/ Elementary mathematics	D&P vs. regular classroom instruction	CTBS—Grades 3–6	-0.118	(ns)	0.304
Saunders & Bell, 1980	RCT	101 advanced Algebra students in four classes in one public high school	<1/2 hr per week for the school year/ Algebra II	Algebra problems using BASIC vs. regular Instruction	Cooperative Mathematics Test: Algebra II	0.136	(ns)	0.201
Watkins, 1986	RCT	82 first-grade students from a suburban Southwestern school	3, 15 min sessions per week (October through June)/ Mathematics Machine D&P	Mathematics D&P vs. Reading D&P	California Achievement Test	0.432	~	0.221
Heterogeneity								
Q-value	*Df (Q)*	*P-value*	*I-squared*	*Pooled ES (12 studies, 18 effect sizes)*		*Hedge's g*	*Standard Error*	
14.678	*17*	*0.619*	*0.000*			*0.320*	***	*0.078*

~ p < .10, * p < .05, ** p < .01, *** p < .001

[a] Data were adjusted for clustering that occurred within classrooms.

Contextual variables. The studies in the Task Group's meta-analysis yielded the following results regarding contextual variables:

- *Age or grade.* The effect of drill and practice software was confirmed as significantly effective at the elementary level, but there are not enough studies at the other levels to make any comparisons or other conclusions.
- *Ability.* There is no evidence that children with and without learning disabilities benefit differently from use of drill and practice software.

Implementation variables. These studies yielded the following results regarding implementation variables:

- *Duration.* Results of the Task Group meta-analysis indicate that duration of the intervention is not a significant moderator. This is consistent with findings from prior syntheses and meta-analyses.

- *Substitute versus supplement.* Consistent with other review findings, the effect sizes calculated in the Task Group meta-analysis were higher for drill and practice interventions that substituted for, rather than supplemented, classroom practice. Effect sizes were significant only for substitution implementations, but the difference between the two did not reach statistical significance.

- *Experimenter or teacher vs. commercial developer.* Effects sizes were larger for experimenter or teacher-developed, compared to commercial, drill and practice software, with no significant difference between them. This is consistent with prior review findings, both yielded significant effects.

In summary, the Task Group's meta-analysis of rigorous studies about the effects of drill and practice software produced a mean pooled effect size of 0.320 that is statistically significant. This finding is consistent with the conclusions of prior syntheses and meta-analyses. There is no solid evidence that students of different ability levels or disability status benefit differently. Results suggest higher effect sizes when drill and practice software is used as a substitute, rather than supplement, to instruction (although comparisons were not significant).

Table 20: Subgroup Analysis

	Drill and practice				Tutorials				Programming			
	N studies/ ES	Hedges g		Se	N studies/ ES	Hedges g		se	N studies/ ES	Hedges g		Se
Contextual variables												
Grade level		ns [a]				ns				*** [a]		
Elementary	11 / 17	0.352	***	0.084	2 / 4	0.235		0.214	**9 / 22**	**0.854**	*******	**0.134**
Middle School	0 / 0	na		na	3 / 4	0.138		0.088	**7 / 8**	**0.218**		**0.151**
High School	1 / 1	0.136		0.201	4 / 5	0.480	~	0.246	**0 / 0**	**na**		**na**
Mixed	0 / 0	na		na	1 / 1	0.379		0.441	**0 / 0**	**na**		**na**
Ability		ns				ns				na		
Learning disabled	3 / 4	0.303		0.258	4 / 5	0.238		0.143	0 / 0	na		na
Non-LD	10 / 13	0.356	***	0.087	6 / 9	0.356	**	0.136	14 / 30	0.674	***	0.115
Migrant (span speaking)	1 / 1	-0.118		0.698	0 / 0	na		na	0 / 0	na		na
Implementation variables												
Duration		ns				** [a]				ns		
Less than 4 weeks	3 / 8	0.492	**	0.184	**3 / 5**	**0.642**	*******	**0.181**	3 / 3	0.974	~	0.530
4 to 8 weeks	2 / 2	0.550	*	0.243	**0 / 0**	**na**		**na**	2 / 2	0.910	*	0.441
Greater than 8 weeks	7 / 8	0.223	*	0.095	**6 / 9**	**0.141**	**~**	**0.075**	9 / 25	0.625	***	0.124
Supplementation vs. substitution[a]		ns				ns				ns		
Supplement	4 / 4	0.290		0.250	2 / 3	0.425	*	0.175	7 / 9	0.721	**	0.208
Substitute	8 / 14	0.370	***	0.092	7 / 11	0.288	*	0.112	7 / 21	0.655	***	0.141
Curricular Integration[b]		ns				*** [a]				ns		
Low	2 / 2	0.766		0.637	**0 / 0**	**na**		**na**	1 / 1	-0.065		0.444
Medium	7 / 11	0.319	**	0.100	**3 / 6**	**0.037**		**0.074**	9 / 12	0.739	***	0.159
High	3 / 5	0.259	~	0.139	**6 / 8**	**0.503**	*******	**0.108**	4 / 17	0.682	***	0.168
Commercial vs. researcher		ns				** [a]				na		
Commercial	7 / 13	0.268	*	0.104	**5 / 8**	**0.092**		**0.068**	14 / 30	0.674	***	0.115
Researcher-designed	5 / 5	0.441	**	0.172	**4 / 6**	**0.516**	******	**0.165**	0 / 0	na		na
Total	12 / 18	0.320	***	0.078	9 / 14	0.302	**	0.099	14 / 30	0.674	***	0.115

~ p < .10, * p < .05, ** p < .01, *** p < .001
[a] Between group test of significance p-value (ns = not significant; na = not applicable).
[b] Supplementation categories are defined as follows: "supplement" = the technology treatment served as an addition to regular class time in mathematics; "substitute" = time spent on treatment technology substituted for at least some portion of math instruction/class time.
[c] Curricular integration characterizes the level of integration with the regular math curriculum. "Low" is categorized as little to no integration with math curricula; "Moderate" is defined as covering topics related to the regular math curricula and possibly coordinating instruction with technology; "High" is defined as curricula that was designed around the specific technology intervention.

2. Tutorials

a. Prior Syntheses and Meta-Analyses

Prior syntheses and meta-analyses (see summary Table 21) suggest that CAI tutorials improve students' performance compared to conventional instruction, with slightly greater effects on concepts and applications than on computation. The detailed effect size information from prior quantitative syntheses and meta-analyses are presented in Table C-2 in Appendix C. These syntheses most frequently identify tutorials as the most effective software category, when compared to drill and practice, simulations and games, and tools. They suggest that tutorials appear to be effective at all grade levels, particularly the higher grades and that tutorials are more effective when they supplement, rather than replace, conventional instruction, when they involve experimenter or teacher-developed, rather than commercially-developed, software, and when they are developed for a specific audience rather than a general audience. These findings come from syntheses and meta-analyses with different inclusion criteria than those used by the Instructional Practices Task Group.

Table 21: What Prior Reviews Say About Tutorials

- General findings
  - — Generally improves students' performance when compared to conventional classroom instruction, with a median pooled effect size of 0.38 (Table C-2)
  - — More researchers have claimed that tutorials are more effective than drill and practice (Burns & Bozeman, 1981; Khalili & Shashaani, 1994; Lee, 1990)
  - — Somewhat higher effect sizes for concepts and applications than computation
  - — No effects on attitudes
  - — Often low fidelity of program implementation
- Contextual variables
  - — Slight advantage for higher grade levels
  - — No consistent differences by ability level
  - — No consistent differences by gender
- Implementation variables
  - — Differences favoring programs that supplement instruction versus substitute
  - — Differences favoring experimenter or teacher developed programs vs. commercially developed software
  - — Differences favoring specific vs. a general audience
  - — No consistent differences by program duration

b. The Task Group's Meta-Analysis of Tutorial Software

Table 22 presents the studies in the Task Group's meta-analysis of high-quality experimental and quasi-experimental studies on the effects of tutorial and mixed tutorial and drill and practice software on achievement. From all the studies reviewed, only nine met the criteria for inclusion. These studies yielded a total of 14 effect sizes. Of these, 10 were positive (two of which were statistically significant), one negative, and three near zero, with a significant mean pooled effect size of 0.302. Those studies assessing mathematics achievement only had a mean pooled effect size of 0.288, which was statistically significant. Those that assessed problem-solving ability had a mean pooled effect size of 0.425, which also was statistically significant. Several contextual and implementation variables were examined.

Contextual variables. These studies yielded the following results regarding contexts (see Table 20).

- *Age or grade.* Similar to the results of the prior syntheses and meta-analyses, the IP meta-analysis indicates that tutorials have a slight advantage for high school students, but there are no significant differences between those effects and effects for other grade levels.

- *Ability.* Similar to the results of the prior syntheses and meta-analyses and the results for drill and practice, there was no significant difference between effects for students with and without learning disabilities, although the tendency for effects to be lower in schools with lower achievement needs further study.

Implementation variables. The Task Group's meta-analysis indicates the following regarding implementation variables.

- *Duration.* Tutorials were significantly more effective in studies in which interventions were less than 4 weeks in duration than those in which interventions were greater than 8 weeks. This must be interpreted with caution: Some treatments took place over many weeks, but the time students used the software remained limited. Thus, this finding may have more to do with limiting the confounding effects of other factors.

- *Substitute vs. supplement.* Tutorials were more effective when they supplement, rather than replace, conventional instruction, but the difference was not significant. However, they are significantly more effective when they are highly integrated with the regular mathematics curriculum (compared to medium integration, which had near-zero effects).

- *Experimenter or teacher vs. commercial developer.* Consistent with the prior syntheses and meta-analyses, there are stronger effects for experimenter or teacher-developed, compared to commercial, software.

The Task Group's meta-analysis of rigorous studies is consistent with the conclusions from the prior syntheses and meta-analyses. The Task Group analysis suggests that there is a suggestion that high school students may benefit more from tutorials than students at other grade levels (although comparisons were not significant). Tutorials were significantly more effective if they were highly integrated with the regular mathematics curriculum than when they were less integrated. Tutorials developed by a researcher or teacher had significantly greater positive effects than commercial software.

One of the studies in the Task Group set examining tutorials, Dynarski et al. (2007), includes two recent large randomized trial evaluations, and warrants particular attention because of the scale of these two studies (3,136 students in one study, 1,402 in a second study). The results suggest caution. The near-zero effect sizes in Dynarski et al. (0.071, -0.064) suggest that results of using tutorials are not guaranteed to be superior to standard instruction. Moreover, the results suggest additional questions that must be addressed in future research. Scaling up software interventions may be particularly difficult, and the more encouraging results from earlier and smaller studies (e.g., Fuchs et al., 2002, nonsignificant effect size of 0.586; Henderson et al., 1985, significant ES of 0.976; Thompson & Rickhuss, 1992, nonsignificant effect size of 0.774; or Wheeler & Regian, 1999, significant effect size of 0.517) may reflect efficacy under advantageous (i.e., closer to "ideal") conditions more than effectiveness at scale.

The Task Group findings indicate that tutorials are more effective if they are highly integrated into the curriculum (see Table 20), which requires that such integration be done either by the curriculum or software developers or by teachers. This is an extensive task that may demand additional time for both professional development and for work with colleagues on curriculum. A final issue is amount of use of tutorial software in classrooms. In the Dynarski et al. (2007) study, it considers teacher reports of tutorial software usage in the classroom. But when the study considered software recorded usage, usage in the classroom was much lower; compare teachers' report of 51 hours of usage to the products' reports of 17 hours for sixth grade; or teachers' report of 46 hours of usage to the products' reports of 15 hours for ninth grade. Even the teacher data are substantially lower than publishers' recommendations. This is consistent with the Panel's National Survey of Algebra Teachers that indicated low frequency of the use of technology (averaging "less than once a week;" (Hoffer, Venkataraman, Hedberg, & Shagle, 2007). These are issues in scaling up software use and suggest important questions for future research.

The direct implications of the Dynarski study are serious cautions to anyone who believes merely introducing technology will raise students' scores. This was a rigorous randomized control trials design conducted in 33 districts and 1,232 schools. The products being evaluated had been identified as being effective and widely used. Teachers were trained. There were no significant effects. Thus, educators must consider not only empirical evidence of effectiveness of a particular software package but also issues of scale-up, including integration with the extant curriculum, fidelity of implementation, including amount of use, and technological and pedagogical support.

To return to the software per se, studies also show that fine-tuning the mathematics and pedagogy in software can make a significant difference in learning. For example, in a study of another cognitive tutor (geometry), holding time-of-instruction constant, one group discussed why and how they used the strategy they used, and the other practiced more problems. The authors report that the former group had significantly greater understanding and showed greater transfer (Aleven & Koedinger, 2002).

Table 22: Studies That Examine Effects of Tutorials or Tutorials Plus Drill and Practice on Mathematics Achievement

Study	Design	Sample	Duration/Content	Contrast	Measure	Hedge's g	Standard Error	
Tutorial + Drill & Practice								
Clements, 1986	RCT	24 first-grade students from a middle-class midwestern school system	44 sessions (22 weeks)/ Elementary mathematics using CAI drill, tutorial, and problem-solving software (mathematics and reading)	CAI drill, tutorial, and problem-solving software (mathematics and reading) vs. traditional instruction	WRAT Mathematics score	0.397	(ns)	0.398
		24 third-grade students from a middle-class midwestern school system				0.142	(ns)	0.395
Dalton & Hannafin, 1988	RCT	22 low-achieving eighth-grade students from five sections in same school	Two lessons/ Geometry & area of circle	Computer initial and remedial instruction vs. traditional initial and remedial instruction	Mastery quiz: area of circle	-0.001	(ns)	0.426
		25 high-achieving eighth-grade students from five sections in same school				0.571	(ns)	0.395
Dynarski et al., 2007: Study 1[a]	RCT	3,136 sixth-grade students in 10 different districts across U.S., focused on lower achievement districts	One academic year, wide variation, but overall average use of the CAI was 17hrs/yr/ General mathematics	CAI vs. Control (standard instruction)	SAT-10 mathematics battery	0.071	(ns)	0.106
Dynarski et al., 2007: Study 2[a]	RCT	1,404 algebra students in 10 different districts across U.S., focused on lower achievement districts	One academic year, average use of the CAI was 15hrs/yr/ Algebra	CAI vs. Control (standard instruction)	ETS End-of-Course Algebra Assessment	-0.064	(ns)	0.117
Fuchs et al., 2002	RCT	18 fourth-grade students with mathematics disabilities in three schools in a southeastern city	24 sessions (twice per week for 12 weeks)/ Problem-solving	Computer vs. Control	Pooled problem-solving score (three subtests)	0.586	(ns)	0.486
		20 fourth-grade students with mathematics disabilities in three schools in a southeastern city	48 sessions (four times per week for 12 weeks)/ Problem-solving	Tutor + Computer vs. Tutor only (ES of Tutor + Computer vs. Control was 1.281)		-0.147	(ns)	0.448
Henderson et al., 1985	RCT	81 students attending five general mathematics or intro to algebra classes in one high school with large proportion of Latino students	Three modules (Three sessions)/ Factors and prime numbers	Computer-Video vs. Control	Combined Recognition and Constructed scores on Factors and Prime numbers test	0.976	***	0.234
Moore, 1988	RCT	117 seventh- and eighth-grade students in the lowest level of remedial mathematics in four middle schools	School year (Sept - May)/ Middle school mathematics instruction	Milliken Mathematics Sequences + written assignments vs. Direct Instruction using Mathematics for Individual Achievement	District Mathematics Placement Test	0.273	(ns)	0.185

Continued on p. 6-125

Table 22, continued

Study	Design	Sample	Duration/Content	Contrast	Measure	Hedge's g	Standard Error	
Tutorial + Drill & Practice								
Thomas & Rickhuss, 1992	RCT	17 high school students (average age 15) in one algebra class	1 week/ Algebra—solving equations	CAI MuMath/Solving Equations vs. noncomputer instruction	Solving equations	0.364	(ns)	0.465
			1 week/ Algebra—factorization	CAI MuMath/ Factorization vs. noncomputer instruction	Factorization	0.774	(ns)	0.480
Triffiletti et al., 1984	RCT	20 learning disabled students (ages 9–15) in a private school in Jacksonville, FL	school year (Sept - May)/ SPARK-80 Computerized Mathematics System	SPARK-80 vs. Resource Room	Key Mathematics Diagnostic Arithmetic Test (grade equiv)	0.379	(ns)	0.441
Wheeler & Regian, 1999[a]	RCT	493 ninth-grade students in 40 traditional mathematics instruction classes in Texas, New Mexico, and Ohio	one session per week, for the school year/ Word Problem Solving (WPS) Tutor	WPS vs. Control	Word Problem solving combo (concrete & abstract)	0.517	*	0.206
Heterogeneity								
Q-value	*df (Q)*	*P-value*	*I-squared*	*Pooled ES (9 studies, 14 effect sizes)*		*Hedge's g*	*Standard Error*	
24.385	*13*	*0.028*	*46.688*			*0.302*	**	*0.099*
Mathematics outcomes only								
Heterogeneity								
Q-value	*df (Q)*	*P-value*	*I-squared*	*Pooled ES (7 studies, 11 effect sizes)*		*Hedge's g*	*Standard Error*	
20.519	*10*	*0.025*	*51.264*			*0.288*	*	*0.112*
Problem-solving outcomes only								
Heterogeneity								
Q-value	*df (Q)*	*P-value*	*I-squared*	*Pooled ES (two studies, three effect sizes)*		*Hedge's g*	*Standard Error*	
1.939	*2*	*0.379*	*0.000*			*0.425*	*	*0.175*

Note: The 2 studies with problem-solving outcomes are Fuchs et al. (2002) and Wheeler & Regian (1999), all others have mathematics outcomes.
~ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$
[a] Data were adjusted for clustering that occurred within classrooms.

3. Tools: Calculators and Graphing Calculators

Among the many categories of technology, calculators, including graphing calculators, have probably generated the greatest amount of debate. Some have championed their use in developing problem-solving ability by allowing students to perform far more, and more complex, arithmetic operations than would have been possible without technology. Others have bemoaned their misuse as simple fact checkers. A concern is that calculators have an insidious effect on paper-and-pencil arithmetic and algebraic skills.

Calculators have been used in mathematics education for 70 years, since Emmett Betts engaged students with calculating machines in 1937. They metamorphosed from the original bulky and expensive machines to the electronic calculators of the 1960s, the inexpensive handheld, four-function calculators of the 1970s, and the wide variety of basic, scientific, and graphing calculators available today.

The usefulness of calculators in homes and businesses may seem clear, but their use in education, at first blush, seems equally problematic—students should *learn* to compute without calculators. The Panel's survey of the nation's algebra teachers indicated that the use of calculators in prior grades was one of their concerns (Hoffer et al., 2007).

a. Prior Syntheses and Meta-Analyses

Previous reviews (see summary in Table 23) have suggested that calculators of all types, basic, scientific and graphing, may benefit students' achievement in and attitudes toward mathematics (see the detailed effect size information from prior quantitative syntheses and meta-analyses in Tables C-3 and C-4 in Appendix C). Effects are usually more positive when students are allowed to use calculators during testing. Effects on concepts, contrary to perhaps the most common concern, are near zero but positive, and effects on problem solving were positive.

Table 23: What Prior Reviews Say About Calculators

- General findings
  - — Generally improve students achievement and attitudes (median pooled effect size on computation, 0.41, Table C-3)
  - — Generally improve mathematical problem solving (median pooled effect size, 0.19) but little or no effect on conceptual development (median pooled effect size near zero)
  - — Most effective in facilitating learning of operational skills ("operational" indicating that the report was unclear as to whether the instrument assessed the computational, conceptual, or both domains)
  - — Graphing calculators particularly effective for conceptual skills (Table C-4)
  - — All the effects mentioned are lower when testing without calculators
- Contextual variables
  - — No consistent differences by grade levels
  - — Some differences by ability level
- Implementation variables
  - — Special calculator instruction may have more positive effects (pedagogical uses vs. merely providing calculators) on computational, operational, and problem solving competencies

b. The Task Group's Meta-Analysis of Calculators

Turning to the Task Group's meta-analysis of rigorous studies, Tables 24, 25, and 26 provide individual study and pooled effect sizes for each of the three focal outcomes: computation (Table 24), problem solving (Table 25), and concepts (Table 26).[14] The tables disaggregate studies analyzed at the student level from studies analyzed at the classroom level.

[14] Many of the studies had a concern that calculators may impede computational achievement, and thus were testing to see whether use of calculators had a positive or negative effect on computation. The studies that assessed the effects on problem solving and concepts often hypothesized that calculators would improve these outcomes.

Meta-analytic pooled effect sizes and accompanying statistics are based on pooling of similarly aggregated effect sizes. In other words, studies that analyzed data at the student-level are pooled together, and similarly, studies that analyzed at the classroom level are pooled together.

It was the Task Group's hope to discern through this meta-analysis any differences that might exist between the effects of graphing calculators and non-graphing calculators. However, nearly all of the peer-reviewed published studies using graphing calculators examine the effects on students in advanced mathematics courses (such as Algebra 2, Trigonometry, Precalculus, and Calculus). As a result, only one of the included studies (Graham & Thomas, 2000) used a graphing calculator.

Table 24 presents studies that contrast treatment condition using calculators with a non-calculator control condition on computational outcomes. Seven of the studies (one with effects at four different grade levels) analyzed data at the individual student level, and the remaining three (including nine comparisons) used the classroom or teacher as the unit of analysis. For the student-level set, then, ten effect sizes were calculated. The pooled effect size is 0.319, which borders on statistical significance. In only one of the included studies (Wheatley, 1980), students were allowed to use calculators during assessment. Once that study is removed from the analysis, the pooled effect size is 0.307, and is not statistically significant. For the classroom-level set, nine effect sizes were calculated. The mean effect size is -0.085, which is not statistically significant.

Outcomes of studies that examine the effects of calculator use on problem solving are presented in Table 25. The seven comparisons in the student-level set (note that four were from a single study, Szetela (1982)) yielded a mean pooled effect size of 0.304, which borders on statistical significance. The four comparisons from the classroom-level set yielded a mean pooled effect size of -0.063, which was not statistically significant.

Regarding outcomes on measures of conceptual development, presented in Table 26, the four comparisons in the student-level set yielded a mean pooled effect size of 0.278, which is not statistically significant. The three comparisons from the classroom-level set yielded a mean pooled effect size of 0.128, which was not statistically significant.

Several contextual and implementation variables were examined. These are summarized in Table C-8 in Appendix C.

Contextual variables. These studies yielded the following results regarding contexts.

- *Age or grade.* There were no statistically significant differences among the effect sizes for elementary school-aged students (ES = 0.367, ns, five effect sizes within four studies) versus secondary school-aged students (ES = 0.113, ns, four effect sizes within three studies) for computation, nor for applications or concepts. However, this is based on a small sample of studies and thus there may not be sufficient power to detect differences in effect sizes (e.g., differences in effect sizes for applications, which were larger for secondary than for elementary, and statistically significant only for secondary, should be evaluated in future research).

Implementation variables. The Task Group's meta-analysis indicates the following regarding implementation variables.

- *Duration.* Studies that provided interventions for shorter periods (less than 3 months) had stronger effects on computation than studies that extended over longer time periods (3 months or longer). Specifically, the pooled effect size, under random effects assumptions, for interventions taking place for less than three months was a statistically significant 0.503 ($p < .05$, seven effect sizes within five studies); while the effect for studies taking place for longer than 3 months was not statistically significant (ES = -0.134, three effect sizes within three studies).[15] Although this would be an interesting finding if valid, it is based on a small sample, and thus it is likely that other factors unrelated to program duration may have led to this result. There were no such significant differences for applications or concepts.

- *Special calculator instruction.* Using alternative interventions or enhancing the intervention did not, as a whole, yield significantly higher effect sizes.

In summary, effect sizes of the Task Group's meta-analysis to examine the effects of calculator use on computation skills are smaller than those reported in prior syntheses and meta-analyses.

Concerning the impact of calculator use on problem-solving competencies, the Task Group's meta-analysis at the student level yielded a borderline significant, positive effect, but classroom-level analyses were near zero. The results in Table 25 are mainly for the outcomes in which students were not allowed to use calculators to solve problems on the assessments, Wheatley (1980) is the only study that includes outcomes where calculators were allowed. When looking at outcomes in which calculators were permitted on the assessments, effects were more positive (e.g., two of the four contrasts examined from Szetela (1982) reached statistical significance, see Table C-9 in Appendix C). Assessing proficiency with the same tools available as were available during instruction may be viewed as constituting a valid comparison, perhaps especially for problem-solving outcomes. Comparing these conclusions to those in the syntheses of previous reviews, the pattern is similar to what was found for computation: The effect sizes in the present meta-analyses are smaller.

Effect sizes on conceptual development tended to be positive, favoring the calculator treatments, but generally small and all nonsignificant (see Table 26). This is consistent with the prior syntheses and meta-analyses, which reported near-zero pooled effect sizes.

[15] The five studies taking place for less than three months include: Schnur & Lang (1976), Standifer & Maples (1981), Szetela (1980), the Grade 3, 5, and 7 sample of Szetela (1982), and Wheatley (1980). The three studies taking place for three or more months include: Campbell & Virgin (1976), Standifer & Maples (1982), and the Grade 8 sample of Szetela (1982).

The two specific alternative interventions or enhancements to calculators (programmed feedback calculators and supplementary materials) used in studies identified as high quality by the Task Group (Standifer & Maples, 1981, 1982; Duffy & Thompson, 1980) did not yield any significant effect sizes. This is in contrast to the findings of the prior syntheses and meta-analyses (and to the findings for formative assessment discussed in the Task Group report), which include a wider range of enhancements, including more recent interventions. More research needs to be conducted, for example, on essential distinctions such as between functional and pedagogical use.

Finally, there are several important caveats. Effects of calculator use, especially appropriate versus inappropriate pedagogical use in the early grades, have not been adequately researched. Similarly, long-term effects of inappropriate calculator use may be negative (Wilson & Naiman, 2004); there is no reliable evidence. The Task Group's meta-analysis could not include adequate research on graphing calculators; high-quality research is needed regarding this type of calculator.

Table 24: Studies That Investigate the Effects of Calculators on Computation Outcomes

Study	**Design**	**Sample**	**Duration/ Content**	**Contrast**	**Measure**	**Subgroup**	**Hedge's g**	**Standard Error**	
Computation Outcomes									
Student-level analyses									
Campbell & Virgin, 1976[b]	Quasi	252 fifth- and sixth-graders in two North York elementary schools (Canada).	7 months/ Basic computation	Calculators to check work vs. No calculators	Metropolitan Achievement Test computation score	Overall	0.022	(ns)	0.642
Schnur & Lang, 1976[a]	RCT	60 youths ages 9 to 14 in four summer compensatory education classes in rural Iowa.	26 weeks/ Computation	Calculators to check work and compute subset of problems vs. Compensatory education program	Computational Skills Program Computational Test	Overall	0.855	~	0.512
Standifer & Maples, 1981[a]	RCT	141 students in 6 third-grade classrooms in Monroe, Louisiana	11 weeks/ Computation	Hand-held, four function calculator vs. No calculator in regular mathematics curriculum (see Table C-9 for effects of programmed feedback calculator vs. No calculator in regular mathematics curriculum)	Science Research Associates Assessment: computation score	Overall	0.635	(ns)	0.398
Standifer & Maples, 1982[a]	RCT	113 students in 10 third- and fourth-grade Title I compensatory mathematics classrooms in Monroe, Louisiana	5 months/ Computation	Hand-held, four function calculator vs. General remedial mathematics curriculum (see Table C-9 for effects of experimental group 2 using programmed-feedback calculators + regular remedial curriculum)	Science Research Associates Assessment: computation score	Overall	0.023	(ns)	0.329
Szetela, 1980	RCT	39 students in two seventh-grade classes in a middle class elementary school (likely in Canada)	3 weeks/ Focus on learning the concept of ratios	Calculator-based instruction with four-function calculator vs. Instruction without calculators	Researcher-designed test on ratios	Overall	0.322	(ns)	0.316

Continued on p. 6-130

Table 24, continued

Study	Design	Sample	Duration/ Content	Contrast	Measure	Subgroup	Hedge's g	Standard Error	
Szetela, 1982	RCT for Grades 3, 5 and 7 and Quasi for Grade 8	46 third-grade students in a middle income school in Richmond, British Columbia	8 weeks/ All grades focused on problem solving. Grade specific foci included: Grade 3: whole number operations in multiplication, basic division; Grade 5: introduction to decimals, operations with decimals; Grades 7 and 8: decimals, ratios, and percents.	Regular instruction plus calculator-specific materials vs. Regular instructional activities	Researcher-designed computational skills (16 items); tailored to grade level	Third grade	1.337	***	0.323
		33 fourth-grade students in a middle income school in Richmond, British Columbia				Fifth grade	-0.307	(ns)	0.342
		47 seventh-grade students in a middle income school in Richmond, British Columbia				Seventh grade	0.279	(ns)	0.288
		54 eighth-grade students in a middle income school in Richmond, British Columbia				Eighth grade	-0.267	(ns)	0.270
Wheatley, 1980[c]	Quasi	44 sixth-grade students in two classes (same teacher) in an elementary school in a Midwestern university town	6 weeks/ Problem solving	Problem solving with calculators vs. Problem solving intervention without calculators	Measure of computational errors (reverse coded) on five researcher-designed problems	Overall	0.573	(ns)	0.691
Heterogeneity									
Q-value	*df (Q)*	*P-value*	*I-squared*	*Pooled ES: student level (7 studies, 10 effect sizes)*			*Hedge's g*	*Standard Error*	
20.822	*9*	*.013*	*56.776*				*0.319*	*~*	*0.077*
Classroom-level analyses									
Duffy & Thompson, 1980	RCT	Approx. 135 students in 20 fourth-grade classrooms in Columbus, Ohio	26 weeks/ Application problems, decimals, rounding, estimation	Calculators only plus regular mathematics program vs. Regular mathematics curriculum (see Table C-9 for effects of calculator plus instructional packages for teachers, plus regular mathematics program)	CTBS computation score	Fourth grade	0.037	(ns)	0.428
		Approx. 122 students in 18 fifth-grade classrooms in Columbus, Ohio				Fifth grade	0.325	(ns)	0.452
		Approx. 129 students in 19 sixth-grade classrooms in Columbus, Ohio				Sixth grade	-0.395	(ns)	0.444
Szetela & Super, 1987	Quasi	Approx. 424 students in 21 seventh-grade classrooms in an urban-rural district in Canada	One school year/ Problem solving	Problem solving with calculators vs. Problem solving intervention without calculators	Rational Numbers test—40 item test used in British Columbia	Overall	-0.076	(ns)	0.423

Continued on p. 6-131

Table 24, continued

Study	Design	Sample	Duration/ Content	Contrast	Measure	Subgroup	Hedge's g	Standard Error	
Wheatley & Shumway, 1979	RCT	Students in 50 classrooms in second through sixth grade in five Midwestern states. Ten classrooms in each grade level.	7 months/ Basic four-function calculator/ computation	General calculator use (teachers trained but determine how they will implement) vs. No calculators/regular mathematics program	Stanford Achievement Test - Computation score	Second grade	-0.603	(ns)	0.587
						Third grade	0.352	(ns)	0.577
						Fourth grade	-0.460	(ns)	0.580
						Fifth grade	-0.434	(ns)	0.579
						Sixth grade	0.315	(ns)	0.576
Heterogeneity									
Q-value	*df (Q)*	*P-value*	*I-squared*	*Pooled ES: student level (three studies, nine effect sizes)*			*Hedge's g*	*Standard Error*	
4.010	*8*	*0.856*	*0.000*				*-0.085*	*(ns)*	*0.167*

~ p < .10, * p < .05, ** p < .01, *** p < .001
[a] Data were adjusted for clustering that occurred within classrooms.
[b] Data were adjusted for clustering that occurred within schools.
[c] The treatment group was allowed to use a calculator during assessment for this outcome.

Table 25: Studies That Investigate the Effects of Calculators on Problem Solving Outcomes

Study	Design	Sample	Duration/ Content	Contrast	Measure	Subgroup	Hedge's g	Standard Error	
Problem Solving Outcomes									
Student-level analyses									
Campbell & Virgin, 1976[b]	Quasi	150 fifth- and sixth-graders in two North York elementary schools (Canada)	7 months/ Basic computation	Calculators to check work vs. No calculators	MAT computation score	Overall	*0.238*	*(ns)*	*0.642*
Szetela, 1980	RCT	39 students in two seventh grade classes in a middle class elementary school (likely in Canada)	3 weeks/ Focus on learning the concept of ratios	Calculator-based instruction with four-function calculator vs. Instruction without calculators	Researcher-designed ratio problems test	Overall	0.869	**	0.329
Szetela, 1982	RCT for Grades 3, 5 and 7 and Quasi for Grade 8	46 third-grade students in a middle income school in Richmond, British Columbia	8 weeks/ All grades focused on problem solving. Grade specific foci included: Grade 3: whole number operations in multiplication, basic division; Grade 5: introduction to decimals, operations with decimals; Grades 7 and 8: decimals, ratios, and percents.	Regular instruction plus calculator-specific materials vs. Regular instructional activities	Researcher designed problem-solving post-test (10 items)—correct answer measure was used to calculate effect sizes (other measure available were problems attempted and correct operation used)	Third grade	0.522	~	0.296
		33 fourth-grade students in a middle income school in Richmond, British Columbia				Fifth grade	-0.507	(ns)	0.345
		47 seventh-grade students in a middle income school in Richmond, British Columbia				Seventh grade	0.227	(ns)	0.288
		54 eighth-grade students in a middle income school in Richmond, British Columbia				Eighth grade	0.344	(ns)	0.270
Wheatley, 1980[c]	Quasi	44 sixth-grade students in two classes (same teacher) in an elementary school in a Midwestern university town	6 weeks/ Problem solving	Problem-solving with calculators vs. Problem-solving intervention without calculators	Process score (processes used to solve problems)	Overall	0.353	(ns)	0.689
Heterogeneity									
Q-value	*df (Q)*	*P-value*	*I-squared*	*Pooled ES: student level (four studies, seven effect sizes)*			*Hedge's g*	*Standard Error*	
9.105	*6*	*0.168*	*34.101*				*0.304*	*~*	*0.167*

Continued on p. 6-133

Table 25, continued

Study	Design	Sample	Duration/ Content	Contrast	Measure	Subgroup	Hedge's g		Standard Error
Classroom-level analyses									
Duffy & Thompson, 1980	RCT	Approx. 135 students in 20 fourth-grade classrooms in Columbus, OH	26 weeks/ Application problems, decimals, rounding, estimation	Calculators only plus regular mathematics program vs. Regular mathematics curriculum (see Table C-9 for effects of calculator plus instructional packages for teachers, plus regular mathematics program)	CTBS applications score	Fourth grade	-0.413	(ns)	0.433
		Approx. 122 students in 18 fifth-grade classrooms in Columbus, OH				Fifth grade	-0.161	(ns)	0.440
		Approx. 129 students in 19 sixth-grade classrooms in Columbus, OH				Sixth grade	0.178	(ns)	0.440
Szetela & Super, 1987	Quasi	Approx. 424 students in 21 seventh-grade classrooms in an urban-rural district in Canada	One school year/ Problem solving	Problem solving with calculators vs. Problem-solving intervention without calculators (see Table C-9 for larger positive effects where calculator group was able to use calculators)	Combination of two researcher designed problem solving measures: translation problems (20 items) and process problems (20 items)	Overall	0.140	(ns)	0.424
Heterogeneity									
Q-value	*df (Q)*	*P-value*	*I-squared*	*Pooled ES: student level (two studies, four effect sizes)*			*Hedge's g*		*Standard Error*
1.230	*3*	*0.746*	*0.000*				*-0.063*	*(ns)*	*0.217*

~ p < .10, * p < .05, ** p < .01, *** p < .001
[a] Data were adjusted for clustering that occurred within classrooms.
[b] Data were adjusted for clustering that occurred within schools.
[c] The treatment group was allowed to use a calculator during assessment for this outcome.

Table 26: Studies That Investigate the Effects of Calculators on Concept Outcomes

Study	Design	Sample	Duration/ Content	Contrast	Measure	Subgroup	Hedge's g	Standard Error	
Concepts Outcomes									
Student-level analyses									
Campbell & Virgin, 1976[b]	Quasi	252 fifth- and sixth-graders in two North York elementary schools (Canada).	7 months/ Basic computation	Calculators to check work vs. No calculators	MAT concepts score	Overall	0.129	(ns)	0.642
Graham & Thomas, 2000[a]	Quasi	84 students in Grades 9 and 10 in two schools in New Zealand	3 weeks/Algebra	Graphic calculator used to learn algebraic variables vs. Standard algebra instruction	Kuchmann (1981) designed to measure algebraic understanding	Overall	0.328	(ns)	0.489
Standifer & Maples, 1981[a]	RCT	141 students in six third-grade classrooms in Monroe, LA	11 weeks/ Computation	Hand-held, four function calculator vs. No calculator in regular mathematics curriculum (see Table C-9 for effects of programmed feedback calculator vs. No calculator in regular mathematics curriculum)	Science Research Associates Assessment: computation score	Overall	-0.076	(ns)	0.395
Standifer & Maples, 1982[a]	RCT	113 students in 10 third- and fourth-grade classrooms in Monroe, LA	5 months/ Computation	Hand-held, four function calculator vs. General remedial mathematics curriculum (see Table C-9 for effects of experimental group 2 using programmed-feedback calculators + regular remedial curriculum)	Science Research Associates Assessment: computation score	Overall	0.546	(ns)	0.332
Heterogeneity									
Q-value	*df (Q)*	*P-value*	*I-squared*	*Pooled ES: student level (4 studies, 4 effect sizes)*			*Hedge's g*	*Standard Error*	
1.519	*3*	*0.678*	*0.000*				*0.278*	*(ns)*	*0.213*
Classroom-level analyses									
Duffy & Thompson, 1980	RCT	Approx. 135 students in 20 fourth-grade classrooms in Columbus, OH	26 weeks/ Application problems, decimals, rounding, estimation	Calculators only plus regular mathematics program vs. Regular mathematics curriculum (see Table C-9 for effects of calculator plus instructional packages for teachers, plus regular mathematics program)	CTBS concepts score	Fourth grade	0.063	(ns)	0.428
		Approx. 122 students in 18 fifth-grade classrooms in Columbus, OH				Fifth grade	0.221	(ns)	0.440
		Approx. 129 students in 19 sixth-grade classrooms in Columbus, OH				Sixth grade	0.103	(ns)	0.439
Heterogeneity									
Q-value	*df (Q)*	*P-value*	*I-squared*	*Pooled ES: student level (one study, three effect sizes)*			*Hedge's g*	*Standard Error*	
0.071	*2*	*0.965*	*0.000*				*0.128*	*(ns)*	*0.252*

~ p < .10, * p < .05, ** p < .01, *** p < .001
[a] Data were adjusted for clustering that occurred within classrooms.
[b] Data were adjusted for clustering that occurred within schools.
[c] The treatment group was allowed to use a calculator during assessment for this outcome.

4. Computer Programming

One of the early uses of educational technology in education was engaging students in programming computers as a way to explore, learn, or apply and practice mathematical ideas. For example, the original developers of Logo developed this programming language to serve as a conceptual framework for learning mathematics (Feurzeig & Lukas, 1971; Papert, 1980). Classroom observations suggested that children use certain mathematical concepts in Logo programming. As an illustration, first-graders use such mathematical notions as number, arithmetic, estimation, measure, patterning, proportion, symmetry, inversion, and compensation (Kull, 1986). Similar observations of intermediate graders indicated that Logo may make it possible to explore certain mathematical concepts, such as angle measure or recursion, earlier than is currently believed (Carmichael, Burnett, Higginson, Moore, & Pollard, 1985; Papert, Watt, diSessa, & Weir, 1979). Here the Task Group investigates whether engaging students in computer programming has significant effects on their mathematics achievement and problem-solving ability.

a. Prior Syntheses and Meta-Analyses

Detailed effect size information from prior quantitative syntheses and meta-analyses are presented in Table C-5 in Appendix C. Previous reviews (see summary Table 27) indicate that programming improves students' performance compared to conventional instruction, with the greatest effects on concepts and applications, especially geometric concepts, and weaker effects on computation. They also have indicated that programming positively affects problem solving, as well as attitudes toward mathematics and instruction in mathematics, more so than other software categories. On the basis of prior syntheses and meta-analyses, computer programming appears to have the same effectiveness at various grade levels. There is some evidence it is more effective for students of average, rather than low or high socioeconomic status (SES). Earlier syntheses and meta-analyses have argued that programming is somewhat more effective when it supplements, rather than replaces, conventional instruction, consistent with suggestions for mediated instruction of programming. Certain computer languages, especially the Logo computer language, have stronger positive effects than other computer languages. Other syntheses have similarly concluded that direct teacher involvement and better designed languages result in better instruction, and provide more guidance for instruction (Clements & Sarama, 1997). As with other types of software, Logo programming can be particularly effective when embedded in a curriculum and then in a context that includes professional development for teachers.

Table 27: What Prior Reviews Say About Programming Interventions

- General Findings
  - — Logo programming can increase students' mathematical achievement, especially if it is integrated into a coherent curriculum with teacher mediation (Clements & Sarama, 1997)
  - — The median pooled effect size for mathematics achievement across the meta-analyses is 0.35; for problem solving, the median is 0.285 (see Table C-5)
  - — Impacts more likely on concepts and applications as opposed to computation
  - — Positive effects on attitudes toward mathematics and instruction
- Contextual variables
  - — Differences favoring elementary school-age (vs. secondary) in achievement (similar for problem solving)
  - — Differences favoring average SES students vs. either high or low SES
  - — No consistent differences in ability level
- Implementation variables
  - — Differences favoring shorter duration programs (up to 18 weeks; based on only one meta-analysis)
  - — Differences favoring programs that supplement rather than substitute for other mathematics instruction for problem solving (substitution is slightly higher for achievement). Narrative reviews conclude that better outcomes result from curriculum integration and mediated teaching
  - — Differences favoring computer programs designed to support learning (such as Logo)

b. The Task Group's Meta-Analysis of Computer Programming Interventions

Table 28 presents the studies in the Task Group's meta-analysis of high-quality experimental and quasi-experimental studies on the effects of students' engaging in computer programming on their achievement. From all the studies reviewed, only 14 met the criteria for inclusion. These 14 studies yielded a total of 30 effect sizes. Of these, 24 were positive, 1 negative, and 5 near zero, with a mean pooled effect size on combined outcome measures of 0.674, which was statistically significant. Those assessing mathematics achievement only had a mean pooled effect size of 0.698, which also was statistically significant. (An important note is that some of these interventions involved changes in curriculum, using technology, but also altering content and teaching.) Those that assessed problem solving ability had a mean pooled effect size of 0.518, which was also statistically significant.

Although only two studies (Johnson-Gentile et al., 1994; Ortiz & MacGregor, 1991) reported effects on retention, both reported a larger effect size for the delayed, compared to the immediate, posttests (1.901 immediate, significant ES; 2.410 delayed for Johnson-Gentile et al.; 0.437 immediate, bordering on significant .898 for Ortiz & MacGregor). These findings suggest that computer programming, possibly due to the more extensive processing (due to the programming activity per se) over multiple modalities (e.g., numeric or symbolic and visual or graphic) or the ability to actively submit one's ideas for evaluation and feedback (e.g., did the program run as expected) facilitates students' development of higher level conceptual structures. That is, computer programming requires a complete, precise, and abstract explication, potentially leading to conceptually richer concepts. Students specify steps to a noninterpretive agent, with thorough specification and detail, then observe, reflect on, and correct. The computer serves as an explicative agent.

Several of these studies also compared computer programming to a CAI-based treatment, and so were not included in the basic meta-analysis in Table 28. Showing consistently higher scores for the computer programming than the CAI groups (but none reaching levels of statistical significance), these contrasts can be found in Appendix C in Table C-10. Several contextual and implementation variables may have contributed to the inconsistency.

Contextual variables. These studies about programming yielded the following results regarding contexts (Table 20).

- *Age or grade.* Effects were significantly higher when used with elementary school students than with middle school students, consistent with previous reviews.

Implementation variables. These studies yielded the following regarding implementation variables.

- *Duration.* There were no significant differences for interventions of different durations.

- *Substitute versus supplement.* Both substitution and supplementation programming treatments had statistically significant positive effects (ES 0.721 and ES 0.655), and the differences in effects between these two types of treatments were not significant.

- *Level of integration.* The differences across subcategories of curricular integration also did not reach statistical significance, but there is a clear pattern of effect sizes in which stronger effects are related to high (0.682, significant) or medium (0.739, significant), compared to low (-0.065, not significant) integration.

The Task Group's meta-analysis of rigorous studies on the effects of computer programming on mathematics achievement supports the conclusions of the previous syntheses, with a significant mean pooled effect size of 0.698 for mathematics achievement and 0.518 for problem solving (See Table 28) (compare to the median pooled effect sizes of 0.35 and 0.258, respectively, for the previous meta-analyses). Effects were higher for elementary school students than for older students. There is a suggestion that greater curricular integration yields stronger positive effects. Further, this meta-analysis suggested a result not previously revealed—that results for delayed posttests might be greater than those for immediate posttests.

Table 28: Studies That Examine Effects of Computer Programming on Mathematics Achievement

Study	Design	Sample	Duration/ Content	Contrast	Measure	Hedge's g	Standard Error	
Programming								
Battista & Clements, 1986	RCT	12 fourth-grade students in a midwestern middle school	42 sessions (two 40-min per week)/ LOGO	Logo vs. C (computer literacy)	Problem-solving Tests 1&2, Combined	0.660		0.609
		26 sixth-grade students in a midwestern middle school				0.049		0.392
Blume & Schoen, 1988	QED	50 eighth-graders in two midwestern junior high schools	A semester-long class/ BASIC	Basic vs. C	Combined problem solving and logic	-0.065		0.444
Clements, 1986	RCT	24 first-grade students from a middle-class midwestern school system	44 sessions (22 weeks)/ LOGO	Logo vs. C	WRAT Mathematics score	1.072	*	0.423
		24 third-grade students from a middle-class midwestern school system				0.636		0.405
Clements et al., 2001	QED	51 Kindergarten students in a school near Kent, OH	Incorporated into classes over entire academic year/ LOGO and geometry	Logo vs. Control	Geometry	2.842	***	0.609
		71 Kindergarten students in a school near Buffalo, NY				0.121		0.495
		87 first-grade students in a school near Kent, OH				0.938	~	0.495
		92 first-grade students in a school near Buffalo, NY				0.394		0.486
		103 second-grade students in a school near Kent, OH				0.009		0.481
		96 second-grade students in a school near Buffalo, NY				1.457	**	0.502
		56 third-grade students in a school near Kent, OH				0.571		0.511
		47 third-grade students in a school near Buffalo, NY				0.674		0.524
		158 fourth-grade students in a school near Kent, OH				0.184		0.470
		92 fourth-grade students in a school near Buffalo, NY				0.353		0.486
		103 fifth-grade students in a school near Kent, OH				0.093		0.481
		95 fifth-grade students in two schools, one near Kent, OH (Site 1) and one in Buffalo, NY (Site 2)				0.982	*	0.492
		141 sixth-grade students in a school near Kent, OH				0.526		0.474
		108 sixth-grade students in a school near Buffalo, NY				0.011		0.479
Clements & Battista, 1989	RCT	48 third-grade students of seven teachers from a middle class midwestern school	78 sessions, three 45-55 min per week (26 weeks)/ LOGO	Logo vs. C (computer composition/music + some Logo)	Combined posttest	1.495	***	0.328
Degelman et al., 1986	RCT	15 Kindergarten students attending a private day care center	15 minutes/day for five weeks/ LOGO	Logo vs. C	Problem solving (proportion correct)	1.284	*	0.576

Continued on p. 6-139

Table 28, continued

Study	Design	Sample	Duration/ Content	Contrast	Measure	Hedge's g	Standard Error	
Emihovich & Miller, 1988	RCT	24 first-grade students in five classrooms in an elementary school in the southeast	20, 30-min sessions (3 months)/ LOGO	Logo vs. C	CTBS - Mathematics	0.719	~	0.408
Johnson-Gentile et al., 1994	QED	150 fifth- and sixth-graders in six classrooms in two schools, one urban and one suburban	8 class days/ LOGO and geometry	Logo vs. control	Logo Geometry Motions Unit Posttest	1.901	***	0.420
Kapa, 1999	RCT	15 fifth-grade students from four classes in two elementary schools in Tel Aviv, Israel, working individually	twice per week for 45 min for semester/ LOGO-STAT-programming and graphing	individual: LOGO-STAT vs. C (Q-Text, linguistic problem-solving)	Problem solving (range 1–6)	0.520		0.496
Kapa, 1999[a]		30 fifth-grade students from four classes in two elementary schools in Tel Aviv, Israel, working in pairs		pairs: LOGO-STAT vs. C (Q-Text, linguistic problem-solving)		0.697	~	0.412
Lehrer & Randle, 1987	RCT	24 first-grade students in a low SES New York City school	35 sessions, twice per week 20–25 min, 5 months/ LOGO	Logo vs. C	TOH (avg of TOH 1–3)	1.254	**	0.434
Oprea, 1988[a]	QED	54 sixth-grade students in three schools in a small midwestern city	6 weeks/ BASIC applied to mathematics content	Wholistic BASIC vs. C	Mathematical Generalization Instrument	0.391		0.679
Ortiz & MacGregor, 1991	RCT	59 sixth-grade students from four classrooms in two metropolitan area public schools	5, 50-min sessions/ LOGO and the concept of a variable	Logo vs. textbook-based instruction on concept of variable	Concept of variable instrument	0.437	~	0.260
Thompson & Wang, 1988[a]	QED	40 sixth-grade students from two classrooms (taught by the same teacher) in one school	3, 45-min sessions/ LOGO and graphing skills	Logo vs. C	Posttest-Cartesian coordinates	0.538		0.696
Turner & Land, 1988[a]	QED	153 middle school students in seven classrooms in four inner-city midwestern public schools	1hr/week, 16 weeks/ LOGO and angles and distance, variables, rectangular coordinate systems, negative numbers, etc.	Logo vs. C	Mathematics Multiple Choice Posttest	-0.267		0.471

Heterogeneity

Q-value	*df (Q)*	*P-value*	*I-squared*		*Hedge's g*	*Standard Error*	
53.503	*29*	*0.004*	*45.797*	*Pooled ES (14 studies, 30 effect sizes)*	*0.674*	*****	*0.115*

Mathematics outcomes only

Heterogeneity

Q-value	*df (Q)*	*P-value*	*I-squared*		*Hedge's g*	*Standard Error*	
45.078	*22*	*0.003*	*51.196*	*Pooled ES (9 studies, 23 effect sizes)*	*0.698*	*****	*0.138*

Problem-solving outcomes only

Heterogeneity

Q-value	*df (Q)*	*P-value*	*I-squared*		*Hedge's g*	*Standard Error*	
8.512	*7*	*0.290*	*17.765*	*Pooled ES (six studies, eight effect sizes)*	*0.518*	****	*0.169*

~ p < .10, * p < .05, ** p < .01, *** p < .001
[a] Data were adjusted for clustering that occurred within classrooms.

Note: The studies with mathematics outcomes are Clements (1986), Clements et al. (2001), Clements & Battista (1989), Emihovich & Miller (1988), Johnson-Gentile et al. (1994), Oprea (1988), Ortiz & MacGregor (1991), Thompson & Wang (1988), and Turner & Land (1988).

The studies with problem-solving outcomes are Battista & Clements (1986), Blume & Schoen (1988), Clements (1990) (not part of main pooled analysis; it was based on the same study or sample as Clements & Battista (1989), Degelman et al. (1986), Kapa (1999), and Lehrer & Randle (1987).

Based on the small number of studies in the subsequent categories, the Task Group did not conduct a meta-analysis of studies in these categories. The findings of prior syntheses and reviews are briefly presented to guide future research.

5. Tools: Computer—Existing Reviews

Studies in a broader and more ill-defined category of technology, software tools and exploratory environments (excluding calculators which were discussed above), were found to have inconsistent effects on student performance when compared to conventional classroom instruction in the synthesis of existing reviews. Detailed effect size information from prior quantitative syntheses and meta-analyses are presented in Table C-6 in Appendix C. One review reported that problem solving software appeared as effective as other categories of software (Edwards et al., 1975). However, pooled effect sizes reported in other meta-analyses have tended to be low, including 0.04 for tool and exploratory environments (Lou et al., 2001, who emphasize that commercial tests may underestimate effects), 0.10 for "computer-enhanced instruction" (a broad interpretation, Kulik & Kulik, 1991), and 0.24 for secondary students' use of problem-solving software (Kuchler, 1999).

In contrast to these limited effects, a recent randomized trials evaluation of a middle-school mathematics approach in which software is a key component reported a larger effect size (Roschelle et al., 2007). The approach focuses on proportionality, with software that connects different representational systems; for example, linking visual forms such as graphs and simulated motions to linguistic forms such as algebraic symbols and narrative stories of motion in an interactive and expressive context. The approach also embeds the software within a curriculum and includes professional development for the teachers, which may account for its success. Caveats include the short duration of the study (less than a month) and the participation of all volunteer teachers; for these reasons this study did not meet the Task Group's inclusion criteria.

Based on the small number of studies for any particular subcategory of tools and exploratory environments, the Task Group did not conduct a meta-analysis of this category except for one specific type of tool, the calculator, discussed previously.

6. Simulations and Games—Existing Reviews

Detailed effect size information from prior meta-analyses on simulation and games are presented in Table C-7 in Appendix C. The prior syntheses of the effects of simulation and game software revealed inconsistent effects on student performance when compared to conventional classroom instruction, with three previous meta-analyses providing a median pooled effect size of 0.23. All specific findings come from only one of these meta-analyses; thus, all results are tentative.

Table 29: What Prior Reviews Say About Simulations and Games

- General findings
  - Larger effects for computation or a combination of goals than for concepts and applications (but based on one meta-analysis, with small numbers of effect sizes; see Table C-7)
  - Arithmetic and “general” subjects showed higher effects than geometry and algebra
  - Attitudes toward mathematics and instruction positively affected by use of simulation software
- Contextual variables
  - Junior high students benefited more than elementary students
  - Simulations appear more effective for males
- Implementation variables
  - Effects were greater for studies of 1–18 weeks compared to those of 19–36 weeks duration
  - Higher effects of supplemental use on achievement than substitution.
  - Substitutions shows a negative effect on problem solving
  - No differences between experimenter or teacher-developed and commercial software
  - Higher gains in a context that combines guidance both with the subject matter content (e.g., other forms of instruction) and with students’ interaction with the simulation

7. Internet

There is no consistent empirical research base on the many types of learning and teaching that can be delivered or supported over the Internet. Two categories of software that appear to have tentative support, based on previous syntheses, are online learning and Web-based inquiry (e.g., Fadel & Lemke, 2006). Possible negative effects of using the Internet for mathematics and mathematics instruction also need to be researched. There were an insufficient number of original empirical studies to conduct an original meta-analysis on the use of the Internet in mathematics instruction.

8. Tools for Teachers

Such tools as electronic blackboards and quick-response devices have mostly descriptive studies to support them (e.g., Fadel & Lemke, 2006). The application of computer-managed instruction (CMI) has already been discussed as a component of ILSs. In addition, direct studies of CMI show a pooled effect size of 0.14 in a previous meta-analysis (Kulik, 1994). There were an insufficient number of original empirical studies to conduct an original meta-analysis on this topic.

D. Conclusions and Implications

This review summarized what is known about the role of technology, including different categories of computer software and calculators, in mathematics instruction and learning. Before reviewing the findings, several general issues are discussed.

Some reviewers have decried that the literature on educational technology is too inconsistent and uneven to make "sweeping conclusions about the effectiveness of instructional technology" (Kulik, 2003). Both a conceptual analysis and empirical review concur that any such sweeping conclusions are not warranted, but also suggest that such conclusions should not be sought as guides for educational practice. "Technology" is not a single, monolithic entity (Clements & Sarama, 2003). This review has shown different effects for different categories of software, has identified contextual and implementation variables, and whenever possible has distinguished between different applications of computer technology. However, the present research corpus is weak in distinguishing the effects of specific features of software categories and specific software applications (such as in Table 17; this major gap in research will be discussed in the succeeding section, "Instructional Software: Features and Pedagogical Strategies"). There are too few studies on documented implementations of specific strategies for educational technology, and even fewer studies on particular educational technology programs. Longitudinal studies are also needed.

Although some previous meta-analyses identified their effects (e.g., 0.19 to 0.24) as "weak," any such classification is dubitable, because the importance of any pooled effect size depends on a variety of factors (Lipsey & Wilson, 2001). For CBI, one particular issue is that students are often maximally engaged with the computer materials for 15–30 minutes two to three times per week. Pooled effect sizes must be interpreted in that context (Slavin & Lake, 2007).

Existing research, and the many available reviews of this body of research, suggests that specific categories and uses of educational technology can make a significant, positive contribution to students' learning of mathematics. The Task Group conducted its own meta-analyses to evaluate those conclusions of previous reviews.

1. Drill and Practice

Prior syntheses and meta-analyses suggest that CAI drill and practice generally improves students' performance compared to conventional instruction, with the greatest effects on computation, and more limited on concepts and applications. It is the only category of instructional software that shows, in previous reviews, higher effects for serving as a substitute for conventional instruction, rather than as a supplement to it. It may be that such programs address students' instructional needs for practice adequately and efficiently, making substantial teacher intervention less important.

The Task Group's meta-analysis of rigorous studies supports these conclusions. Drill and practice software had a significant positive effect on mathematics achievement. When analyzed for different ages and grades, positive effects were confirmed for the elementary level, but there were too few studies at other levels to make comparisons or conclusions. Effect sizes were higher for interventions that substituted for, rather than supplemented, classroom practice.

In summary, drill and practice through high-quality CAI, implemented with fidelity, can be considered a useful tool in developing students' automaticity, or fast, accurate, and effortless performance on computation, freeing working memory so that attention can be directed to the more complicated aspects of complex tasks. A caveat is that older studies may have used software better designed to use research-based strategies (and fewer "bells and whistles," graphics and sound not related to instruction) than many more recently published programs. Using such strategies to incorporate features such as those in Table 17 will likely maximize positive effects. The following section includes additional caveats relevant to drill and practice.

2. Tutorials

Prior syntheses and meta-analyses suggest that CAI tutorials improve students' performance compared to conventional instruction, with slightly greater effects on performance on concepts and applications measures than on computation measures. Based on these prior syntheses and meta-analyses, tutorials appear to be effective at all grade levels, particularly the higher grades. Reviews indicate that they are more effective when they supplement, rather than replace, conventional instruction, when they involve experimenter or teacher-developed, rather than commercially developed, software, and when they are developed for a specific audience rather than a general audience.

The Task Group's meta-analysis of rigorous studies similarly indicates that tutorials can increase mathematics performance, both overall achievement and, possibly more so, mathematical problem-solving ability. It supported the conclusion that tutorials are more effective as supplements, rather than replacements, for conventional instruction and when they are highly integrated with the regular mathematics curriculum. Finally, tutorial software developed by researchers or teachers was more effective than that developed by commercial companies. Findings of individual studies provide serious caveats, however, including the need to consider empirical evidence of effectiveness of a particular software package, and issues of scale-up, including integration with the extant curriculum, and fidelity of implementation, including amount of use, and technological and pedagogical support.

In summary, tutorials, as well as software packages that combine tutorials with drill and practice, that are well designed (e.g., including features in Table 17; see also Clements, 2007; Clements & Battista, 2000) and implemented can be considered as potentially useful tools in introducing and teaching specific subject-matter content to specific populations, especially at the junior and senior high school levels. Research suggests that tutorials be designed to develop specific educational goals for specific populations. Caveats are that results are not guaranteed, and care must be taken that there is evidence that the software increases learning and that the requisite support conditions to use the software effectively are in place.

3. Tools: Calculators and Graphing Calculators

Prior syntheses and meta-analyses suggest that calculators of all types, basic, scientific and graphing, may benefit students' achievement in (and attitudes toward) mathematics, and effects are more positive when calculators are used during testing. Previous reviews also indicate that effects of calculator use on calculation, contrary to perhaps the most common concern, are near zero but positive (even when calculators are not allowed on the assessments), and effects on problem solving were positive.

The Task Group's meta-analyses of 11 studies that met the Panel's rigorous criteria (only one study less than 20 years old) found limited to no impact of calculators on calculation skills, problem-solving, or conceptual development. Effect sizes of these studies are lower than those in prior syntheses and meta-analyses. On the basis of the high quality studies identified in this category by the Task Group, it is reasonable to conclude that there is no significant negative impact of calculators on students' calculation competence (only one of the studies allowed students to use calculators on the assessment). However, there are several important caveats. These findings are limited to the effect of calculators as used in the 11 studies, including studies up to a year in duration. Also, tests of computational skills did not measure the more basic processes, such as retrieval or decomposition, that students use to solve arithmetic problems, nor did they measure automaticity or procedural execution as might be assessed with timed paper-and-pencil tests (see the Learning Processes Task Group report). This is especially important when arithmetic skills are being formed, because inappropriate calculator use may interfere with the development of these skills. On the other hand, it is possible that appropriate calculator use could provide useful feedback and build a stronger association between addends and their sum, strengthening these associations. Especially given these conflicting possibilities, and the importance of this early development, the lack of rigorous studies with students earlier than third grade is especially unfortunate. Also, research on calculator use over several years—especially comparing inappropriate and appropriate use—is direly needed.

Further, given that the basic computational skills of many Americans are poor, as described in the Learning Processes report, a finding of no effect is not a promising one; more powerful instructional approaches are needed. The synthesis of previous reviews suggests that more recent calculator interventions, especially those putting calculators to "pedagogical use" as an essential element in the teaching and learning of mathematics, have a greater positive effect (the studies in the Task Group's meta-analysis did not report such comparisons). "Pedagogical use" usually implies extending mathematics learning in certain situations (and perhaps using calculators to check the accuracy of mental or other calculations), rather than using calculators when other methods would be appropriate. The overuse and inappropriate use of calculators, decried by many, may be more harmful than these (relatively short-term) studies indicate. On the other hand, an emphasis on mental arithmetic may ameliorate such problems. There is much researchers still need to study.

This report has not addressed several important educational issues. There is a dearth of research not only on broad categories of calculator use such as "functional vs. pedagogical" use, but on specific uses of calculators that may lead to negative effects (e.g., overdependence), null effects, or specific positive effects. Such research should also fill the gap in the literature if studies included *observation* of how, how much, and how well, calculators are used (this includes, for planned interventions, "fidelity of implementation" measures).

In a similar vein, the older studies in the Task Group's meta-analysis, and more recent calculator studies, classify measures and findings by broad categories only, such as "calculation," "problem solving," and "concepts." Greater specificity in terms of grade level and topic, instructional goals, and pedagogical strategy would yield more useful research results and implications. Would specifically targeted use, in which the calculator's unique characteristics are used intentionally, result in greater benefits? For example, one might only introduce calculators in work with arithmetic with numbers of six or more digits, square roots, or scientific notion. Another project might introduce calculators in earlier grades, not to replace computational practice (mental and paper-and-pencil arithmetic) but rather to extend computational and problem-solving proficiency.

Even more fundamental, although some may argue against calculator use because it circumvents the mathematics they wish students to perform, others believe that in an age of calculators and computers, it is inappropriate to continue to focus the elementary school mathematics curriculum on pencil-and-paper arithmetic (Ralston, 1999). Research cannot address such curriculum issues of goals and values, although, it should be explicit about its assumptions. Research can clarify the ramifications of various approaches, but the work of discussing these approaches, and evaluating them through empirical research, largely remains to be done.

In summary, most of the effects in the Task Group's meta-analysis have a similar pattern of results to those in the prior syntheses and meta-analyses, but with smaller, and usually near-zero, statistically insignificant effect sizes. Given the design flaws noted in some studies included in previous meta-analyses, this may indicate that the smaller effect sizes represent more accurate estimates of calculators' effects. However, there are different, but still substantial, limitations to the pool of studies that met the criteria for inclusion in the present meta-analyses. First, only 11 studies of the hundreds in the literature are included in the Task Group's meta-analysis. Only one was published after 1987, and that had but one comparison, at Grades 9 and 10. This could be important, as a previous meta-analysis indicated that effects of calculators may be becoming more positive with time (Ellington, 2003), which may suggest that technology, and especially support materials and professional development related to technology use, are improving since the introduction of calculators. Recent calculator interventions that use research-based approaches (e.g., embedding technology within a curriculum and targeting calculator use to particular pedagogical ends) to incorporate newer technologies provide suggestive results (Stroup, Pham, & Alexander, 2007). Second, most of the studies included in the Task Group's meta-analysis measured computational skills, but only half assessed the learning of problem solving or concepts. Thus, many of the comparisons for a specific effect are from a small number of studies (with effects pooled from multiple comparisons frequently originating from a single study). Given the different limitations of each report, conclusions that they share appear trustworthy—that is, calculators as used in these studies have little or no effect on most measured outcomes in calculation or concepts, given the manner in which calculators were used and the duration of the studies—but, especially when outcomes differ, a larger body of more recent, rigorous studies that documents *how* calculators are used, including research that examines multiyear use of calculators, is needed before firm conclusions can be reached.

4. Computer Programming

Computer programming by students can be employed in a wide variety of situations using distinct pedagogies. Prior syntheses and meta-analyses indicate that programming improves students' performance compared to conventional instruction, with the greatest effects on concepts and applications, especially geometric concepts, and weaker effects on computation. Previous reviews also indicate that programming positively affects problem solving, as well as attitudes toward mathematics and instruction in mathematics, more so than other software categories. Prior reviews also provide some evidence that use of computer programming is more effective for students of average, rather than low or high SES. Earlier reviews have claimed that programming is somewhat more effective when it supplements, rather than replaces, conventional instruction, consistent with suggestions for mediated instruction of programming. Certain computer languages, especially the Logo computer language, were reported to have stronger positive effects than other computer languages.

The Task Group's meta-analysis of rigorous studies supports the conclusions of previous reviews about the impact of computer programming on mathematics performance. Further, the meta-analysis suggested that results for delayed posttests might be greater than those for immediate posttests. Additional research is needed to ascertain whether this finding is generalizable.

In summary, computer programming can be considered an effective tool, especially for elementary school students, for developing specific mathematics concepts and applications and mathematical problem-solving abilities. Effects may be larger the more computer programming is integrated into the curriculum. Although there was insufficient research on such issues, the Task Group notes that instructional use of programming has fewer "bells and whistles" than other categories of software and demands thoughtful curricula and knowledgeable teachers, all of which may have contributed to the lack of frequency in U.S. classrooms (it is more widely used in other countries, Clements & Sarama, 1997). Dissemination of research, including research-based curricula and professional development, could lead to a reversal of this trend.

5. Tools: Computer Tools

Software tools and exploratory environments (excluding calculators) have inconsistent effects on student performance. Prior syntheses and meta-analyses suggest that problem-solving software may have potential, but effect sizes have been small. Recent rigorous studies suggest that new approaches may have promise, but there are an inadequate number of such studies for the Task Group to conduct a meta-analysis of this software category.

6. Simulations and Games

Prior syntheses and meta-analyses suggest that simulation and game software packages may have positive, but relatively small, effects on student performance when compared to conventional classroom instruction. Previous studies also have shown them to have a positive effect on attitudes. Junior high, more than elementary, students may benefit from working with simulations and games. Supplemental use is indicated, consistent with the intrinsically unguided nature of simulations and games.

In summary, there is only slight evidence—based on studies of unknown rigor—indicating that simulations may be useful, especially at the middle or junior high level, to develop skills, concepts, and applications of knowledge in problem-solving settings. More needs to be known about developing and using this category of software, but it is likely that careful integration into a well-structured curriculum is critical to facilitate learning.

7. Instructional Software: Features and Pedagogical Strategies

Many questions essential to designing and selecting educational technology applications cannot be answered, because studies and reviews do not distinguish such applications on their use of specific features. Similar situations exist for practice, the role of the teacher (especially specific pedagogical strategies).

a. Software Features

The Task Group's reviews found that the previous meta-analyses and rigorous studies did not permit generalizations about critical features of software, such as those identified in Table 17. That is, prior syntheses and meta-analyses do not sufficiently distinguish such applications on their use of specific features that theoretically should contribute to learning. Such findings would be invaluable to the field, both because decisions could be guided by any software program's inclusion of critical features and because the development of new software programs could be similarly guided.

Only for the sake of illustration, a few studies that did not meet the Task Group's criteria are described here that compare CAI conditions, most of which had varied conclusions. One study reported that enhancing drill by placing it within a game context does not yield significantly different outcomes overall, but the game may distract students with learning disabilities (Christensen & Gerber, 1990). Enhancement with multimedia significantly improved learning in one study (Macaulay, 2003) but CAI with animated vs. static pictures or with or without the presentation of a cognitive strategy were equally effective (Shiah, Mastropieru, Scruggs, & Mushinski Fulk, 1994). Verbal guidance (in students' first language) may support learning from multimedia educational games (Moreno & Duran, 2004). These are single studies with little conceptual overlap; the field needs more complete and reliable guidance.

Few software programs are designed based on explicit (i.e., published) theoretical and empirical research foundations (but see Clements, 2007; Clements & Sarama, 2007a; Ritter, Anderson, Koedinger, & Corbett, 2007). More continuous, committed, iterative research and development projects are needed in this area. Research-based iterative cycles of evaluation and development, fine tuning software's mathematics and pedagogy within each cycle, can make substantial differences in learning (e.g., see Aleven & Koedinger, 2002; Clements & Battista, 2000; Clements et al., 2001; Laurillard & Taylor, 1994; Steffe & Olive, 2002).

Such research could identify how and why software designs could be improved. As one example, the pooled effect sizes in the Task Group's meta-analysis actually might be an underestimate of what can be achieved if drill and practice software were more carefully designed. Few studies use empirically validated strategies such as adaptive feedback and increasing ratio review (Siegel & Misselt, 1984).

8. Final Words

In most cases, specific uses of technology will not facilitate learning optimally unless they are implemented with fidelity. Unfortunately, information is lacking on this critical issue because reviewers and researchers generally have not measured fidelity. A similar situation exists for many specific pedagogical issues.

In addition, from the subtleties of designing features of software, to the complexities of scaling up approaches to work with entire educational systems, substantive challenges face researchers and other educators. These challenges must be met, and findings integrated across levels, before conclusions about the effectiveness of educational technology can be offered with confidence. Many difficulties stand in the way of conducting high-quality work in the field of technology in mathematics education. Applications that go beyond using the simplest features of technology to deliver a traditional curriculum face both (a) challenges of redesigning scope and sequences, pedagogies, software, and assessments (Kulik & Kulik, 1991), along with financial and logistical hurdles, and (b) barriers of a priori negative evaluation of the goals and assessment instruments they may wish to employ. Such barriers may have dampened innovative research and development in educational technology. This is unacceptable; concerted efforts are needed to meet these challenges and provide educators with clear guidelines from research. This is especially important given the poor implementation of educational technology in the field (Clements & Sarama, 1997; Cuban, 2001; Hoffer, Venkataraman, Hedberg, & Shagle, 2007).

Finally, technological advances continue to challenge practitioners and researchers. There is no research on questions that have arisen only recently. What technologies are most appropriate for students for whom multiple hand-held devices are a ubiquitous presence? How has the presence of Internet sites affected students (e.g., mathematics as presented on Wikipedia)? Both new questions and old must be better addressed with high-quality studies of high-quality implementations of computer-based tools if educational technology is to fulfill its potential.

BIBLIOGRAPHY

Aleven, V.A.W.M.M., & Koedinger, K.R. (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer-based cognitive tutor. *Cognitive Science, 26(2),* 147–179.

Ball, S. (1988). Computers, concrete materials and teaching fractions. *School Science and Mathematics, 88(6),* 470–475

Battista, M.T., & Clements, D.H. (1986). The effects of Logo and CAI problem-solving environments on problem-solving abilities and mathematics achievement. *Computers in Human Behavior, 2*, 183–193.

Becker, H.J. (1992). Computer-based integrated learning systems in the elementary and middle grades: A critical review and synthesis of evaluation reports. *Journal of Educational Computing Research, 8*, 1–41.

Blume, G.W., & Schoen, H.L. (1988). Mathematical problem-solving performance of eighth-grade programmers and nonprogrammers. *Journal for Research in Mathematics Education, 19(2),* 142–156.

Burns, P.K., & Bozeman, W.C. (1981). Computer-assisted instruction and mathematics achievement: Is there a relationship? *Educational Technology, 21(10),* 32–39.

Campbell, D.L., Peck, D.L., Horn, C.J., & Leigh, R.K. (1987). Comparison of computer-assisted instruction and print performance: a research note. *Educational Communication and Technology Journal, 35(2),* 95–103.

Campbell, P., & Virgin, A.E. (1976). *An evaluation of elementary school mathematics programs utilizing the mini-calculator*. Toronto, Canada: Ontario Department of Education.

Carmichael, H.W., Burnett, J.D., Higginson, W.C., Moore, B.G., & Pollard, P.J. (1985). *Computers, children and classrooms: A multisite evaluation of the creative use of microcomputers by elementary school children.* Toronto, Ontario, Canada: Ministry of Education.

Carrier, C., Post, T.R., & Heck, W. (1985). Using microcomputers with fourth-grade students to reinforce arithmetic skills. *Journal for Research in Mathematics Education, 16(1),* 45–51.

Chambers, E.A. (2002). *Efficacy of educational technology in elementary and secondary classrooms: A meta-analysis of the research literature from 1992–2002.* Southern Illinois University Carbondale.

Christensen, C.A., & Gerber, M.M. (1990). Effectiveness of computerized drill and practice games in teaching basic mathematics facts. *Exceptionality, 1(3),* 149–165.

Christmann, E., Badgett, J., & Lucking, R. (1997). Progressive comparison of the effects of computer-assisted instruction on the academic achievement of secondary students. *Journal of Research on Computing in Education, 29(4),* 325–337.

Clements, D.H. (1986). Effects of Logo and CAI environments on cognition and creativity. *Journal of Educational Psychology, 78(4),* 309–318.

Clements, D.H. (1990). Metacomponential development in a Logo programming environment. *Journal of Educational Psychology, 82(1),* 141–149.

Clements, D.H. (2007). Curriculum research: Toward a framework for 'research-based curricula'. *Journal for Research in Mathematics Education, 38(1),* 35–70.

Clements, D.H., & Battista, M.T. (1989). Learning of geometric concepts in a Logo environment. *Journal for Research in Mathematics Education, 20(5),* 450–467.

Clements, D.H., & Battista, M.T. (2000). Designing effective software. In A.E. Kelly & R.A. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp. 761–776). Mahwah, NJ: Lawrence Erlbaum Associates.

Clements, D.H., Battista, M.T., & Sarama, J. (2001). Logo and geometry. In *Journal for Research in Mathematics Education Monograph Series, 10.* Reston, VA: National Council of Teachers of Mathematics.

Clements, D.H., & Sarama, J. (1997). Research on Logo: A decade of progress. *Computers in the Schools, 14(1–2),* 9–46.

Clements, D.H., & Sarama, J. (2003). Strip mining for gold: Research and policy in educational technology—A response to "Fool's Gold". *Educational Technology Review, 11(1),* 7–69.

Clements, D.H., & Sarama, J. (2007). Effects of a preschool mathematics curriculum: Summative research on the *Building Blocks* project. *Journal for Research in Mathematics Education, 38(2),* 136–163.

Clements, D.H., & Sarama, J. (2008). Experimental evaluation of the effects of a research-based preschool mathematics curriculum. *American Educational Research Journal* (forthcoming 2008).

Cuban, L. (2001). *Oversold and underused*. Cambridge, MA: Harvard University Press.

Dalton, D., & Hannafin, M.J. (1988), The effects of computer-assisted and traditional mastery methods on computation accuracy and attitudes. *Journal of Educational Research, 82(1),* 27–33.

Degelman, D., Free, J.U., Scarlato, M., & Blackburn, J.M. (1986). Concept learning in preschool children: Effects of short-term logo experience. *Journal of Educational Computing Research, 2(2),* 199–205.

Duffy, J.A., & Thompson, G.E. (1980). *Improving mathematical achievement of elementary students through the use of electronic hand calculators*. Columbus, OH: Columbus Public Schools, Ohio Department of Management, Planning and Information.

Dynarski, M., Agodini, R., Heaviside, S., Novak, T., Carey, N., Campuzano, L., et al. (2007). *Effectiveness of reading and mathematics software products: Findings from the first student cohort*. Washington, DC: U.S. Department of Education, Institute of Education Sciences.

Edwards, J., Norton, S., Taylor, S., Weiss, M., & Dusseldorp, R. (1975). How effective is CAI? A review of the research. *Educational Leadership, 33(2),* 147–153.

Ellington, A.J. (2003). A meta-analysis of the effects of calculators on students' achievement and attitude levels in precollege mathematics classes. *Journal for Research in Mathematics Education, 34(5),* 433–463.

Ellington, A.J. (2006). The effects of non-CAS graphing calculators on student achievement and attitude levels in mathematics: A meta-analysis. *School Science and Mathematics, 106(7)*, 16–26.

Emihovich, C., & Miller, G.E. (1988). Effects of Logo and CAI on Black first graders' achievement, reflectivity, and self-esteem. *The Elementary School Journal, 88(5),* 472–487.

Fadel, C., & Lemke, C. (2006). *Technology in schools: What the research says*. San Jose, CA: Metiri Group, commissioned by Cisco Systems.

Fletcher, J.D., Hawley, D.E., & Piele, P.K. (1990). Costs, effects, and utility of microcomputer-assisted instruction in the classroom. *American Educational Research 27(4)*, 783–806.

Fletcher-Flinn, C.M., & Gravatt, B. (1995). The efficacy of computer-assisted instruction (CAI): A meta-analysis. *Journal of Educational Computing Research, 12(3),* 219–242.

Fuchs, L.S., Fuchs, D., Hamlett, C.L., & Appleton, A.C. (2002). Explicitly teaching for transfer: Effects on the mathematical problem-solving performance of students with mathematics disabilities. *Learning Disabilities Research & Practice, 17(2),* 90–106.

Fuchs, L.S., Fuchs, D., Hamlet, C.L., Powell, S.R., Capizzi, A.M., & Seethaler, P.M. (2006). The effects of computer-assisted instruction on number combination skill in at-risk first graders. *Journal of Learning Disabilities, 39(5)*, 467–475.

Gordon, B.W. (1992). *A quantitative analysis of the relationship between computer graphics and mathematics achievement and problem solving.* Unpublished doctoral dissertation, University of Cincinnati, Ohio.

Graham, A.T., & Thomas, M.O. (2000). Building a versatile understanding of algebraic variables with a graphic calculator. *Educational Studies in Mathematics, 41(3),* 265–282.

Hamilton, W.A. (1995). A meta-analysis of the comparative research on computer-assisted instruction and its effects on elementary and secondary mathematics achievement. *Dissertation Abstracts International, 56(05)*, 1599.

Hartley, S.S. (1978). Meta-analysis of the effects of individually paced instruction in mathematics. *Dissertation Abstracts International, 38*, 4003A. UMI No. 77-29, 926.

Hembree, R. (1984). *Model for meta-analysis of research in education with a demonstration of mathematics on education: Effects of handheld calculators.* Unpublished doctoral dissertation, University of Tennessee, Knoxville.

Hembree, R. (1992). Experiments and relational studies in problem solving: A meta-analysis. *Journal for Research in Mathematics Education, 23(3),* 242–273.

Hembree, R., & Dessart, D.J. (1986). Effects of hand-held calculators in precollege mathematics education: A meta-analysis. *Journal for Research in Mathematics Education, 17(2),* 83–99.

Henderson, R.W., Landesman, E.M., & Kachuck, I. (1985). Computer-video instruction in mathematics: Field test of an interactive approach. *Journal for Research in Mathematics Education, 16(3),* 207–224.

Hoffer, T.B., Venkataraman, L, Hedberg, E.C., & Shagle, S. (2007). *Final report of the national survey of algebra teachers for the National Math Panel*. Chicago: University of Chicago.

Johnson-Gentile, K., Clements, D.H., & Battista, M.T. (1994). The effects of computer and noncomputer environments on students' conceptualizations of geometric motions. *Journal of Educational Computing Research, 11(2)*, 121–140.

Kapa, E. (1999). Problem solving, planning ability and sharing processes with LOGO. *Journal of Computer Assisted Learning, 15(1)*, 73–84.

Khalili, A., & Shashaani, L. (1994). The effectiveness of computer applications: A meta-analysis. *Journal of Research on Computing in Education, 27(1),* 48–61.

Khoju, M., Jaciw, A., & Miller, G.I. (2005). *Effectiveness of graphing calculators in K–12 mathematics achievement: A systematic review*. Palo Alto, CA: Empirical Education Research Report.

Kraus, W.H. (1981). Using a computer game to reinforce skills in addition basic facts in second grade. *Journal for Research in Mathematics Education, 12(2),* 152–155.

Kuchler, J.M. (1999). *The effectiveness of using computers to teach secondary school mathematics.* Unpublished doctoral dissertation. University of Massachusetts, Lowell.

Kulik, C.-L.C., & Kulik, J.A. (1991). Effectiveness of computer-based instruction: An updated analysis. *Journal of Computer-Based Instruction, 7(1–2)*, 75–94.

Kulik, J.A. (1994). Meta-analytic studies of findings on computer-based instruction. In E.L. Baker, & H.F. O'Neil, Jr. (Eds.), *Technology assessment in education and training* (pp. 9–33). Hillsdale, NJ: Lawrence Erlbaum Associates.

Kulik, J.A. (2003). *Effects of using instructional technology in elementary and secondary schools: What controlled evaluation studies say.* Arlington, VA: SRI International.

Laurillard, D., & Taylor, J. (1994). Designing the stepping stones: An evaluation of interactive media in the classroom. *Journal of Educational Television, 20(3),* 169–184.

Lee, W.C. (1990). The effectiveness of computer-assisted instruction and computer programming in elementary and secondary mathematics: A meta-analysis. (Doctoral dissertation, University of Amherst, 1990). *Dissertation Abstracts International, 51(03)*, 775.

Lehrer, R., & Randle, L. (1987). Problem solving, meta-cognition and composition: The effects of interactive software for first-grade children. *Journal of Educational Computing Research, 3(4)*, 409–427.

Lipsey, M.W., & Wilson, D.B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage Publications.

Lou, Y., Abrami, P.C., & d'Apollonia, S. (2001). Small group and individual learning with technology: A meta-analysis. *Review of Educational Research, 71(3)*, 449–521.

Macaulay, M. (2003). The effects of multimedia on learning in third world children. *Journal of Educational Multimedia and Hypermedia, 12 (2),* 185–198.

McCollister, T.S., Burts, D.C., Wright, V.L., & Hildreth, G.J. (1986). Effects of computer-assisted instruction and teaching instruction on arithmetic task achievement scores of kindergarten children. *Journal of Educational Research, 80(2),* 121–125.

Moore, B.M. (1988). Achievement in basic mathematics skills for low performing students—A study of teachers' affect and CAI. *Journal of Experimental Education, 57(1),* 38–44.

Moreno, R., & Duran, R. (2004). Do multiple representations need explanations? The role of verbal guidance and individual differences in multimedia mathematics learning. *Journal of Educational Psychology, 96*(3), 492–503.

National Council of Teachers of Mathematics (NCTM). (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.

Niemiec, R.P., & Walberg, H.J. (1984). Computers and achievement in the elementary schools. *Journal of Educational Computing Research, 1(4),* 435–440.

Oprea, J.M. (1988). Computer programming and mathematical thinking. *Journal of Mathematical Behavior, 7(2),* 175–190.

Ortiz, E., & MacGregor, S.K. (1991). Effects of LOGO programming on understanding variables. *Journal of Educational Computing Research, 7,* 37–50.

Papert, S., Watt, D., di Sessa, A., & Weir, S. (1979). *Final report of the Brookline Logo Project: Parts 1 and 11* (Logo Memos Nos. 53 and 54). Cambridge, MA: Massachusetts Institute of Technology Artificial Intelligence Laboratory.

Podell, D., Tournakirein, N., & Lin, A. (1992). Automatization of mathematics skills via computer-assisted-instruction among students with mild mental-handicaps. *Education & Training in Mental Retardation & Developmental Disabilities, 27(3),* 200–206.

Ralston, A. (1999). Let's abolish pencil-and-paper arithmetic. *Journal of Computers in Mathematics and Science Teaching, 18,* 173–194.

Ritter, S., Anderson, J.R., Koedinger, K.R., & Corbett, A. (2007). Cognitive Tutor: Applied research in mathematics education. *Psychonomics Bulletin & Review, 14(2),* 249–255.

Roschelle, J.M., Tatar, D., Shechtman, N., Hegedus, S., Hopkins, B., Knudsen, J., et al. (2007). *Can a technology enhanced curriculum improve student learning of important mathematics?* (Vol. Technical Report 01). Menlo Park, CA: SRI International.

Ryan, A.W. (1991). Meta-analysis of achievement effects of microcomputer applications in elementary schools. *Educational Administration Quarterly, 27(2),* 161–184.

Saracho, O.N. (1982). The effects of a computer-assisted instruction program on basic skills achievement and attitudes toward instruction of Spanish-speaking migrant children. *American Educational Research Journal, 19(2),* 201–219.

Saunders, J., & Bell, F.H. (1980). Computer-enhanced algebra resources: Their effects on achievement and attitudes. *International Journal of Mathematical Education in Science and Technology, 11(4),* 465–473.

Schnur, J.O., & Lang, J.W. (1976). Just pushing buttons or learning?—A case for mini-calculators. *The Arithmetic Teacher, 23(7),* 559–562.

Shiah, R.L., Mastropieri, M.A., Scruggs, T.E., & Mushinski-Fulk, B.J. (1994). The effects of computer-assisted instruction on the mathematical problem solving of students with learning disabilities. *Exceptionality,* 5(3), 131–161.

Siegel, M.A., & Misselt, A.L. (1984). Adaptive feedback and review paradigm for computer-based drills. *Journal of Educational Psychology, 76(2),* 310–317.

Slavin, R.E., & Lake, C. (2007). *Effective programs in elementary mathematics: A best-evidence synthesis*. Baltimore, MD: Johns Hopkins University.

Smith, B.A. (1997). *A meta-analysis of outcomes from the use of calculators in mathematics education.* Unpublished doctoral dissertation, Texas A&M University.

Standifer, C.E., & Maples, E.G. (1981). Achievement and attitude of third-grade students using two types of calculators. *School Science and Mathematics, 81(1),* 17–24.

Standifer, C.E., & Maples, E.G. (1982). Mathematical achievement of third- and fourth-grade students in compensatory education when using two types of calculators. *Psychological Reports*, 82(51), 11–18.

Steffe, L.P., & Olive, J. (2002). Design and use of computer tools for interactive mathematical activity (TIMA). *Journal of Educational Computing Research, 27(1&2),* 55–76.

Stroup, W., Pham, V., & Alexander, C. (2007). *Richardson MathForward Project: Second Year Final Report: Mathematics TAKS Results.* Austin, TX: University of Texas at Austin.

Szetela, W. (1980). Calculators and the teaching of ratios in Grade 7. *Journal for Research in Mathematics Education, 11(1),* 67–70.

Szetela, W. (1982). Story problem solving in elementary school mathematics: What differences do calculators make? *Journal for Research in Mathematics Education, 13(5),* 381–389.

Szetela, W., & Super, D. (1987). Calculators and instruction in problem solving in Grade 7. *Journal for Research in Mathematics Education, 18(3),* 215–229.

Thomas, P.G., & Rickhuss, M.G. (1992). An experiment in the use of computer algebra in the classroom. *Education and Computing, 8(3),* 225–263.

Thompson, A.D., & Chen-Wang, H. (1988). Effects of a Logo microworld on student ability to transfer a concept. *Journal of Educational Computing Research, 4(3),* 335–347.

Trifiletti, J.J., Frith, G.H., & Armstrong, S.W. (1984). Microcomputers versus resource rooms for LD students: A preliminary investigation of the effects on mathematics skills. *Learning Disability Quarterly, 7(1),* 69–76.

Turner, S.V., & Land, M.L. (1988). Cognitive effects of a Logo-enriched mathematics program for middle school students. *Journal of Educational Computing Research, 4(4),* 443–452.

Watkins, M.W. (1986). Microcomputer-based mathematics instruction with first-grade students. *Computers in Human Behavior,* 2, 71–75.

Wheatley, C.L. (1980). Calculator use and problem-solving performance. *Journal for Research in Mathematics Education, 11(5),* 323–334.

Wheatley, G.H., & Shumway, R.J. (1979). *Impact of calculators in elementary school mathematics: final report*. Washington, DC: National Science Foundation.

Wheeler, J.L., & Regian, JW (1999). The use of a cognitive tutoring system in the improvement of the abstract reasoning component of word problem solving. *Computers in Human Behavior*, *15(2),* 243–254.

Wilson, W.S., & Naiman, D.Q. (2004). K–12 calculator usage and college grades. *Educational Studies in Mathematics, 56(1),* 119–122.

VI. Instructional Practices and Mathematics Achievement: The Case of the Gifted Student

Students arrive at school with different skills and knowledge levels as well as capacities for benefiting from the opportunities provided by schools; these differences remain throughout schooling (Benbow & Stanley, 1996). This conclusion has been documented widely in the literature, going back as far as Learned and Wood (1938). Learned and Wood were among the first to show the wide range in knowledge among students in the same grade. For example, approximately 10% of high school seniors had more scientific knowledge than the average college senior. Such individual differences in knowledge and skills are evident even before entry into kindergarten, are reflected by the variance of test scores, and persist in every grade thereafter (Paterson, 1957; Pressley, 1949; Seashore, 1922; Terman, 1954; Tyler, 1965; Willerman, 1979; also see Learning Processes Task Group report). Moreover, there are differences in rate of learning. Those 13-year-olds who are in the top 1% of ability, for example, can assimilate, in three intensive weeks of schooling, a full year of high school biology, chemistry, Latin, physics, or mathematics (e.g., Lynch, 1992; Stanley & Stanley, 1986; VanTassel-Baska, 1983). Those who are in the top 1 in 10,000 in ability can accomplish even more in this time frame. Moreover, highly mathematically able students, with their exceptionally strong short-term working memory (Dark & Benbow, 1990, 1991, 1994), enjoy abstract, unstructured problems and thrive with complexity, which is different from the learning environment that is typical in the “regular” classroom. At the other end are students who need intensive work and much structured support and scaffolding over a long period of time to master basic skills in reading and mathematics. A challenge in teaching, then, is to be responsive to these individual differences so that all students make progress and are allowed to achieve their potential (National Research Council, 2000, 2002; Stanley, 2000). Particularly challenging for teachers are those students who are advanced or so challenged that the typical age-grade curriculum becomes inappropriate. In the case of the advanced student, serious adjustment is required if to teach them only what they already do not know (Stanley, 2000). In this report, the Task Group begins by briefly describing the strategies that are typically used to meet the learning needs of the advanced learner, often labeled the gifted student, and then move on to assess their effectiveness.

In American schools there are a plethora of programs that have been developed to meet the needs of gifted students. They represent the varied results obtained when the four, theoretically derived principles for adjusting the educational experiences or, more precisely, differentiating the curriculum are employed. The curriculum can be differentiated by level (e.g., grade level), complexity (e.g., abstract, unstructured), breadth or depth, and pacing to meet the learning needs of gifted students and ensure developmental appropriateness, according to an extensive literature in gifted education (Kaplan, 1986; Renzulli, 1986; VanTassel-Baska, 1998; Olszewski-Kubilus, 2007). Depending upon the relative emphasis of each one of these principles and the social context, the resulting programs fall into four broad categories: enrichment, acceleration, homogeneous grouping, and individualization. Enrichment often is seen in the regular classroom or in pullout programs or supplemental classes. It represents attempts to make the curriculum more appropriate for gifted students by adding to it or providing more depth and complexity while keeping students with their same-age peers. Acceleration and homogeneous grouping are attempts at forming groups for instruction that are at the same approximate achievement level, either by moving the advanced student to a higher grade in a (or many) subject(s) or by forming groups of same-age students on the basis of their demonstrated achievement. Indirectly, complexity is enhanced. Of course, all of these options can be used in some combination and that is what textbooks and articles in gifted education suggest (VanTassel-Baska, 1998). As well, the amount of adjustment required depends upon the level of giftedness and the difference between the individual gifted student and the average of the class. Acceleration that involves grade-skipping or putting individual students in a higher grade for a specific subject, for example, is typically reserved for the highly gifted (e.g., top 1% or even more extreme ability levels) as students much below that level often do not require such extreme adjustments.

A big debate in gifted education has been between the use of enrichment and acceleration. Most view this as a false dichotomy. For the highly gifted especially, it is recommended that both be utilized (VanTassel-Baska, 1998; Olszewski-Kubilus, 2007). There is, however, great resistance in K–12 schools toward using acceleration, even with the highly gifted (Benbow, 1991; Colangelo et al., 2004). This is not the case at the collegiate level when course placement is dependent upon having met prerequisites or scores on placement exams. The resistance by K–12 educators and fears of parents in terms of social and emotional development, however, have served to stimulate much research, admittedly of varying quality, to assess acceleration’s effectiveness and whether it actually produces harm. Thus, there is an imbalance in the existing research literature in gifted education, with most of the research focused on accelerative strategies. Moreover, acceleration itself is a profound thing to do as it puts usual intellectual and social trajectories out of synchrony and often involves just one or possibly just a few students in a given school. This means that special considerations and individualization are required to make it possible and ensure its success. For example, special efforts are made to place the to-be-accelerated child with a teacher supportive of the acceleration if at all possible (VanTassel-Baska, 1998), given the frequent hostility toward such students and any interventions provided (Benbow & Stanley, 1996; Coleman, 1960; Cramond & Martin, 1987; Hofstadter, 1963; Tannenbaum, 1962). This, coupled with other issues (e.g., the child wanting to accelerate—motivation as a criterion)—makes it challenging to conduct carefully controlled research.

Previous meta-analyses have tried to make sense of this literature with all of its limitations and the varying quality of studies. They identified acceleration as the most promising strategy, followed by homogeneous grouping involving differentiation of the curriculum and adjustment of methods of teaching (Kulik & Kulik, 1982, 1984, 1992; Rogers, 2007; Olszewski-Kubilus, 2007). This represents what the field of gifted education thought the state of knowledge was before the Instructional Practices Task Group began its work.

As discussed in the introduction of this report and in Appendix A, the Instructional Practices Task Group developed criteria for which studies it would consult as part of its deliberations. The Task Group’s charge was to assess the effects of instructional practices on mathematics achievement and establish a warranted claim of causality. It posed the following question: Can the Task Group conclude, without much doubt, that an intervention or mode of teaching is more effective than conventional practice or another approach? To draw such conclusions requires that studies meet a high standard of methodological rigor. Experimental and high-quality quasi-experimental studies would be consulted and confounds had to be carefully assessed to determine if valid inferences could be drawn even from this select group of studies. The Task Group also decided that non-Category 1 studies could be used only to support the conclusions of well-designed experimental or quasi-experimental studies. Groups of compromised studies (e.g., analyses, where weaknesses in one, for example, are off-set by findings in another) could provide context for the analysis conducted by the Task Group and the strength of its recommendation. The Panel also chose to limit itself for the most part to published, peer-reviewed journal articles. The approach is perhaps most similar to that used by the What Works Clearinghouse. This process eliminated all but a few relevant studies in several topical areas. This was true here as well—for the report on strategies used to serve gifted students.

Using the criteria established, the Task Group conducted a literature search for studies that assessed effectiveness of various options for serving gifted students. Key terms used included enrichment, differentiated curriculum, and acceleration. Only studies that compared gifted students participating in an intervention with a comparison group composed of nonparticipating gifted students were included. Studies that employed other comparison groups (e.g., students several grades above the treatment group, norms, or non-gifted) are not included. Finally, the Task Group generally used the term gifted to refer to students at the 90th percentile or above on standardized mathematics achievement tests, although most of the studies included here used much more selective criteria. The literature search, operating within these constraints, initially produced 11 studies, one of which was immediately eliminated due to methodological design weaknesses. The remaining ten were then reviewed by an independent methodologist, who also assessed them in relation to the Panel criteria. The Task Group followed his guidance, which resulted in two more studies being eliminated. Additional suggestions for studies to be consulted that emerged from the review process or in discussion were followed up and subjected to the same review criteria.

Of the eight studies that were included in this report on serving the needs of gifted children, all were either Category 1 or 2 studies (as described in the Methodology document in Appendix A). One was a randomized control trial (RCT) and seven were quasi-experimental. The methodological limitations of each study are clearly presented below. The Task Group organized the studies based on the type of approach toward instruction into two main categories: (i) Acceleration practices, including individualized, self-paced learning and (ii) Enrichment with or without acceleration.

In the following sections, the Task Group describes the practices used, provides study characteristics for each of the studies, and calculates effect sizes for the outcomes when possible.

A. The Role of Acceleration in Gifted Students' Math Achievement and Math-Related Outcomes

Acceleration of the curriculum, as noted above, is one form of adapting the instructional experiences received by gifted students. The curriculum is adjusted to meet the needs of the individual learner or, rather, the individual is placed in the curriculum at the approximate level of his or her functioning. Some call this placement according to competence or developmental placement (Benbow & Stanley, 1996). Acceleration may include presenting subject matter content earlier (e.g., algebra in 7th grade) or at a faster pace, or both, self-paced learning or compacting of the curriculum, participating in Advanced Placement programs (i.e., college-level classes in high school), taking college classes while in high school, skipping grades, and graduating early from high school and subsequently entering college early. It provides a differentiated curriculum for gifted students by using curricula designed for older students. The opinion of most educators in the field of gifted education, however, is that good acceleration does not stop there (VanTassel-Baska, 1998). It also should explore topics more deeply, probe interconnectedness of concepts, and adjust the content to make it more complex and abstract. This can occur in special accelerated classes for gifted students or in the regular classroom with a truly excellent teacher.

Several points need to be considered when evaluating the value of acceleration for gifted students. Acceleration, beyond self-paced learning or offering algebra to eighth-graders, is reserved for the highly gifted. Second, because of social and academic disruptions it causes, acceleration is used only with students who want to accelerate (VanTassel-Baska, 1998). No matter how positive the effects of acceleration could be, it is a widely held professional opinion that it is inadvisable to accelerate a child if there is significant resistance (Benbow, 1998). Thus, this educational intervention is different from others (e.g., choosing a specific text-book or teaching method) because student choice is a factor in its use. It may be that those students who choose to accelerate are more academically motivated or desire academic challenges more than those who choose not to. Alternatively, those who choose to not accelerate may need more of other, nonacademic, factors in structuring a satisfying life. That is, accelerates and non-accelerates may have different priorities and this is confounding when the aim is to assess effects of acceleration specifically. Thus, any recommendations would pertain only to academically motivated students, the very ones for whom acceleration is to be used according to practice guidelines developed on the basis of professional judgment.

Third, when gifted students are accelerated by putting them together for special classes, this creates a different academic and social environment that appears to be highly valued by and motivating for gifted students (Benbow, Lubinski, & Suchy, 1996). In this descriptive study, students report feeling affirmed and challenged in ways that the regular classroom does not provide. Also, the nature of the discourse changes, becoming much more high-level and intellectually challenging (Fuchs, Fuchs, Hamlett, & Karns, 1998). So, accelerated classes are more than just content taught at a fast pace. This makes it hard, if not impossible, to separate out the effects attributable only to the acceleration in these types of programs.

The analyses presented in Table 30 include six quasi-experimental studies that looked at acceleration and include both short- and long-term outcomes; that is, students were assessed shortly after having been accelerated (e.g., completion of self-paced learning program) or several years later. In the latter studies, the short-term impact on learning is not assessed (e.g., covering two years of mathematics in one) but rather subsequent participation in mathematics in the years following the advancement (e.g., college course-taking four years later) is assessed. In the analyses, the Task Group presents individual but not pooled effects sizes of the acceleration practices. The "interventions" were seen as sufficiently different to preclude pooling. Findings associated with both the short-term and long-term outcomes are presented in Table 30.

Table 30: Studies That Examine the Impact of Acceleration on Gifted Students' Math Achievement and Math Related Outcomes

Study	Design	Sample	Duration/ Content	Contrast	Measure	Hedge's g	Standard Error	
Short-Term Effects								
Brody & Benbow, 1990: Study 1[c]	Quasi	80 seventh-grade participants in the Talent Search sponsored by the Center for Talented Youth (CTY) at Johns Hopkins University. All subjects were screened with SAT-M to meet eligibility requirements for participation in gifted educational programs	Three weeks/ Algebra 1 or above	Fast-paced accelerated summer math class vs. No summer program	SAT-Math	0.241	(ns)	0.227
Ma, 2005[b d e]	Quasi	276 gifted seventh-grade students randomly selected from the Longitudinal Study of American Youth (LSAY)	One school year/ Algebra 1	Took Algebra 1 in grades 7 or 8 vs. Did not take Algebra 1 in grades 7 or 8	LSAY Math Achievement (combination of sub-tests: math basic skills, algebra, geometry)[f]	0.167	(ns)	0.166
Parke, 1983	Quasi	44 gifted students in Grades K–2 from two elementary schools	10 weeks/ Addition, subtraction, place value, sets, and measurement	Used self-instructional math materials vs. Control	Skill Mastery[g]	Insufficient data to calculate effect sizes		
Ysseldyke et al., 2004[a]	Quasi	100 gifted students in Grades 3–5 that were part of a larger study in which instructional software was implemented in 15 different states in the U.S.	Four months/ Individually based mathematics curriculum tailored to grade level/skill level	Personalized Computer Instruction vs. Control	STAR math computer adaptive test of mathematics skills in numeric concepts, computation, and math applications	0.449	~	0.271

Continued on p. 6-160

Table 30, continued

Study	Design	Sample	Duration/ Content	Contrast	Measure	Hedge's g	Standard Error	
Long-Term Effects								
Swiatek & Benbow, 1991[a]	Quasi	37 qualifying participants and 58 nonparticipants of a fast-paced/self-paced accelerated math classes that were followed up 10 year after participation. The subjects were initially identified through the Study of Mathematically Precocious Youth (SMPY) referrals and screened through additional testing	Saturday mornings for approximately one year/ Algebra 1 and 2, plane geometry, trigonometry, and analytic geometry	Participation in extracurricular fast-paced and self-paced accelerated math classes vs. No participation	Percent taking elective undergraduate math courses (ES = 0.066), Percent undergraduate majors in math (ES = 0.271), Percent graduate majors in applied math (ES = 0.514~)	0.284	(ns)	0.317
Swiatek & Benbow, 1991[b]	Quasi	107 pairs of gifted students that were followed up 10 years after participation in SMPY	N/A	Students who chose to undergo acceleration and enter college at least one year early vs. Students who chose traditional educational route	Number of non-required math courses (ES = 0.381**), Number of undergraduate math courses (ES = 0.091)	0.236	~	0.137

~ p < .10, * p < .05, ** p < .01, *** p < .001
[a] Data were adjusted for clustering that occurred within classrooms.
[b] Data were adjusted for clustering that occurred within schools.
[c] The standard deviations used for the calculation of effect size were estimates provided by the author.
[d] The number of students used for the gifted analytic sample was adjusted by the 12% attrition that the overall study had in the first two years.
[e] The seventh and ninth grade achievement tests were used as pre- and posttests to estimate effect of taking Algebra 1 in seventh or eighth grade.
[f] The three sub-measures of LSAY Math Achievement Test—math basic skills, algebra, geometry—were combined to create a score comparable to the SAT-M.
[g] The assessment included 170 items testing mastery of 82 skills in five areas: addition, subtraction, place value, sets, and measurement.

Brody and Benbow (1990; Study 1), in a quasi-experimental study, investigated whether short-term, accelerative academic training had an effect on SAT scores of middle school students who were in the top 1% in ability. Program participants were enrolled in a fast-paced, three-week summer academic program that was focused on increasing content knowledge in pre-algebra. Their performance on the SAT was compared to a nonrandomized control group not participating in any accelerative learning experience during the summer but in the top 1% in ability (with lower SAT-M scores initially but not SAT-V) or to students enrolled in other, nonmathematics accelerative classes in the academic summer program. The results from the ANCOVA, adjusting for relatively large initial differences in ability across groups, revealed that in-depth instruction over a short period of time in specific mathematical or verbal areas had little or no impact on SAT scores at the conclusion of the program. The data reported in Table 30 reflects no significant effect between the participants in the fast-paced mathematics class and those not enrolling in any special mathematics class (ES = 0.241).

Ysseldyke et al. (2004), in a quasi-experimental study, compared third through sixth grade gifted students whose mathematics curriculum was differentiated and adjusted to the needs of the students through a instructionally-based curriculum management system, called Accelerated Math (Renaissance Learning, 1998). It was a self-directed, four-month long mathematics program with assessment of skill level, tailoring of the instruction to match skill level, individual pacing and goal setting, ample practice, and immediate feedback to student and teacher on performance. The effect size of 0.449 bordered on statistical significance, and favored the personalized instruction group on outcome measures of mathematics skills, numeric concepts, computation, and mathematics applications.

Parke (1983), in another quasi-experimental study, used a 10-week self-instruction program in mathematics with gifted Kindergarten through second-graders to differentiate the curriculum and compared their performance to another equally able, high-ability sample and a comparison group, both of whom were enrolled in a regular class with no differentiation. In terms of design limitations, there was no random assignment; the sample size was small; and there were large pretest ability differences among the groups. Data needed to compute an effect size also were not reported. The ANCOVA results reported by the author, adjusted for initial differences in ability, were statistically significant. The intervention group mastered significantly more concepts and skills than the comparison groups. The adjusted means were 52 learned concepts for the participants, 38 for the high-ability comparison group, and 41 for the random comparison groups. This finding lends support to the value of differentiation through individualization via self-paced, accelerative learning.

Ma (2005) compared the mathematics achievement at the end of high school for students in the top 10% in ability who took formal algebra either in seventh or eighth grade, an increasing trend in the U.S. for capable students, and equally able students who took such algebra in ninth grade or beyond. Ma used a subsample of the Longitudinal Study of American Youth that was divided into gifted, honors, and regular students. For each, differences between students who took Algebra I early (accelerated) versus those who did not, mainly reflecting practices in different schools, were examined. There were relatively balanced numbers of accelerated versus not accelerated for the gifted (49%) and honors (21%) students, but few “regular” students were accelerated in this way (4%) as would be expected. The mathematics achievement outcome variable, which captured performance on a combination of basic mathematical skills, algebra, geometry, and quantitative literacy items, was a growth curve of achievement measures from Grade 7 to 12. The effect size for the score differences favoring accelerates was 0.167 and not statistically significant. All accelerated students seemed to perform well on this test, however, even though reservation has been expressed about learning algebra early (e.g., Prevost, 1985).

Swiatek and Benbow (1991a), in a quasi-experimental study, assessed participants 10 years after the completion of two homogeneously grouped and fast-paced mathematics classes. The individuals in these classes had learned algebra and possibly all the content up through precalculus at a rapid rate. These classes were the model for the fast-paced programs that have sprung up across the country in the past 35 years and now serve over 100,000 gifted students annually. The initial class was taught by an experienced math teacher at a rate dictated by the capacity for learning of the most able students in the class. Most students in the class completed four years of mathematics in 14 months and their standardized achievement test scores were well above the 90th percentile on relevant tests of mastery. A subgroup completed just two years of math in that time frame. They were less able initially and, thus, experienced difficulty in keeping up the pace of the faster moving group. The participants in the two fast-paced mathematics classes were compared to students who had been matched on ability but did not attend the class and students who dropped out of the class. All were at least in the top 1% in ability, but there may have been motivational and other differences between participants and nonparticipants. Another limitation was that the same teacher taught both classes (reassuringly, similar results in mathematics has been found with other teachers, e.g., Lunny, 1983; Mezynski, Stanley, & McCoart, 1983). At the end of high school, the participants scored higher on standardized mathematics achievement tests, such as the College Board Math Achievement test, than the nonparticipants or dropouts, despite their younger age, and did not regret their acceleration (Benbow, Perkins, & Stanley, 1983). Ten years after the initiation of the class few statistically significant differences on academic achievement variables emerged between the participants and the comparison group (Table 30). The one effect that borders on significance (ES = 0.514) favored the participants (who also tended to be several years ahead in their educational progress and so were younger at time of comparison on specific variables than nonparticipants). This comparison was the percent of students at age 23 who were attending graduate school in applied mathematics, engineering, and computer science (50% of participants vs. 28% of comparison group).

Swiatek and Benbow (1991b), in a quasi-experimental study, compared, via a 10-year follow-up, mathematically talented students (at least top 1% in ability) who had managed to accelerate their education so that they entered college at least one year early with equally able students who had not entered college early. This was a nonrandomized comparison, but the groups had been matched on gender and pretest SAT scores (within 10 points for mathematics, 20 for verbal). The mathematics achievement outcomes were indirect—number of undergraduate mathematics courses taken, number of non-required mathematics courses taken, mathematics major as an undergraduate or graduate student, and interest, confidence, and perceived ease of mathematics. Only one statistically significant effect size was found on the various outcome variables (see Table 30)—on the number of non-required mathematics courses taken (ES = 0.381). The difference favored accelerates who, of course, also had the advantage of being advanced in their education.

In a correlational study, Sadler and Tai (2007) have demonstrated that learning mathematics at an earlier age than typical or at a faster pace is related to allowing students to become more advanced in their mathematics education and to be better prepared for college science classes. No long-term negative consequences have been found and the evidence suggests that there are possibly some small additional advantages.

B. The Role of Enrichment on Gifted Students' Mathematics Achievement

The following section presents findings from the remaining two studies that utilized primarily enrichment to differentiate the curriculum for gifted students (Robinson et al., 1990; Robinson, 1997). Because the interventions differed in that one also explicitly adjusted the pace of instruction, the effects are presented individually (see Table 31).

.

Table 31: Studies That Examine the Role of Computer Instruction, Enrichment, and Cooperative Learning on Gifted Students' Math Achievement

Study	Design	Sample	Duration/ Content	Contrast	Measure	Hedge's g	Standard Error	
Robinson et al., 1990	Quasi	78 elementary age students who participated in a special program for mathematically talented children and 185 program alternates	One school year/ Variety of math content	Curriculum replacement program with a focus on enrichment, self-paced computer instruction, and acceleration vs. Regular curriculum	Math Applied to Novel Situations (MANS)[c]	0.648	***	0.114
Robinson et al., 1997	RCT	310 gifted Kindergarten and 1st-grade students in 158 schools who scored at or above the 98th percentile on a screening test	14 2 1/2 hour sessions per year for two years/ Kindergarten and first-grade curriculum	Extracurricular constructivist enrichment activities (Saturday Club) vs. No enrichment	Pooled measures (Stanford-Binet IV quantitative subtest, Number knowledge test; Woodcock-Johnson calculation subtest)	0.401	**	0.127

~ p < .10, * p < .05, ** p < .01, *** p < .001
[a] Data were adjusted for clustering that occurred within classrooms.
[b] Data were adjusted for clustering that occurred within schools.
[c] Standardized measure MANS measures skills in computation, estimation, mental arithmetic, number representations, relations, number patterns, elucidation, word problems.

Robinson et al. (1997), the only experimental study to emerge out of the literature search, randomly assigned equally able gifted kindergarten and first-grade students to supplemental enrichment mathematics classes conducted on Saturdays over two years or to no treatment. The enrichment classes, with 28 sessions in all, were described by the authors as constructivist in philosophy, "developmentally appropriate," and adhering to NCTM (1989) guidelines. Teachers created social communities that engaged in open-ended problem-solving. At the end of two years, the participants significantly outperformed nonparticipants on a combined mathematics achievement measure. However, there was differential attrition—5% in the control condition, 20% in treatment condition—and it is possible that the least able students left the program at higher rates than the most able. A statistically significant effect size of 0.401 was found favoring the students who participated in the enrichment program. Here, the regular curriculum in the school was not differentiated in any way. Rather, gifted children were challenged through the provision of extra activities, a pull-out model of sorts, and were not explicitly accelerated.

The Robinson et al. (1990) study is similar to the Ysseldyke et al. (2004) study in that it utilized CAI to adjust pace and is quasi-experimental. However, this after-school mathematics program for gifted elementary students also provided specific enrichment activities for the class that allowed students to add breadth and depth to their learning. Hence, the curriculum was differentiated even further than what was possible in Ysseldyke et al. and Parke (1983). This was, however, one single class and hence involved just one teacher. Performance of participants was compared to non-randomized control groups comprised of students who were selected but did not attend or were selected as alternates. Although there was no reporting on pretest mean differences in ability among the groups, a regression discontinuity analysis was used with pretest proxy measures as covariates. With a statistically significant effect size of 0.648, the results lend support to the value of differentiating and enhancing the pace of the curriculum for gifted students.

C. Conclusions

It is generally agreed that good teaching is responsive to individual differences, tailoring instruction to meet the needs of individual learners (Robinson, 1983). In the case of gifted students who are advanced in their skill and concept attainment and can learn new material at a much more rapid rate than their same-age peers (e.g., Lynch, 1992; Stanley & Stanley, 1986; VanTassel-Baska, 1983), it is the professional opinion of those in gifted education that these students need a curriculum that is differentiated (by level, complexity, breadth and depth), developmentally appropriate, and conducted at a more rapid rate (Van Tassel-Baska, 1998). This is typically accomplished to some degree through some combination of acceleration, homogeneous grouping, enrichment, or individualization.

As the Instructional Practices Task Group began its work, it was aware that there were hundreds of studies over decades evaluating the effectiveness of acceleration, in which results have been interpreted as indicating positive academic benefits and no negative effects social-emotionally (see Colangelo, Assouline, & Gross, 2004; and Rogers, 2007 for the latest syntheses). The Task Group did not know the overall quality of these studies or their usefulness for drawing causal attributions. So, it was impossible to decipher the strengths of the signal emitted from these studies and into which category the support for this instructional practice fell (see the Panel Standards of Evidence Document). From a descriptive study, the Task Group learned, however, that gifted students report satisfaction with acceleration (even wishing as adults that they had accelerated more) and that they feel they would not have achieved as much without it (Benbow et al., 2000; Benbow, Lubinski, & Suchy, 1996). But, such data, although valuable, are from the world of perceptions and beliefs and cannot speak to effectiveness.

Enrichment, which attempts to add breadth and depth to the regular curriculum, as well as complexity, also has been studied and has exhibited some positive effects under the same circumstances, limitations, or conditions affecting the interpretability of findings from the literature on acceleration. Yet, many seemingly excellent enrichment programs have not been rigorously evaluated, perhaps because this option for meeting the needs of gifted students has faced less negativity and resistance than is the case for acceleration.

Homogeneous grouping is an educational approach that meets with much controversy as well. Enrichment tends to dominate in homogeneously grouped classes, but it often includes some increased pace of learning. So, there can be settings wherein both acceleration and enrichment is utilized, which most professionals in gifted education would prefer. Before the Task Group began our review of this literature, the state of knowledge, as captured by the results of meta-analyses, revealed positive effects of homogeneously grouped classes, with the value-added gain in one year being about four to five months (Kulik & Kulik, 1982, 1987, 1992). These meta-analyses, however, lumped together studies of various methodological quality, making them less rigorous tests of effectiveness and hence compromised generalizability (also see Delcourt, Cornell, & Goldberg, 2007; Robers, 2007). In terms of perceptions, nonetheless, gifted students when reflecting back as an adult 20 years later seem to favor homogeneous grouping (Benbow et al., 2000). Finally, although often utilized to stimulate gifted children, the effects of mathematics contests on the mathematics achievement of gifted students have not been well studied. This, then, was the state of knowledge before Instructional Practices undertook its analysis.

The Task Group's review of the literature assessing the effectiveness of the various means for tailoring instruction to meet the needs of gifted students yielded surprisingly few studies that met the methodologically rigorous criteria for inclusion adopted by the Task Group. The Task Group actually had to use somewhat less stringent criteria than in other instructional practices reports in order to fulfill the charge of evaluating the "best available scientific evidence." The Task Group could formulate recommendations only on the basis of one randomized control trial study and seven quasi-experimental studies that met the Category 1 and 2 criteria. This was disappointing, especially because even the few studies included in the analyses contained some methodological limitations. For example, almost all studies on acceleration, although essentially positive in their reported outcomes (Colangelo, Assouline, & Gross, 2004 and Rogers, 2007 provide a comprehensive review), are limited to students who are highly gifted and motivated to accelerate. Thus, motivation is a confound just as it is a selection criterion for being considered a candidate for acceleration.

Nonetheless, the studies reviewed above that met our criteria provided some support for the value of differentiating the mathematics curriculum, especially when acceleration is a component (i.e., pace and level of instruction is adjusted). Individualized instruction in which pace of learning is increased, often managed via computer instruction, also showed positive benefits.

The challenge of implementing random assignment or well-matched comparison groups in programs for gifted students is substantial. Parents are unlikely to agree to let their child participate in anything but the treatment that is designed for acceleration or enrichment. Thus, to gain insights about the impact and nature of different approaches to the mathematical education of the gifted, it would be useful to look at some research that does not meet the inclusion criteria because it has been designed to be more descriptive or correlational. That research when coupled with the analyses reported above suggests several positive directions. For instance, there is evidence that gifted students who are accelerated by other means gained time and reached educational milestones earlier (e.g., college entrance) than their equally able same-age peers. They also demonstrate comparable or stronger performance than their same-age peers (although with small effect sizes) on a variety of indicators, at younger ages. Together these studies help to illuminate the conclusions drawn from the scientific literature as summarized above. The Task Group has no evidence that acceleration harms the mathematical achievement of the gifted student.

Gifted students who are accelerated also appear to become more strongly engaged in science, technology, engineering, or mathematics areas. This finding fits well with the results of a recent correlational study showing that the more mathematics courses taken in high school, which is facilitated through acceleration, the more likely students are to perform well in science (Sadler & Tai, 2007). Although some have seen acceleration as a cause for concern, there is no evidence in the studies that met our criteria that gaps and holes in knowledge have occurred as a result of acceleration.

Support also was found for supplemental enrichment programs. Of the programs analyzed here, one explicitly utilized acceleration as a program component and the other did not. Of the two studies that met our criteria for inclusion as Category 1 or 2, both studies had significant effect sizes favoring the enrichment treatment. So, although the evidence is somewhat mixed, it suggests a positive effect of enrichment approaches. Other research (e.g., Benbow, 1998; Delcourt, Cornell, & Goldberg, 2007; Rogers, 2007; VanTassel-Baska, 1998; VanTassel-Baska & Brown, 2007) has examined varied enrichment approaches. Clearly understanding the nature of the enrichment activity is crucial to efforts to improve opportunities for gifted students. Self-paced instruction supplemented with enrichment seemed to have a positive impact on student achievement. This supports the widely held view in the field of gifted education that acceleration and enrichment combined should be the intervention of choice.

Underscored by the analysis undertaken by the Task Group is the need for more high-quality experimental and quasi-experimental research to study effectiveness of interventions designed to meet the learning needs of gifted students. Especially missing are evaluations of academically rigorous enrichment programs, the mathematical content explored in such programs, and their goals. The Task Group concludes, however, that it is important for school policies to support appropriately challenging work in mathematics for gifted and talented students. Acceleration, combined with enrichment, is certainly a promising, possibly moderately supported (if the entire literature is considered), practice.

BIBLIOGRAPHY

Benbow, C.P. (1991). Meeting the needs of gifted students through use of acceleration: A neglected resource. In M.C. Wang, M.C. Reynolds, & H.J. Walberg (Eds.), *Handbook of special education* (Vol. 4, pp. 23–36). Elmsford, NY: Pergamon Press.

Benbow, C.P. (1998). Grouping intellectually advanced students for instruction. In J. VanTassel-Baska (Ed.), *Excellence in educating the gifted* (pp. 261–278). Denver, CO: Love.

Benbow, C.P., Lubinski, D., Shea, D.L., & Eftekhari-Sanjani, H. (2000). Sex differences in mathematical reasoning ability: Their status 20 years later. *Psychological Science, 11(6),* 474–480.

Benbow, C.P., Lubinski, D., & Suchy, B. (1996). Impact of the SMPY model and programs from the perspective of the participant. In C. P. Benbow & D. Lubinski (Eds.), *Intellectual talent: Psychometric and social issues* (pp. 266–300). Baltimore, MD: Johns Hopkins University Press.

Benbow, C.P., Perkins, S., & Stanley, J.C. (1983). Mathematics taught at a fast pace: A longitudinal evaluation of SMPY's first class. In C. P. Benbow & J.C. Stanley (Eds.), *Academic precocity: Aspects of its development* (pp. 51–78). Baltimore, MD: Johns Hopkins University Press.

Benbow, C.P., & Stanley, J.C. (1996). Inequity in equity: How current educational equity policies place able students at risk. *Psychology, Public Policy, and Law, 2,* 249–293.

Brody, L.E., & Benbow, C.P. (1990). Effects of high school course-work and time on SAT scores. *Journal of Educational Psychology, 82(4),* 866–875.

Colangelo, N., Assouline, S.G., & Gross, M.U.M. (Eds.) (2004). *A nation deceived: How schools hold back America's brightest students*. Iowa City, IA: University of Iowa.

Coleman, J. S. (1960). The adolescent sub-culture and academic achievement. *America Journal of Sociology*, *65(4)*, 337–347.

Cramond, B., & Martin, C.E. (1987). In-service and preservice teachers' attitudes towards the academically brilliant. *Gifted Child Quarterly, 31(1)*, 15–19.

Dark, V.J., & Benbow, C. (1990). Mathematically talented students show enhanced problem translation and enhanced short-term memory for digit and spatial information. *Journal of Educational Psychology, 82,* 420–429.

Dark, V.J., & Benbow, C.P. (1991). Differential enhancement of working memory with mathematical and vertical precocity. *Journal of Educational Psychology, 83(1),* 48–60.

Dark, V.J., & Benbow, C.P. (1994). Type of stimulus mediates the relationship between performance and type of precocity. *Intelligence, 19(3),* 337–357.

Delcourt, M.A.B., Cornell, D.G., & Goldberg, M.D. (2007). Cognitive and affective learning outcomes of gifted elementary school students. *Gifted Child Quarterly, 51(4),* 359–381.

Fuchs, L.S., Fuchs, D., Hamlett, C.L., & Karns, K. (1998). High-achieving students' interactions and performance on complex mathematical tasks as a function of homogeneous and heterogeneous pairings. *American Educational Research Journal, 35(2),* 227–268.

Hofstadter, R. (1963). *Anti-intellectualism in American life*. New York: Knopf.

Kaplan, S. (1986). The grid: A model to construct differentiated curriculum for the gifted. In J.S. Renzulli (Ed.), *Systems and models for developing programs for the gifted and talented* (pp. 180–193). Mansfield Center, CT: Creative Learning Press.

Kulik, C.C., & Kulik, J.A. (1982). Effects of ability grouping on secondary school students: A meta-analysis of evaluation findings. *American Educational Research Journal, 19(3)*, 415–428.

Kulik, J.A., & Kulik, C.C. (1984). Effects of accelerated instruction on students. *Review of Educational Research, 54(3)*, 409–425.

Kulik, J.A., & Kulik, C.C. (1987). Effects of ability grouping on school achievement. *Equity and Excellence, 23,* 22–30.

Kulik, J.A., & Kulik, C.C. (1992). Meta-analytic findings on grouping programs. *Gifted Child Quarterly, 36(2),* 73–77.

Learned, W.S., & Wood, B.D. (1938*). The student and his knowledge. A report to the Carnegie foundation on the results of the high school and college examinations of 1928, 1930, and 1932*. New York: Carnegie Foundation for the Advancement of Teaching.

Lunny, J.F. (1983). Fast-paced mathematics classes for a rural county. In C.P. Benbow & J.C. Stanley (Eds.), *Academic precocity: Aspects of its development* (pp. 79–85). Baltimore, MD: Johns Hopkins University Press.

Lynch, S.J. (1992). Fast-paced high school science for the academically talented. A six-year perspective. *Gifted Child Quarterly, 36(3)*, 147–154.

Ma, X. (2005). A longitudinal assessment of early acceleration of students in mathematics on growth in mathematics achievement. *Developmental Review, 25(1),* 104–131.

Mezynski, K., Stanley, J.C., & McCoart, R.F. (1983). Helping youth score well on AP examinations. In C.P. Benbow & J.C. Stanley (Eds.), *Academic precocity: Aspects of its development* (pp. 86–112). Baltimore, MD: Johns Hopkins University Press.

National Council of Teachers of Mathematics (NCTM) (1989). *Curriculum and evaluation standards for school mathematics.* Reston, VA: National Council of Teachers of Mathematics.

National Research Council (2000). *How people learn: Brain, mind, experience, and school.* Washington, DC: National Academy Press.

National Research Council (2002). *Learning and Understanding*. Washington, DC: National Academy Press.

Olszewski-Kubilus, P. (2007). Best Practices in Gifted Education (Special Issue). *Gifted Child Quarterly, 51*(4).

Parke, B.N. (1983). Use of self-instructional materials with gifted primary-aged students. *Gifted Child Quarterly, 27(1),* 29–34.

Paterson. D.G. (1957). Conservation of human talent. *American Psychologist*, *12*, 134–144.

Pressley, S.L. (1949). *Educational acceleration: Appraisals and basic problems.* Columbus, OH: Ohio State University.

Prevost, F.J. (1985). Soundoff: Let’s not teach algebra to eighth graders! *Mathematics Teacher, 78(8),* 586–587.

Renaissance Learning (1998). *Accelerated math [computer software]*. Wisconsin Rapids, WI: Author.

Renzulli, J.S. (1986). The three-ring conception of giftedness: A developmental model for creative productivity. In R.J. Sternberg & J.E. Davidson (Eds.), *Conceptions of giftedness* (pp. 53–92). New York: Cambridge University Press.

Robinson, A., Bradley, R.H., & Stanley, T.D. (1990). Opportunity to Achieve: Identifying Mathematically Gifted Students. *Contemporary Educational Psychology, 15(1),* 1–12.

Robins, H. (1983). A case for radical acceleration. In J.C. Stanley & C.P. Benbow (Eds.), *Academic Precocity: Aspects of its Development.* Baltimore, MD: Johns Hopkins University Press.

Robinson, N.M., Abbott, R.D., Berninger, V.W., Busse, J., & Mukhopadhyay, S. (1997). Developmental changes in mathematically precocious young children: Longitudinal and gender effects. *Gifted Child Quarterly, 41(4),* 145–158.

Rogers, K.B. (2007). Lessons learned about educating the gifted and talented: A synthesis of research on educational practice. *Gifted Child Quarterly, 51(4),* 382–396.

Sadler, P.M., & Tai, R.H. (2007). The two high-school pillars supporting college science. *Science, 317,* 457–458.

Seashore, C.E. (1922). The gifted student and research. *Science, 56(1458),* 641–648.

Stanley, J.C. (2000). Helping students learn only what they don't already know. *Psychology, Public Policy, and Law, 6(1),* 216–222.

Stanley, J.C., & Stanley, B.S.K. (1986). High school biology, chemistry, and physics learned well in three weeks. *Journal of Research in Science Teaching, 23(3),* 237–250.

Swiatek, M.A., & Benbow, C.P. (1991a). A 10-Year longitudinal follow-up of participants in a fast-paced mathematics course. *Journal for Research in Mathematics Education, 22(2),* 138–150.

Swiatek, M.A., & Benbow, C.P. (1991b). Ten-year longitudinal follow-up of ability matched accelerated and unaccelerated gifted students. *Journal of Educational Psychology, 83(4),* 528–538.

Tannenbaum, A.J. (1962). *Adolescent attitudes towards academic brilliance*. New York: Bureau of Publications, Teachers College, Columbia University.

Terman, L.M. (1954). The discovery and encouragement of exceptional talent. *American Psychologist*, *9*, 221–230.

Tyler, L.E. (1965). *The psychology of individual differences* (3rd Edition). New York: Meredith.

Van Tassel-Baska, J. (1983). Illinois' state-wide replication of the Johns Hopkins Study of Mathematically Precocious Youth. In C.P. Benbow & J.C. Stanley (Eds.), *Academic Precocity: Aspects of its Development* (pp. 179–191). Baltimore: Johns Hopkins University Press.

Van Tassel-Baska, J. (1998). *Excellence in Educating Gifted & Talented Learners.* Denver, CO: Love.

Van Tassel-Baska, J., & Brown, E.F. (2007). Toward best practice: An analysis of the efficacy of curriculum models in gifted education. *Gifted Child Quarterly, 51(4),* 342–358.

Willerman, L. (1979). *The psychology of individual and group differences*. San Francisco: Freedman.

Ysseldyke, J., Tardew, S., Betts, J., Thill, T., & Hannigan, E. (2004) Use of an instructional management system to enhance math instruction of gifted and talented students, *Journal for the Education of the Gifted, 27(4)*, 293–310.

VII. Teachers' Use of Formative Assessments to Improve Learning of Mathematics: Results from a Meta-Analysis of Rigorous Experimental and Quasi-Experimental Research

Formative assessment—ongoing monitoring of student learning to inform instruction—is generally considered a hallmark of effective instruction in any discipline. In the past 20 years, the term has been used in two complementary but distinct traditions of scholarship and research. One tradition—which has played an influential role in the field of mathematics education—is represented, for example, in two recent publications by the National Research Council—one on the teaching and learning of mathematics (National Research Council, 2001) and one on human learning and cognition (National Research Council, 2005). Donovan and Bransford (National Research Council, 2005) established three goals for formative assessments: 1) "to make students' thinking visible to both teachers and students;" 2) "to monitor student progress (in mastering concepts as well as factual information);" and, 3) to "design instruction that is responsive to student progress" (p. 16).

The second tradition, developed primarily within the fields of school psychology, educational psychology, and special education shares only one of the three aforementioned goals. This approach is typified, for example, by the definition provided in the recent *Encyclopedia of School Psychology* (Lee, 2005). In this tradition, formative assessments are tools "used to monitor progress and to provide feedback about the progress being made. ... In the classroom setting, formative evaluation is used to inform students and teachers about progress" (p. 209). The goal of this type of formative assessment is to determine whether specific students—or in some cases, an entire class—require additional instruction devoted to learning a particular concept or acquiring proficiency in a particular mathematical procedure or strategy for problem solving. Typically, measures are administered weekly or biweekly, often using computer-assisted assessment. They are brief and efficient, taking approximately 5 minutes to administer.

Note that both research traditions stress monitoring progress toward an instructional goal and adjusting instruction for students based on the formative measures. However, the school psychology tradition devotes a good deal of attention to empirical establishment of the validity and reliability of the assessment procedures (e.g. Fuchs, 2004; Foegen, Jiban, & Deno, 2007).

The system described by Donovan and Bransford (2005) rarely addresses psychometric issues. It presents an ambitious agenda that includes not only progress monitoring but also interpretation of students' errors and misconceptions to guide the types of questions teachers ask to probe for student understanding. This tradition is embodied for example, in *Adding It Up (*National Research Council, 2001):

> *Information about students is crucial to a teacher's ability to calibrate tasks and lessons to students' current understanding.… In addition to tasks that reveal what students know and can do, the quality of instruction depends on how teachers interpret and use that information. Teachers' understanding of their students' work and the progress they are making relies on … their ability to use that understanding to make sense of what the students are doing. (pp. 349–350).*

Ruiz-Primo, Shavelson, Hamilton, and Klein (2002) discuss a continuum of assessment distance that traverses both traditions. The continuum includes the following distance classifications, with the classifications ranging from formative to summative assessments, by proximity:

- **Immediate—**informal observation or artifacts from a lesson.
- **Close—**embedded assessments and semi-formal quizzes following several activities.
- **Proximal—**formal classroom exams following a particular curriculum.
- **Distal—**criterion-referenced achievement tests such as those required by NCLB.
- **Remote—**broad outcomes measured over time—norm-referenced tests, such as the Scholastic Aptitude Test.

Formative assessment would be identified as immediate, close and perhaps proximal in the continuum above and is used to regularly monitor instruction. Freudenthal (1973) noted, "It is more informative to observe a student during a mathematical activity than to grade his papers" (p. 84). Informal assessments which include observations and informal probes of students to assess their level of understanding, according to Freudenthal, need to inform day-to-day teaching.

Sueltz, Boynton, and Sauble (1946) noted that observation, discussion, and interviews serve better than paper-pencil tests in evaluating a pupil's ability to understand the principles he or she uses. Spitzer (1951) and others have long advocated the interview as a formative assessment strategy that is closely associated with the use of observations. Decades later the National Council of Teachers of Mathematics noted that information is best collected through informal observation as students participate in class discussions, attempt to solve problems, and work on various assignments individually or in groups (National Council of Teachers of Mathematics, 1989, p. 233). However, Glaser and Silver (1994) note that aside from teacher-made classroom tests, the integration of assessment and learning as an interacting system has been too little explored. The meta-analysis completed by the Instructional Practices Task Group revealed no methodologically acceptable studies that examined the impact of using this type of assessment on student performance.

The goal of this section is to review the experimental and quasi-experimental research on the extent to which teachers' use of formative assessments in mathematics enhances students' acquisition of mathematics content. Although the Task Group reviewed the literature for studies using any type of formative assessment from any tradition, the only studies located that met the criteria for adequate experimental design emanated from the school psychology or educational psychology traditions.

This report describes studies that indicate the extent to which use of formative assessments improves students' mathematics proficiency. The Task Group also describes the impacts of various enhancements, i.e., procedures and strategies for helping teachers use this information to provide differentiated instruction, and, thus enhance mathematics achievement. The Task Group centered the meta-analysis on two research questions. The first was whether teachers' use of formative assessments enhanced student achievement in mathematics. The second question explored the effectiveness of various tools or enhancements that can assist teachers in their use of formative assessments. Before discussing the findings of the meta-analysis, the Task Group begins by providing a brief historical overview.

A. Historical Overview

In the 1970s and 1980s, a good deal of research and effort went into the field of formative assessment. Researchers examining high-performing schools invariably found that the school used some system to monitor all students' academic progress on a regular basis. Mastery learning (e.g., Bloom, 1980; Guskey, 1984), a form of differentiated instruction was widely implemented in large school districts such as Chicago and San Diego. Mastery learning calls for frequent assessment of student progress using brief tests at the end of each unit, which is typically once a week. Students who did not reach a mastery level (typically defined as 80% correct) are retaught the material. Those with scores above the mastery level are provided with extension or enrichment activities. During this time in history, publishers of the major mathematics curricula began to include unit tests along with their programs. This practice continues to this day.

B. Validity and Reliability Concerns for Formative Assessments in Mathematics

In developing formative assessments in mathematics, the goal has invariably been to develop measures that are valid and reliable in the psychometric sense (AERA, APA, NCME, 1999) and that are relatively easy to administer and score.

Contemporary conceptions of test validity include indices that a measure is correlated with other measures of mathematics achievement which can include teacher appraisals (*criterion-related validity*), that the mathematical content is valid and important (*content validity*) and that there is evidence concerning the impact of use of the measure, including both intended and unintended consequences (*consequential validity*) (e.g., Messick, 1988). A valid formative assessment system should actually help teachers "make specific instructional decisions" (National Research Council, 2001, p. 35) and according to Gersten, Keating, and Irvin (1995), it should also provide data that indicates that use of the system is beneficial to students.

A group of researchers found the unit mastery tests problematic for several reasons. When they examined the psychometric characteristics of these measures (Fuchs, Tindal, & Fuchs, 1986; Tindal, Fuchs, Fuchs, Shinn, Deno, & Germann, 1985) they found them to be weak. In addition, difficulty levels varied from week to week, and the cut score of 80 or 85% seemed increasingly arbitrary. Unit mastery tests, sometimes called criterion-referenced tests in this era, did nothing to assess retention of previously taught material.

Almost 25 years ago, two seminal articles called for a radically different, seemingly counterintuitive approach to formative assessment (Fuchs, Deno, & Mirkin, 1984; Deno, 1986). This approach entailed a sampling of items representing major instructional objectives for the year and periodic use of assessments with items that randomly sampled across the year's objectives. This approach seems counterintuitive in that, during the early parts of the year, students are asked to solve problems involving material not yet covered. Toward the end of the year, they are asked about material they may have covered 6 to 8 months ago.

Yet, therein lies the power of a formative assessment that contains items from across the year's objectives. It is a far more accurate means to measure progress because the difficulty remains more or less the same across the year, and teachers and students can actually see the progress they have made toward acquiring the material. In contrast, typical mastery learning tests' difficulty level varied from unit to unit, depending on both the difficulty of the topic and the difficulty of the items selected. In addition, with this type of assessment system, students could actually see their progress; from say a score of 20% correct to 90% correct, as the year progressed. Even in the best of unit mastery tests, students will typically score at about the same level from unit to unit. Another advantage of this type of system is that it automatically assesses both retention of material taught months ago and, to some extent, a student's ability to generalize what she learned to unfamiliar material. Because each of the brief measures samples broadly across the years' objectives, criterion-related validity is far superior to assessments that only cover one week's unit. For all these reasons, these measures have consistently shown far superior reliabilities and criterion-related validity than traditional unit mastery tests (see Fuchs (2004) and Foegen et al. (2007) for extensive reviews). They also have consistently demonstrated *construct validity,* in particular, in terms of sensitivity to instruction, i.e. use as a means to reliably monitor student progress.

Especially in the field of reading, a second type of formative assessment was used, which also possessed strong psychometric qualities in terms of criterion-related validity (i.e., correlation with state- or nationally-normed achievement test) and reliability. These measures are typically called *robust indicators*. Foegen et al. (2007) define them as:

> *Measures that represent broadly defined proficiency in mathematics … Effective measures are not necessarily representative of a particular curriculum but are instead characterized by the relative strength of their correlations to various overall mathematics proficiency criteria (p. 4).*

These measures are "not necessarily drawn from the student's … (actual) ... curriculum, yet offer strong correlations to a host of criterion measures of overall subject area proficiency" (p. 4).

A potential advantage of robust indicators is that they can "create a seamless and flexible system of progress monitoring measures in mathematics … across multiple grade levels. The search for robust indicators represents an effort to identify aspects of core competence in mathematics … that are predictive of important outcomes in mathematics, regardless of the vagaries of specific curriculum programs or high stakes state tests."

Several robust indicators have been developed in the field of mathematics, especially for students in the primary grades. These include measures of number naming (e.g., VanDerHeyden, Witt, Naquin, & Noell, 2001; Chard, Clarke, Baker, Otterstedt, Braun, & Katz, 2005), magnitude comparison (e.g., Clarke & Shinn, 2004) and counting proficiency. Validity coefficients tend to be higher for first grade assessments than kindergarten assessments and highest for magnitude comparison measures. A drawback of these measures is that they are only useful for one grade level so that they cannot be used to assess progress over multiple years.

At the middle school level, Foegen (2000) developed two robust indicator measures: one of fluency with basic arithmetic combinations (i.e., facts) and the second, an estimation task. The estimation task was a timed measure and attempted to measure students' number sense (as opposed to computational skill).

Helwig and Tindal (2002) developed a measure that focused on conceptual understanding using an item bank developed for eighth-graders. In general, these measures demonstrated adequate criterion related validity, although the Helwig et al. conceptual measure demonstrated the strongest correlations with high-stakes assessments.

The authors note that, with one or two exceptions, neither the robust indicators approach nor the sampling from annual state curricular objectives approach have generated the same high levels of criterion related validity that oral reading fluency has in the field of reading. A major benefit of sampling from annual objectives is that teachers can use these data to obtain a sense of topics that require additional attention for groups of students. There are, however, several drawbacks.

The first is that, in order to be efficient, the sample of items should be limited. However, a limited sample of items may cause potential reliability issues. The second is that, at the current point in time, state standards in mathematics are quite variable in terms of quality, and different states provide differing emphases to topics and sequence topics differently (Reys, Dingman, Sutter, & Teuscher, 2005). Current efforts to use a common framework such as the *NCTM Focal Points* (National Council of Teachers of Mathematics, 2006) may help alleviate this problem in the future.

Both these types of measures were, in our view, unfortunately given the term, *curriculum based measurement.* That term seems to imply that they are valid, for example, only for a given curricula. Yet, in reality they are aligned to various district or states' mathematics standards. In 2007, Deno reported that the term used to describe this formative assessment approach was an unfortunate choice (Deno, 2007). Unlike unit mastery tests, the curriculum objective can basically be used for any curricula in use in a district because they gauge progress toward state standards.

Virtually all the applied experimental research on formative assessments has involved the use of these types of measures and an understanding of (a) the extent to which providing teachers and students with this information enhances mathematics achievement and, increasingly, (b) the efficacy of various tools and procedures for helping teachers use this information to provide differentiated instruction. This is the focus of the remainder of this section.

Most of the formative assessments used in mathematics demonstrate criterion-related validities in the 0.5 to 0.7 range (Foegen et al., 2007). Although these are weaker than those found in reading, they appear to be reasonable. It is in the area of *consequential validity*, that formative assessment measures have shown their greatest utility (Gersten et al., 1995). In the next section, the Task Group reviews the research on this topic that (a) examines impacts/effects of use of formative assessments and (b) addresses the standards for experimental and quasi-experimental design utilized by the Task Group.

C. Results

Table 32 presents the contrast between use of formative assessment on a regular (typically biweekly) basis versus a control condition. Six of the studies analyzed data at the individual student level, and the remaining three used the classroom or teacher as the unit of analysis. (One study (Calhoon & Fuchs, 2003) did not compare formative assessment to a control condition; it measured the impact of an enhanced version of formative assessment to a control condition. This study is thus excluded from this analysis.)

Table 32: Studies that Investigate the Impact of Formative Assessment (FA) Versus a Control Condition

Study	Design	Sample	Duration/ Content	Contrast	Measure	Subgroup	Hedge's g	Standard Error	
Student-level analyses									
Allinder et al., 2000[a]	RCT	38 learning disabled elementary students and 22 teachers in a large midwestern school district	School year/ Curriculum based measurement in math computation	FA only vs. Control	Math Computation Test-Revised	Overall	-0.012	(ns)	0.349
Fuchs et al., 1990[a]	RCT	56 learning disabled students (Grades 3–9) and 20 elementary special education teachers in a southeastern metropolitan school district	15 weeks/ Individualized math programs	FA with performance indicator only vs. Control	Math Computation Test-Revised (Combined problems and digits)	Overall	0.239	(ns)	0.323
Fuchs et al., 1994	RCT	30 students and teachers in Grades 2–5 in a southeastern district	25 weeks/ Classwide program - general math operations	FA only vs. Control	Math Operations Test-Revised	Average achieving	0.310	(ns)	0.379
		30 students and teachers in Grades 2–5 in a southeastern district				Low achieving	0.015	(ns)	0.377
		30 students and teachers in Grades 2–5 in a southeastern district				Learning disabilities	0.189	(ns)	0.378
Fuchs et al., 1996[a]	RCT	22 learning disabled students (Grades 3–7) and 12 special education teachers in Tennessee metro school district	School year/ Aim was to reintegrate students into mainstream math	FA vs. Control	Math Operations Test-Revised (Digits)	Overall	0.468	(ns)	0.467
		25 learning disabled students (Grades 3–7) and 15 special education teachers in Tennessee metro school district		FA plus TP (trans-environmental programming) vs. TP only			0.220	(ns)	0.429
Fuchs et al., 1999[a]	RCT	272 students (Grades 2–4) and 16 teachers in four schools in one southeastern school district	23 weeks/ Problem solving	Performance Assessments (PA) vs. no PA	Novel problem-solving (based on ITBS problem)	Overall	0.355	(ns)	0.249
Spicuzza et al., 2001[a]	Quasi	495 students in Grades 4 and 5 in multiple schools in a large, midwestern school district	4 months/ Accelerated Math - individualized assignments	Accelerated Math vs. regular math program	NALT (Northwest Evaluation Association) – annual district test	Overall	0.139	(ns)	0.213

Heterogeneity							
Q-value	*Df (Q)*	*P-value*	*I-squared*	*Pooled ES: student level (six studies, nine effect sizes)*	*Hedge's g*	*Standard Error*	
1.509	*8*	*0.993*	*0.000*		*0.206*	*~*	*0.107*

Continued on p. 6-177

Table 32, continued

Study	Design	Sample	Duration/ Content	Contrast	Measure	Subgroup	Hedge's g	Standard Error	
Classroom-level analyses									
Allinder, 1996	RCT	58 students (Grades 3–6) of 29 special education teachers in multiple schools in a large midwest school district	16 weeks/ Math computation	FA vs. Control	Math Computation Test-Revised	Overall	0.558	(ns)	0.387
Fuchs et al., 1989	RCT	40 students (Grades 2–9) of 20 special education teachers in elementary and middle schools in a southeastern metropolitan area	15 weeks/ Individualized math programs	Dynamic goal FA vs. Control	Math Computation Test	Overall	0.600	(ns)	0.439
Fuchs et al., 1991	RCT	42 learning disabled students (Grades 2–8) and 22 teachers in multiple elementary and middle schools in a southeastern metropolitan area	20 weeks/ Math operations	FA only vs. Control	Math Operations Test (combined problems and digits)	Overall	0.045	(ns)	0.426
Heterogeneity									
Q-value	*Df (Q)*	*P-value*	*I-squared*	*Pooled ES: student level (three studies, three effect sizes)*			*Hedge's g*	*Standard Error*	
1.069	*2*	*0.586*	*0.000*				*0.408*	*~*	*0.240*

~ p < .10, * p < .05, ** p < .01, *** p < .001
[a] Data were adjusted for clustering that occurred within teachers or classrooms.

Effect sizes were calculated for studies in which analysis was conducted at the individual student level (N = 6) and those where analysis was conducted at the classroom level (N = 3).

For the student level set, nine effect sizes from the six relevant studies were calculated. There are two reasons for this. The first is that Fuchs, Fuchs, et al. (1994) intentionally sampled students from three strata: those with learning disabilities, a below-average low-performing group, and a group of students performing at or near the class average. For Fuchs, Roberts, Fuchs, and Bowers (1996), there were four conditions, including two involving formative assessment (FA). Thus, effect sizes for two orthogonal contrasts could be calculated.

For the six studies where analyses were conducted at the student level, the mean effect size is 0.206, bordering on significance. For the three studies that use the class as the unit of analysis, the effect size is 0.408, also bordering on significance.

A reasonable inference is that merely providing teachers and students with feedback on how they are progressing is consistently helpful to students. This is a consistent replicable phenomenon across a large number of studies that involve well over a hundred classrooms. The reader should keep in mind that two-thirds of the research has been conducted at the elementary school level and only two include a middle school sample, and a one high school sample. In addition, all have used formative assessment systems that have empirical data to indicate validity and reliability. In addition, all but one (Ysseldyke et al., 2003), have used formative assessments that are based on sample problems selected to represent a randomly selected set of state standards for the year. Thus, there is insufficient evidence to determine whether or not the use of formative assessments is effective in the secondary grades. The next set of studies investigates what additional information is required to assist teachers in how to use these data.

D. Enhancements to Assist Teachers in Use of Formative Assessment

Early on, researchers realized that teachers might not know how to use formative assessment to enhance instruction unless some type of additional guidance was provided. Thus, a set of *enhancements* was developed and field-tested in a series of research studies. These appear in Tables 33 and 34. Table 33 compares the use of formative assessments with enhancement to a control condition (i.e., no formative assessment). Table 34 attempts to estimate the value added to formative assessment by each of these enhancements. Thus, the contrasts in Table 34 compare use of formative assessment with an enhancement to use of formative assessment.

Table 33: Studies That Investigate the Impact of Formative Assessment (FA) Plus Enhancements Versus a Control Condition

Study	Design	Sample	Duration/ Content	Contrast	Measure	Subgroup	Hedge's g	Standard Error	
Student-level analyses									
Allinder et al., 2000[a]	RCT	37 learning disabled elementary students and 20 teachers in a large midwestern school district	School year/ Curriculum based measurement in math computation	FA + teacher self-monitoring of instructional changes vs. Control	Math Computation Test-Revised	Overall	0.588	(ns)	0.369
Calhoon & Fuchs, 2003[a]	RCT	92 high school students with disabilities (Grades 9–12) and three teachers in 10 classrooms in three schools in a southeastern urban district	15 weeks/ Computation, concepts and applications	FA with PALS (Peer-assisted learning strategies) vs. Control	Math Operations Test-Revised (computation)	Overall	0.355	(ns)	0.340
Fuchs et al., 1990[a]	RCT	54 learning disabled students (Grades 3–9) and 20 elementary special education teachers in a southeastern metropolitan school district	15 weeks/ Individualized math programs	FA with performance indicator and skills analysis vs. Control	Math Computation Test-Revised (Combined problems and digits)	Overall	0.398	(ns)	0.325
Fuchs et al., 1994	RCT	30 students and teachers in Grades 2–5 in a southeastern district	25 weeks/ Classwide program - general math operations	FA + instructional recommendations vs. Control	Math Operations Test-Revised	Average achieving	0.292	(ns)	0.379
		30 students and teachers in Grades 2–5 in a southeastern district				Low achieving	0.546	(ns)	0.383
		30 students and teachers in Grades 2–5 in a southeastern district				Learning disabilities	0.172	(ns)	0.377
Fuchs et al., 1996[a]	RCT	24 learning disabled students (Grades 3–7) and 13 special education teachers in Tennessee metro school district	School year/ Aim was to reintegrate students into mainstream math	FA + transenvironmental programming (TP) vs. Control	Math Operations Test-Revised (Digits)	Overall	0.304	(ns)	0.445
Heterogeneity									
Q-value	*df (Q)*	*P-value*	*I-squared*	*Pooled ES: student level (five studies, seven effect sizes)*			*Hedge's g*	*Standard Error*	
0.901	*6*	*0.989*	*0.000*				*0.383*	**	*0.140*
Classroom-level analyses									
Fuchs et al., 1991	RCT	43 learning disabled students (Grades 2–8) and 22 teachers in multiple elementary and middle schools in a southeastern metropolitan area	20 weeks/ Math operations	FA with expert system instructional consultation vs. Control	Math Operations Test (combined problems and digits)	Overall	0.657	(ns)	0.439

~ p < .10, * p < .05, ** p < .01, *** p < .001

[a] Data were adjusted for clustering that occurred within teachers or classrooms.

Table 34: Studies That Investigate the Impact of Formative Assessment (FA) Plus Enhancements Versus Formative Assessment Only

Study	Design	Sample	Duration/ Content	Contrast	Measure	Subgroup	Hedge's g	Standard Error	
Student-level analyses									
Allinder et al., 2000[a]	RCT	33 learning disabled elementary students and 18 teachers in a large midwestern school district	School year/ Curriculum based measurement in math computation	FA + teacher self-monitoring of instructional changes vs. FA only	Math Computation Test-Revised	Overall	0.603	(ns)	0.386
Fuchs et al., 1990[a]	RCT	72 learning disabled students (Grades 3–9) and 20 elementary special education teachers in a southeastern metropolitan school district	15 weeks/ Individualized math programs	FA with performance indicator and skills analysis vs. FA with performance indicator only	Math Computation Test-Revised (Combined problems and digits)	Overall	0.234	(ns)	0.291
Fuchs et al., 1994	RCT	20 students and teachers in Grades 2–5 in a southeastern district	25 weeks/ Classwide program—general math operations	FA + instructional recommendations vs. FA only	Math Operations Test-Revised	Average achieving	-0.021	(ns)	0.428
		20 students and teachers in Grades 2–5 in a southeastern district				Low achieving	0.453	(ns)	0.434
		20 students and teachers in Grades 2–5 in a southeastern district				Learning disabilities	-0.037	(ns)	0.428
Fuchs et al., 1996[a]	RCT	24 learning disabled students (Grades 3–7) and 13 special education teachers in Tennessee metro school district	School year/ Aim was to reintegrate students into mainstream math	FA + transenvironmental programming (TP) vs. FA only	Math Operations Test-Revised (Digits)	Overall	-0.229	(ns)	0.444
Heterogeneity									
Q-value	*df (Q)*	*P-value*	*I-squared*	*Pooled ES: student level (4 studies, 6 effect sizes)*			*Hedge's g*	*Standard Error*	
2.949	*5*	*0.708*	*0.000*				*0.194*	*(ns)*	*0.159*
Classroom-level analyses									
Fuchs et al., 1991	RCT	41 learning disabled students (Grades 2–8) and 22 teachers in multiple elementary and middle schools in a southeastern metropolitan area	20 weeks/ Math operations	FA with expert system instructional consultation vs. FA without expert system	Math Operations Test (combined problems and digits)	Overall	0.750	~	0.443

~ p < .10, * p < .05, ** p < .01, *** p < .001
[a] Data were adjusted for clustering that occurred within teachers or classrooms.

Specific enhancements included:

1) *Providing teachers with detailed analysis indicating strengths and weaknesses based on formative assessment data*. The analysis describes student proficiency in specific mathematics subskills, and visually presents student proficiency in each subskill across the school year. The detailed information provided by the analysis allows teachers to evaluate how students maintain skills over time and helps teachers decide what to teach. Further, specific areas for instructional change can be targeted based on the information provided (Fuchs et al., 1990) (effect size of 0.398; value-added effect size of 0.234; these effect sizes did not reach statistical significance).

2) *Using data from formative assessments and sophisticated software to provide specific instructional suggestions to teachers for individual students.* Instructional consultation on teacher planning and student achievement was provided by a computerized expert system. Using the formative assessment data, the expert consultation system recommended instructional adjustments and provided detailed instructions on how to implement that adjustment. The consultation not only helped teachers isolate what material to re-teach but also how to restructure their instruction (Fuchs et al., 1991) (ES = 0.657, ns; value added ES = 0.750, bordering on significance.).

3) *Using the data from the formative assessments as a basis for the content of peer-assisted learning sessions.* Results of the formative assessments were entered into a computer program that produced a graph of students' progress overtime and a skills profile of each "student's performance on each skill in the annual curriculum" (p. 240). Teachers used this report to group students into pairs for Peer-Assisted Learning (PALS). The teachers also used the reports to determine the content of the PALS lesson (Calhoon & Fuchs, 2003) [ES = 0.355, nonsignificant (ns)]. This is the only study that addressed high school mathematics instruction.

4) *Using the data from formative assessments as the basis for consultation between the classroom teacher and the special educator to determine what content to emphasize.* The data from the formative assessments was used by the special education teacher to provide classroom teachers with specific information on which curricular areas require additional attention. Feedback from the formative assessments was also used to provide teachers with data on the effectiveness of various instructional strategies used to promote math achievement (Fuchs et al., 1996) (ES = 0.304; value added ES = -0.229, neither ES reached statistical significance).

5) *Self monitoring.* The self-monitoring process was completed each time the formative assessment data suggested the need for an instructional change. Teachers answered a set of questions regarding their students' progress and their future instructional plans. Using the information provided by the twice-weekly formative assessment, teachers responded to questions such as, "On what skill(s) has the student done well in the preceding 2 weeks? On what skill(s) has the student improved compared to the previous 2-week period? What skill(s) should be

targeted for the coming 2-week period? How will the teacher attempt to improve student performance on the targeted skill(s)?" (Allinder et al., 2000, p. 5). (ES = 0.588, value added ES = 0.603, neither ES reached statistical significance.)

6) *Using formative assessment data as a means for teachers to provide specific instructional suggestions for small group instruction and computer-assisted instruction.* Each teachers' weekly report, based on the formative assessment data included the following: (a) content that needed to be taught or retaught during whole class instruction, (b) specific suggestions on how to break the class into small groups for small group instruction and which content to cover, (c) individualized computer-assisted problems for each student, and (d) suggested material to cover during peer tutoring sessions (Fuchs et al., 1994). (Effect sizes were 0.546 for low-achieving students, 0.292 for average achieving students and 0.172 for students with learning disabilities; none of these effect sizes were statistically significant.) Note that for value added, effect size was 0.453 for low achievers, -0.021 for average achievers, and 0.37 for students with learning disabilities. However, effect sizes were negligible as none of them reached statistical significance.

The overall picture provided by the data in Table 33 indicates that the set of enhancements are effective in enhancing students' mathematics achievement. The average effect size, in this case, is significant for studies conducted at the student level [average ES = 0.383, statistically significant, ES = 0.657, (ns) for the class level study]. As was the case for the first set of analyses, the same pattern emerges whether or not the full set of studies is included, or only those where an effect size could be computed at the student level.

Note that the effect of formative assessment with enhancements increases dramatically from formative assessment alone. It appears that the approaches that provided specific suggestions directly to the teacher about what to teach during small group instruction or partner work, or provided specific instructional suggestions worked better than this more indirect method.

Table 34 presents effect sizes for the value added by the enhancement. In other words, the comparison condition involves use of formative assessment only. As one would expect, with the more stringent criterion, the mean effect size is much lower, 0.194, which is not statistically significant. In two cases, the enhancement did not provide any additional gain in terms of student achievement to the mere use of formative assessment. On average the effect size doubles when an enhancement is added.

It is important to note that the majority of these studies focus on students with diagnosed learning disabilities. Only two samples (both from the same study) involve students from the general population. It is also important to note that these studies involve only one of two dependent measures, the Mathematics Operations Test, or the Mathematics Concepts and Applications Tests. These tests were developed by the researchers. However, they do possess solid psychometric properties.

The final caveat is that many of the studies involve special education teachers. Thus, one should be cautious in interpreting implications for classroom teachers because few of the enhancements involved only the classroom teacher.

E. Summary and Conclusions

The set of ten well-designed and well-executed studies on formative assessment demonstrates that regular use of formative assessments in mathematics can enhance students' mathematics achievement in the elementary grades, in both the areas of computation and concepts and applications. These studies were conducted with moderately large numbers of teachers in "real-world" settings; thus the external validity is high. The average effect size boost provided by use of formative assessments for studies conducted at the individual level is 0.206 and approaches significance. This corresponds to a boost of 9 percentile points.

In addition, the set of studies describes a set of tools and procedures (what the Task Group calls "enhancements") that can accompany formative assessment. These tools include specific activities that are linked to a student's current needs. Activities range from a list of ideas for alternate means of teaching the material, to specific materials for use in peer tutoring, to a listing of skills and concepts that require additional explanation and discussion.

Although many of the effect sizes doubled in value with these enhancements, the net contribution of 0.194 was not significant. Thus, the Task Group would more cautiously call these practices promising as opposed to evidence-based.

Two other issues need to be considered in framing specific recommendations for improving practice. The first is that the preponderance of studies were conducted at the elementary school. Second, to date, only one type of formative assessment has been studied with rigorous experimentation. These are assessments that include random sampling of items that address state standards. These assessments tend to take between 2 and 8 minutes to administer and thus are feasible for regular use. However, as discussed in the Introduction, many other types of formative assessments have been developed. The Task Group simply cannot comment on how useful these other types are in terms of enhancing students' performance at this point in time since the Task Group was unable to uncover any rigorous experiments involving their use. Hopefully, such research will be conducted in the near future.

F. Proposed Recommendations

Schools should seriously consider regular use of formative assessments in mathematics.

This might entail weekly assessments of students experiencing difficulties and less frequent (perhaps three times a year) assessments for others.

The Task Group would recommend use of formative assessments with known validity and reliability. However, the Task Group is aware that at the current point in time, there is a paucity of such measures. The Task Group advocates serious research and development in this area. It appears such work has already begun and federal support of this effort seems critical. In particular, validity studies of methods other than those that sample from annual state or district goals could and should be explored.

Research findings suggest that several enhancements can help teachers use formative assessment information more effectively.

Here, the research base is smaller, and less consistent. Several major tools appear to be promising. The first is linking formative assessment information (via technology) with specific recommendations for a teacher in areas such as a) content and concepts that require additional work with the majority of the class and b) specific activities that could and should be used by a given student for either tutoring or computer-assisted intervention or intervention work provided by an adult (teacher, mathematics specialist, or trained paraprofessional).

Use of formative assessments in mathematics can lead to increased precision in how time is used in class and assist teachers in providing appropriate instruction to students who need help on topics for which they need help. This should seriously be considered as districts consider the development and implementation of response-to-intervention models in mathematics.

G. Suggestions for Future Research

Several future research areas seem worth pursuing. The first is extending this line of research to the middle school and high school area. The second entails the same type of rigorous research studies of other, more clinical types of formative assessments such as those described in recent publications by NCTM.

The Task Group also needs to know more about how formative assessment measures relate to mathematics tests that include items that are more mathematically sophisticated than those on current standardized achievement tests. It also is important to update the studies of the reliability and validity of publisher-developed tests. That research is now over 20-years-old and the nature of mathematics instruction has changed dramatically.

BIBLIOGRAPHY

Allinder, R.M. (1996). When some is not better than none: Effects of differential implementation of curriculum-based measurement. *Exceptional Children, 62(6)*, 525–535.

Allinder, R.M., Bolling, R., Oats, R., & Gagnon, W.A. (2000). Effects of teacher self-monitoring on implementation of curriculum-based measurement and mathematics computation achievement of students with disabilities. *Remedial & Special Education, 21(4)*, 219–226.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: AERA Publications.

Bloom, B.S. (1976). *Human characteristics and school learning.* New York: McGraw-Hill.

Bloom, B.B. (1980). *All our children learning.* New York: McGraw-Hill.

Calhoon, M.B., & Fuchs, L. (2003). The effects of peer-assisted learning strategies and curriculum-based measurement on the mathematics performance of secondary students with disabilities. *Remedial and Special Education, 24(4)*, 235–245.

Chard, D., Clarke, B., Baker, B., Otterstedt, J., Braun, D., & Katz, R. (2005). Using measures of number sense to screen for difficulties in mathematics: Preliminary findings. *Assessment Issues in Special Education*, *30*(2), 3–14.

Clarke, B., & Shinn, M.R. (2004). A preliminary investigation into the identification and development of early mathematics curriculum-based measurement. *School Psychology Review, 33(2)*, 234–248.

Deno, S.L. (1986). Formative evaluation of individual student programs: A new role for school psychologists. *School Psychology Review*, *15(3)*, 358–374.

Deno, S. (2007, April). Distinguished Research Award. Presentation at the American Educational Research Association Annual Meeting. Chicago, IL.

Donovan, M.S., & Bransford, J.D. (2005). Introduction. In the National Research Council (Ed.), *How students learn: Mathematics in the classroom* (pp. 1–30). Washington, DC: National Academy Press.

Foegen, A. (2000). Technical adequacy of general outcome measure for middle school mathematics. *Diagnostique*, *25*, 175–203.

Foegen, A., Jiban, C., & Deno, S. (2007). Progress monitoring measures in mathematics: A review of the literature. *Journal of Special Education*, *41*, 121–139.

Freudenthal, H. (1973). *Mathematics as an educational task.* Dordrecht, Holland: Reidel.

Fuchs, D., Roberts, P.H., Fuchs, L.S., & Bowers, J. (1996). Reintegrating students with learning disabilities into mainstream: a two-year study. *Learning Disabilities Research and Practice*, *11*, 214–229.

Fuchs, L.S. (2004). The past, present, and future of curriculum-based measurement research. *School Psychology Review*, *33*, 188–192.

Fuchs, L.S., Deno, S.L., & Mirkin, P.K. (1984). The effects of frequent curriculum-based measurement and evaluation on pedagogy, student achievement, and student awareness of learning. *American Educational Research Journal, 21(2),* 449–460.

Fuchs, L.S., Fuchs, D., & Hamlett, C.L. (1989). Effects of alternative goal structures within curriculum-based measurement. *Exceptional Children 55(5),* 429–438.

Fuchs, L.S., Fuchs, D., Hamlett, C.L., & Stecker, P.M. (1990). The role of skills analysis in curriculum-based measurement in math. *School Psychology Review, 19(1),* 6–22.

Fuchs, L.S., Fuchs, D., Hamlett, C.L., & Stecker, P.M. (1991). Effects of curriculum-based measurement and consultation on teacher planning and student achievement in mathematics operations. *American Educational Research Journal, 28(3),* 617–641.

Fuchs, L.S., Fuchs, D., Hamlett, C.L., Thompson, A., Roberts, P.H., Kupek, P., et al.(1994). Technical features of a mathematics concepts and applications curriculum-based measurement system. *Diagnostique, 19(4),* 23–49.

Fuchs, L.S., Fuchs, D., Karns, K., Hamlett, C.L., & Katzaroff, M. (1999). Mathematics performance assessment in the classroom: Effects on teacher planning and student problem solving. *American Educational Research Journal, 36(3),* 609–646.

Fuchs, L.S., Tindal, G., & Fuchs, D. (1986). Effects of mastery learning procedures on student achievement. *Journal of Educational Research, 79(5),* 286–291.

Gersten, R., Keating, T.J., & Irvin, L.K. (1995). The burden of proof: Validity as improvement of instructional practice. *Exceptional Children, 61(6),* 510–519.

Glaser, R., & Silver, E. (1994). Assessment, testing and instruction: retrospect and prospect. *Review of Research in Education, 20,* 393–419.

Guskey, T.R. (1984). The influence of change in instructional effectiveness upon the affective characteristics of teachers. *American Educational Research Journal, 21(2),* 245–259.

Helwig, R., & Tindal, G. (2002). Using general outcome measures in mathematics to measure adequate yearly progress as mandated by Title I. *Assessment for Effective Intervention, 28(1)*, 9–18.

Jiban, C., Deno, S.L., & Foegen, A. (in press). Progress monitoring in mathematics: A review of the literature. *Exceptional Children*.

Lee, S.W. (2005). *Encyclopedia of school psychology*. Thousand Oaks, CA: Sage Publications, Inc.

Messick, S. (1988). The once and future of validity: Assessing the meaning and consequences of measurement. In H. Wainer & H. I. Braun (Eds.), *Test Validity* (pp. 33–46). Hillsdale, NJ: Erlbaum.

National Council of Teachers of Mathematics (NCTM) (1989). *Curriculum and evaluation standards for school mathematics.* Reston, VA: Author.

National Council of Teachers of Mathematics (NCTM). (2006). *Curriculum focal points for prekindergarten through Grade 8 mathematics: A quest for coherence.* Reston, VA: Author.

National Research Council. (2001). *Adding it up: Helping children learn mathematics*. In J. Kilpatrick, J. Swafford, and B. Findell (Eds.), Mathematics learning study committee, center for education, division of behavioral and social sciences, and education. Washington, DC: National Academy Press.

National Research Council (2005). *How students learn: Mathematics in the classroom.* Washington, DC: National Academy Press.

Renaissance Learning (1998). Accelerated Math [computer software]. Wisconsin Rapids, WI: Author.

Reys, B.J., Dingman, S., Sutter, A., & Teuscher, D. (2005) *Development of state-level mathematics curriculum documents: Report of a survey.* Columbia, MO: University of Missouri, Center for the Study of Mathematics Curriculum, 2005. Retrieved on February 14, 2006, from http://www.mathcurriculumcenter.org/resources/ASSMReport.pdf.

Ruiz-Primo, M.A., Shavelson, R.J., Hamilton, L., & Klein, S. (2002). On the evaluation of systemic science education reform: Searching for instructional sensitivity. *Journal of Research in Science Teaching, 39(5)*, 369–393.

Spicuzza, R., Ysseldyke, J., Lemkuil, A., Kosciolek, S., Boys, C., & Telluchsingh, E. (2001). Effects of curriculum-based monitoring on classroom instruction and math achievement. *Journal of School Psychology, 39(6)*, 521–542.

Spitzer, H.F. (1951). Learning and Teaching Arithmetic. In N.B. Henry, & G.T. Buswell, *The teaching of arithmetic—Fiftieth yearbook, part 11, National Society for the Study of Education*. Chicago: University of Chicago Press.

Sueltz, B.A., Boynton, H., & Sauble, I. (1946). The measurement of understanding in elementary-school mathematics. In *Measurement of understanding, 45th yearbook of the national society for the study of education, part I*. Chicago: University of Chicago Press.

Tindal, G., Fuchs, L.S., Fuchs, D., Shinn, M.R., Deno, S.L., & Germann, G. (1985). Empirical validation of criterion-referenced tests. *Journal of Educational Research, 78(4)*, 203–209.

VanDerHeyden, A.M., Witt, J.C., Naquin, G., & Noell, G. (2001). The reliability and validity of curriculum-based measurement readiness probes for kindergarten students. *School Psychology Review, 30(3),* 363–382.

Ysseldyke, J., Spicuzza, R., Kosciolek, S., Teelucksingh, E., Boys, C., & Lemkuil, A. (2003). Using a curriculum-based instructional management system to enhance math achievement in urban schools. *Journal of Education for Students Placed at Risk*, 8(2), 247–265.

Conclusion

Mathematics instruction is a complex professional practice. Researchers in the educational research community have made important forays into several of the most controversial and pressing questions about the effectiveness and impact of various types of instructional practice and, in particular, have conducted some studies that examine the effects of various interpretations and implementations of practices that have been advocated in the "reform" documents in mathematics education during the past two decades.

The question we asked is: *What can be learned from a review of the best available evidence in six important aspects of practice?* These practices included: the use of cooperative groups and peer instruction, the use of direct instruction with learning disabled students, the use of "real-world" problems in mathematics teaching, the use of technology, the enrichment and acceleration of instruction for mathematically precocious students, and the use of formative assessment.

For none of the areas examined did the Task Group find sufficiently strong and comprehensive bodies of research to support all-inclusive policy recommendations of any of the practices addressed. Nor did the Task Group find sufficient evidence to support policy recommendations favoring the status quo in mathematics teaching.

Across all of the areas, the Task Group found that **several instructional practices in mathematics teaching show some promise, in comparison to typical practice, for affecting student learning.** In each case the "promising" practice is clearly specified, somewhat prescriptive, and involves a mix, or combination, of particular, distinct practices. Thus, for example, it cannot be said that cooperative learning is a practice whose effectiveness is supported by research—but the Team Assisted Individualization (TAI) approach, with particular students in a particular area of mathematics, is effective. Although formative assessment to inform instruction is useful, it is enhanced when teachers use assessment tools with known validity and reliability. For students performing in the lower third of their grade level expectations, explicit instruction involving clear models of proficient performance, many opportunities to verbalize their problem solving strategies, and adequate practice and review should be a part of the mathematics program. It is not surprising that what the Task Group found about effective instructional practice is far more subtle and nuanced than direct answers to the starkly stated questions investigated.

The Task Group found some rather robust findings but they must be accompanied by a caveat. When a practice is demonstrated by high-quality experimental research to have some promise, it is critical to be clear about the promise "for what aspects of mathematics proficiency?" Different practices and approaches impact different kinds of outcomes, ranging from computational performance, to "real-world" problem solving, to identifying extraneous problem information, to long-term participation and interest in studying mathematics.

Because researchers and practitioners use different definitions to describe their interventions, it is conceptually problematic to place too much stock in either generalizing that a broad category of practice (e.g., using technology, or using "real-world" problems) has impact because a set of studies working on the same particular component of this category have impact, which was the case in some of the Task Group's reviews.

The Task Group's process included asking mathematicians and mathematics education reviewers to examine the mathematical content of the research studies—to look at the assessments and interventions, to the extent possible, based on the published reports. They expressed important concerns, including the possibility that an outcome measure item purported to measure computation might not do so because it really measured ability to use the context, for instance. They expressed concern that some topics were underdeveloped (i.e., failed to help students access the underlying mathematics in the topic covered), or that items were mislabeled (e.g., as "problem solving") when a mathematics expert might classify them otherwise. However, they also did note that several of the studies we reviewed seemed to help students increase their knowledge of mathematics and their ability to apply that knowledge to novel situations is a fashion that is valid from a mathematical perspective.

The reader may feel disappointed at this juncture, seeing how few robust findings emanated from a review of the rigorous research on the topics addressed. Yet even the inconclusive and limited findings can provide a real service to the profession. If an administrator, a curriculum developer or a parent comments, "Research says that lessons must start with 'real-world' problems," or "Students will really learn mathematics only if they are taught using direct instruction," consumers and professionals now know that research is inconclusive on these topics. This is a necessary step in the evolution of educational research into a more mature science. The paucity of findings and the paucity of high-quality experimental research in the field led the Task Group to realize, early on in the process, that few definitive answers to the research questions posed would be found.

However, the Task Group did see this work as the starting point for creating a base of knowledge to answer the questions posed at the onset of this work. We also see the application of the rigorous standards (developed in large part through earlier work of the Institute of Education Sciences of U.S. Department of Education) as serving as guidelines for the next generation of researchers.

The questions and topics studied and findings are briefly summarized below.

A. How Effective Is Teacher-Directed Instruction in Mathematics in Comparison to Student-Centered Approaches, Including Cooperative and Collaborative Groups, in Promoting Student Learning? Is One Approach Preferable to Another? If So, in Which Areas of Mathematics?

A controversial issue in the field of mathematics teaching and learning is whether classroom instruction should be more teacher-directed or student-centered. These terms have come to incorporate a wide array of meanings, with teacher-directed ranging from highly scripted direct instruction approaches to interactive lecture styles, and with student-centered ranging from students having primary responsibility for their own mathematics learning to highly structured cooperative groups. Schools and districts must make choices about curricular materials or instructional approaches that often seem more aligned with one instructional orientation than another. This leaves teachers wondering about when to organize their instruction one way or the other, whether certain topics are taught more effectively with one approach or another, and whether certain students benefit more from one approach than the other.

In the review, the Task Group limited the search to studies that directly compared these two extreme positions. Teacher-directed instruction was defined as instruction in which it is the teacher who is primarily communicating the mathematics to the students directly, and student-centered instruction as instruction in which primarily students are doing the teaching.

Only eight studies were found that met the Task Group's standards for quality that were consistent with this definition. The studies presented a mixed and inconclusive picture of the relative impact of these two forms of instruction. High-quality research does not support the contention that instruction should be either entirely "child-centered" or "teacher-directed." Research indicates that some forms of particular instructional practices can have a positive impact under specified conditions. All-encompassing recommendations that instruction should be entirely "child-centered" or "teacher-directed" are not supported by research. The limited research base of rigorous research does not support the exclusive use of either approach.

1. Cooperative and Collaborative Groups

One of the major shifts in education over the past 25–30 years has been advocacy for the increased use of cooperative learning groups and peer-to-peer learning (e.g., structured activities for students working in pairs) in the teaching and learning of mathematics.

Research has been conducted on a variety of cooperative learning approaches. One such approach, Team Assisted Individualization (TAI) has been shown to significantly improve students' computation skills. This instructional approach involves heterogeneous groups of students helping each other, individualized problems based on student performance on a diagnostic test, and rewards based on both group and individual performance. Effects on conceptual understanding and problem solving were not significant. There is evidence suggesting that working in dyads with a clear structure also improves computation skills in the elementary grades. However, additional research is needed.

B. What Is the Impact of Use of Formative Assessment in Mathematics Teaching?

Formative assessment—the ongoing monitoring of student learning to inform instruction—is generally considered a hallmark of effective instruction in any discipline. The Task Group's review of the high-quality studies of this topic produced several conclusions.

Teachers' regular use of formative assessment is marginally significant in improving their students' learning. This is especially true if teachers have additional guidance on using the assessment to design and individualize instruction.

Although the research base is smaller, and less consistent than that on the general effectiveness of formative assessment, the research does suggest that several specific tools and strategies can help teachers use formative assessment information more effectively. The first promising strategy is providing formative assessment information to teachers (via technology) on content and concepts that require additional work with the whole class. The second promising strategy involves using technology to specify activities needed by individual students. Both of these aids can be implemented via tutoring, computer-assisted instruction, or help provided by a professional (teacher, mathematics specialist, trained paraprofessional).

We caution that only one type of formative assessment has been studied with rigorous experimentation. These are assessments that include random sampling of items that address state standards. These assessments tend to take between 2 and 8 minutes to administer and thus are feasible for regular use.

The regular use of formative assessment particularly for students in the elementary grades is recommended. These assessments need to provide information not only on their content validity but also on their reliability and their criterion-related validity (i.e., correlation of these measures with other measures of mathematics proficiency). For struggling students, frequent (e.g., weekly or biweekly) use of these assessments appears optimal, so that instruction can be adapted based on student progress.

Research is needed regarding the content and criterion-related validity and reliability of other types of formative assessments (such as unit mastery tests included with many published mathematics programs, performance assessments, and dynamic assessments involving "think alouds"). This research should include studies of consequential validity (i.e., the impact they have on helping teachers improve their effectiveness).

Use of formative assessments in mathematics can lead to increased precision in how instructional time is used in class and can assist teachers in identifying specific instructional needs. Formative measures provide guidance as to the specific topics needed for assistance. Formative assessment should be an integral component of instructional practice in mathematics.

C. What Instructional Strategies for Teaching Mathematics to Students With Learning Disabilities and to Low-Achieving Students Show the Most Promise?

A review of 26 high-quality studies, mostly using randomized control designs, was conducted. These studies provide a great deal of guidance concerning some defining features of effective instructional approaches for students with learning disabilities (LD) as well as low-achieving (LA) students.

Explicit systematic instruction typically entails teachers explaining and demonstrating specific strategies, and allowing students many opportunities to ask and answer questions and to think aloud about the decisions they make while solving problems. It also entails careful sequencing of problems by the teacher or through instructional materials to highlight critical features. More recent forms of explicit systematic instruction have been developed with applications for these students. These developments reflect the infusion of research findings from cognitive psychology, with particular emphasis on automaticity and enhanced problem representation.

Our analysis of the body of research indicated that explicit methods of instruction are consistently and significantly effective with students with learning disabilities in computation, solving word problems, and solving problems that require the application of mathematics to novel situations.

Only a small number of studies were located that investigated the use of visual representations or student "think alouds." Therefore no inferences about their effectiveness can be drawn. The research suggests that they are most useful when they are integrated with explicit instruction.

Based on this admittedly small body of research, we conclude that students with learning disabilities and other students with learning problems should receive some time on a regular basis with explicit systematic instruction. There is no reason to believe that this type of instruction should comprise all the mathematics instruction these students receive. However, it does seem essential for building proficiency in both computation and the translation of word problems into appropriate mathematical equations and solutions. Some of this time should be dedicated to ensuring that students possess the foundational skills and conceptual knowledge necessary for understanding the mathematics they are learning at their grade level.

D. Do "Real-World" Problem Approaches to Mathematics Teaching and Efforts to Ensure that Students Can Solve 'Real-World' Problems, Lead to Better Mathematics Performance Than Other Approaches?

The meaning of the term "real-world" problem varies by mathematician, researcher, developer, and teacher. Conducting research in this area is complex; fidelity of the teachers' implementation of the instructional materials or instructional strategy is difficult to assess. Although not addressed in the studies we examined, teachers' knowledge and capacity to use such problems effectively varies greatly. Given these caveats, the Task Group addressed the question of whether using "real-world" contexts to introduce and teach mathematical topics and procedures is preferable to more typical instructional approaches.

The body of high-quality studies for this topic is small. Five studies addressed the question of whether the use of "real-world" problems as the instructional approach led to improved performance on outcome measures of ability to solve "real-world" problems, as well as on more traditional assessments. Four of these studies were similar enough to combine in a meta-analysis. The meta-analysis revealed that if mathematical ideas are taught using "real-world" contexts, then students' performance on assessments involving similar problems is improved. However, performance on assessments of other aspects of mathematics learning, such as computation, simple word problems, and equation solving, is not improved.

For certain populations (upper elementary and middle grade students and remedial ninth-graders) and for specific domains of mathematics (fraction computation, basic equation solving, and function representation), instruction that features the use of "real-world" contexts can have a positive impact on certain types of problem solving. Additional research is needed to explore the use of "real-world" problems in other mathematical domains, at other grade levels, and with varied definitions of "real-world" problems.

E. What Is the Relative Impact on Mathematics Learning When Students Use Technology Compared to Instruction That Does Not Use Technology?

1. Calculators

A review of 11 studies that met the Task Group's rigorous criteria (only one study was less than 20 years old) found limited to no impact of calculators on calculation skills, problem-solving, or conceptual development over periods of up to one year. Unfortunately, these studies cannot be used to judge the advantages or disadvantages of multiyear calculator use beginning in the early years, because such long-term use has not been adequately investigated. The Panel cautions that to the degree that calculators impede the development of automaticity, fluency in computation will be adversely affected.

2. Computer-Assisted Instruction and Computer Programming

We found that CAI drill and practice, if of high quality, can improve students' performance compared to conventional instruction, with the greatest effect on computation, and less effect on concepts and applications. Drill and practice programs **can** be considered as a useful tool in developing students' automaticity, or fast, accurate, and effortless performance on computation, freeing working memory so that attention can be directed to the more complicated aspects of complex tasks.

Research has demonstrated that tutorials (CAI programs, often combined with drill and practice) that are well designed and implemented can have a positive impact on mathematics performance, particularly at the middle and high school levels. CAI tutorials have been used effectively to introduce and teach new subject-matter content. However, these studies also suggest several important caveats. Care must be taken to ensure that there is evidence that the software to be used has been shown to increase learning in the specific domain and with students who are similar to those who are under consideration. Educators should critically inspect individual software packages and studies that evaluate them critically. Furthermore, the requisite support conditions to use the software effectively (sufficient hardware and software; technical support; adequate professional development, planning, and curriculum integration) should be in place, especially in large-scale implementations, to achieve optimal results.

Research indicates that computer programming improves students' performance compared to conventional instruction, with the greatest effects on understanding of concepts and applications, especially geometric concepts, and weaker effects on computation. However, computer programming by students can be employed in a wide variety of situations using distinct pedagogies, not all of which may be effective. Therefore, the findings are limited to the careful, targeted application of computer programming for learning used in the studies reviewed.

F. What Instructional Arrangements for Engaging with Mathematics Are Most Promising for Mathematically Precocious Students?

The Task Group's review of the literature about what kind of mathematics instruction would be most effective for gifted students focused on the impact of programs involving acceleration, enrichment, and the use of homogeneous grouping. The extensive literature searches we conducted yielded few studies that met the Task Group's methodologically rigorous criteria for inclusion. Thus for this topic—and this topic only—we relaxed these criteria in order to fulfill our charge of evaluating the "best available scientific evidence." One randomized control trial study and seven quasi-experimental studies were located. All but one of these studies have limitations.

Despite the flaws in any one study, the set of studies suggests there is value to differentiating the mathematics curriculum for students who are gifted in mathematics and possess sufficient motivation, especially when acceleration is a component (i.e., pace and level of instruction are adjusted). A small number of studies suggest that individualized instruction, in which pace of learning is increased and often managed via computer instruction, produces gains in learning.

Gifted students who are accelerated by other means not only gained time and reached educational milestones earlier (e.g., college entrance) but appear to achieve at levels at least comparable to those of their equally able same-age peers on a variety of indicators even though they were younger when demonstrating their performance on the various achievement benchmarks. One study suggests that gifted students also appear to become more strongly engaged in science, technology, engineering, or mathematical areas of study.

Some support also was found for supplemental enrichment programs. Of the two programs analyzed, one explicitly utilized acceleration as a program component and the other did not. Self-paced instruction supplemented with enrichment seemed to have a positive impact on student achievement. This supports the view in the field of gifted education that acceleration and enrichment combined should be the intervention of choice. We believe it is important for school policies to support appropriately challenging work in mathematics for gifted and talented students.

G. What Would the Instructional Practices Task Group Say to the Practitioner?

There is no one ideal approach to teaching mathematics; the students, the mathematical goals, the teacher's background and strengths, and the instructional context, all matter. The findings here do suggest that it is especially important:

- to monitor what students understand and are able to do mathematically;
- to design instruction that responds to students' strengths and weaknesses, based on research when it is available; and
- to employ instructional approaches and tools that are best suited to the mathematical goals, recognizing that a deliberate and conscious mix of strategies will be needed.

Also, it is important for teachers, school administrators, and the public to understand the importance of helping to formulate research questions and being willing to participate in the types of experimental and quasi-experimental studies that are described here.

H. What Would the Instructional Practices Task Group Say to the Researcher?

More research that can identify causal claims is needed to guide both policy and practice. Building the mathematics education research portfolio to include this work will involve:

- Formulation of research questions that are of interest to practitioners and policymakers;
- Collaborations among mathematicians, mathematics education researchers, methodologists, and psychometricians; and
- Motivation to design and undertake rigorous studies.

The work of this Task Group has substantiated our understanding of the complexity and challenge of effective mathematics instruction. It is now up to practitioners, policymakers, mathematicians, and mathematics education researchers to take up the challenges of clarifying the definitions of mathematics instructional practices, debunking myths about mathematics instruction, and formulating the types of research studies that can answer the pressing questions that need to be addressed.

APPENDIX A: Methodological Procedures

Methodology for the Instructional Practices Task Group Research Reviews

From the onset, the Instructional Practices Task Group was committed to assembling the most rigorous scientific research addressing questions of effectiveness about the types of interactions that occur in mathematics classrooms relative to student performance. The Task Group was aware that there might be a paucity of such studies. This issue of understanding the quality of evidence and design needed to lead to causal inference is discussed in the *Standards of Evidence* document approved by the National Mathematics Advisory Panel. However, it is particularly germane to this topic, in that before requiring widespread implementation of a particular instructional practice or intervention, or committing significant resources toward such implementation, it seems critical to know that it will in all likelihood lead to higher levels of mathematics proficiency than alternatives. Of the six topics investigated as part of the Panel report, the topic of instructional practices was the topic for which the most experimental research was available. The Task Group thus chose to review and synthesize only the highest quality experimental and quasi-experimental research, research that can lead to causal inference, as the primary goal. In some cases, the Task Group also relied on the best available evidence suitable to a particular issue.

The recent report by the National Research Council (NRC), *Scientific Research in Education* (2002), was influential in the decision. The authors note, "[They] believe that attention to the development and systematic testing of theories and conjectures across multiple studies and using multiple methods—a key scientific principle … is currently undervalued in education relative to other scientific fields" (p. 124). They go on to note, "While large-scale education policies and programs are constantly undertaken… they are typically launched without an adequate evidentiary base to inform their development, implementation or refinement over time…" (p. 124). The report also states: "Randomized experiments are not perfect.… For instance, they typically test complex causal hypotheses, they may lack generalizability to other studies, and they can be expensive. *However, we believe that these and other issues do not generate a compelling rationale against their use in education research and that issues related to ethical concerns, political obstacles and other potential barriers often can be resolved"* (p. 125, emphasis added). Whereas the field of reading instruction has made great strides through a combination of randomized controlled trials (RCTs), longitudinal research, descriptive research and qualitative research, there is less of a history in mathematics education research of using RCTs until recently.

Selection of Topics

The original members of the Task Group[16] devoted the first two meetings to decisions about the array of topics to study. The Task Group followed a process similar to that used by the National Reading Panel. After extended brainstorming and discussion, a list of approximately 20 topics was developed and then each member selected the top six.

No particular theoretical framework was used to generate this list. Panelists selected topics that were perceived as: a) high interest to the teachers and policymakers, b) areas requiring additional attention in terms of implementation of recent federal policies such as No Child Left Behind (NCLB) and Individuals with Disabilities Act (IDEA) of 2004, or c) topics deemed critical by organizations such as National Council of Teachers of Mathematics (NCTM).

In addition, based on cumulative knowledge of the research literature, the Task Group wanted to include at least one or two topics for which adequate research was available to provide empirically based recommendations. This seemed particularly important because three recent reports by the National Research Council (2001, 2002, 2004)[17] noted an extremely limited amount of rigorous research in this field; too few studies were available to draw causal inferences.

This resulted in a list of 12 topics. Due to time constraints, the Task Group was unable to address all of the 12 topics. The following eight received the most support:

1) “Real-world” problem solving
2) Relative effectiveness of explicit or teacher-centered instruction vs. child-centered or inquiry based instruction
3) Formative assessment
4) Cooperative, collaborative learning and peer-assisted instruction
5) Instructional strategies for students with learning disabilities
6) Instructional strategies for low-performing students
7) Instructional strategies for mathematically precocious students
8) Technology with a particular focus on use of graphing calculators and single function calculators

Among the topics that generated a good deal of interest, but were excluded due to time constraints, were: a) importance of time spent engaged in mathematics, b) guidelines for developing homework assignments, c) best practices in terms of review of previously taught material, and d) types of practice problems and the sequencing of practices.

[16] One of the original members of the Task Group (Diane Jones) left when she was assigned to another position in the federal government; she was replaced by Irma Arispe in June, 2007. In April, 2007, Bert Fristedt and Douglas Clements joined the group and Joan Ferrini-Mundy replaced Kathie Olsen in January 2007. By that point, most of the topics had been finalized, although three topics were subsequently eliminated.
[17] *Adding It Up (2001), Scientific Research in Education (2002), On Evaluating Curricular Effectiveness (2004).*

The topic of curriculum and instructional materials was assigned to a small task force to preclude even the appearance of conflict of interest, upon advice from the ethics attorney of the U.S. Department of Education.

Instructional Practices Task Group Methodology Statement

The Instructional Practices Task Group organized the available scientific evidence into several categories of evidence for consideration as they reviewed studies related to each topic of mathematics instruction investigated. A discussion of the categories for studies with quantitative designs that were considered for inclusion as rigorous evidence is provided below followed by a discussion of the procedures for identifying relevant research and synthesizing the research. Any deviations from the general practices outlined below are specified in the individual sections of the report.

Category 1: Experimental and Quasi-Experimental Studies that Meet or Meet with Reservations What Works Clearinghouse (WWC) Standards

Studies in this category provide evidence of causal claims and include randomized control trials (RCT; use random assignment to create experimental groups) and strong quasi-experimental studies (QED; experimental groups created by a method other than random assignment) that meet WWC Criteria. These criteria can be found at http://ies.ed.gov/ncee/wwc/overview/review.asp?ag=pi and http://ies.ed.gov/ncee/wwc/twp.asp.

The only cases where exceptions to WWC criteria were allowed are:

- Differential attrition rates of up to 30% are permitted for RCTs and for QEDs if there is evidence that attrition does not affect the nature of the sample on a salient pretest variable.
- Studies that assign only one school per condition are acceptable provided that there are several teachers per condition.

In all other areas, the Task Group followed the WWC policies expressed on www.whatworks.org. QEDs were excluded if they fail to either provide evidence of pretest comparability or control for pretest differences. Thus, the Task Group downgraded RCTs and exclude QEDs if a) there is evidence of contamination; b) there is only one teacher per experimental condition. However, if both the treatment and control conditions were taught by the same teacher, these were reviewed on a case-by-case basis, and the study may have been included is there was reason to believe that there was no bias in delivery.

Category 1 studies are the core of the results section for the Instructional Practices Task Group as they represent clear evidence to support causal claims. Category 1 studies correspond to high and moderate quality studies, as defined by the National Mathematics Advisory Panel Guidelines for Standards of Evidence.

Category 2: Weak Group Comparison Studies and Other Quantitative Designs that Attempt to Infer Causality

Category 2 consisted of weak group comparison studies (failed RCTs and weak nonequivalent comparison designs). Category 2 studies are always open to multiple interpretations with regard to causal inferences, however, they are not necessarily weak studies for other purposes (e.g., descriptive). Category 2 studies correspond to moderate and low-quality studies, as defined by the National Mathematics Advisory Panel Guidelines for Standards of Evidence.

Category 2: Weak Group Comparison Studies. These are attempts at experiments or quasi-experiments that are seriously flawed (e.g., one teacher per condition, widely differential attrition across the experimental groups, quasi-experiments with no evidence of pretest equivalence). This category would also include studies in which all the dependent measures are closely aligned to the instructional content of the intervention and not at all covered in the control condition. These studies are considered biased. Studies in which the experimental sample consists only of volunteers and the control group only of those who declined to participate would also be considered a weak comparison study. Because of the serious nature of the flaws, the Task Group would *not* consider these as providing valid causal evidence.

Category 2 studies are used only when there is an insufficient body of information from the evidence provided by Category 1-level studies. Flawed studies can never compensate for high-quality experimental or quasi-experimental studies. However, if there are no acceptable experimental studies, the report may include brief discussion of Category 2 studies. If there is a pattern of findings across the studies—and if the design flaws that compromise the studies are dissimilar (e.g., one study has differential attrition, another compares volunteers to non-volunteers)—the report may indicate that a pattern emerges that might be considered worthy of mentioning. Studies in this category, however, are highly variable in the nature of their flaws and will be assessed case by case by two Panelists and a researcher at Abt Associates before being used for this purpose.

These two categories are studies that attempt to determine causal inference. However, panelists were free to use any type of research (descriptive, correlational, qualitative) to set the context for their meta-analysis. The reader will note that all of these types of research have been used to help explain the concepts examined in the chapter, and to help interpret findings from the experiments. However, these studies were not used to make claims of causality or effectiveness.

Procedures

Literature search and study inclusion

Literature searches were conducted to locate studies on evidence-based practices and learning in mathematics. Electronic searches were made in PsycInfo and the Social Sciences Citation Index (SSCI) using search terms identified by the Instructional Practices Task Group. A full list of the search terms used follows on page 206. A total of 1,733 studies[18] were identified based on these search terms. The identification of studies using formative assessments was based on work conducted by the Urban Institute and is described in Appendix B. Any other deviations from this general literature search procedure are specified in the report for each topic. Additional studies were identified through manual searches of relevant journals, reference lists, and recommendations from experts. Abstracts from these searches were screened for relevance to research questions and appropriate study design. For each of the 381 studies that met the screening criteria, the full study report was examined to determine whether it met the inclusion criteria specified below. Additionally, citations from relevant articles and research syntheses in each of the areas were reviewed to identify additional candidate studies.

Criteria for Inclusion:

- Study was published between 1976 and 2007.
- Study involved K–12 students studying mathematics through algebra.
- Study was available in English.
- Study was published in peer-reviewed journal or government report.
- Study design was (a) a randomized experiment or (b) a quasi-experiment with techniques to control for bias (matching, statistical control) or demonstration of initial equivalence on a salient pretest variable.
- The study included at least two classrooms per condition if the intervention was performed at the classroom level. In cases where a single teacher or investigator administered both the treatment and control, one classroom may have been sufficient if there was evidence that no bias existed.
- The intervention was not confounded with teacher, instructional time, or any other variable. Studies with potential confounds were reviewed on a case-by-case basis.
- There was no evidence of contamination (i.e., that control group teachers were using experimental curriculum or ideas from the experimental curriculum).
- Attrition was less than 30% or evidence showed that the remaining sample was equivalent to the original sample on a salient variable.

Effect size calculations

For all studies that met the criteria for inclusion, the Panel applied the WWC guidelines to calculate standardized mean differences in mathematics achievement (see http://www.whatworks.ed.gov/reviewprocess/conducted_computations.pdf). Using *Comprehensive Meta-Analysis, Version 2,* software, Hedges g standardized mean

[18] This number does not include studies that were identified from searches using combinations of terms that led to hundreds of largely irrelevant citations, studies that were identified from manual reviews, or studies that were recommended from experts.

differences were calculated for each of the studies. The standardized mean difference is defined as the difference between the mean score for the treatment group minus the mean score for the comparison group, divided by the pooled standard deviation of that outcome for both the treatment and comparison groups.

For all quasi-experiments and for randomized controlled trials that showed differences in pretest scores at baseline, the effect size measure was calculated as an adjusted mean difference as per WWC guidelines. Specifically, whenever possible, the numerator in the effect size was calculated as the difference between the posttest means of the treatment and control groups minus the difference in the pretest means for those groups, divided by the pooled unadjusted between-student standard deviation on the posttest.

In cases in which schools, teachers, or classrooms were assigned (either randomly or nonrandomly) into intervention and comparison groups and the unit of assignment was not the same as the unit of analysis, the effect size and accompanying standard error were adjusted for clustering within schools, teachers, or classrooms. This analysis used WWC guidelines to adjust for clustering,[19] applying an intraclass correlation (ICC) adjustment of 0.20 when actual ICC values were unavailable, which is the default ICC for achievement outcomes recommended by the WWC.

Pooling effect sizes across study samples

When judged appropriate, the Task Group pooled effect sizes across studies meta-analytically using random effects models. Specifically, weighted mean effect sizes were computed using inverse variance weights to reflect the statistical precision of the respective studies stemming from both the subject-level and study-level sampling error.

Multiple contrasts: For each study that included at least three conditions, effect sizes were calculated for all relevant contrasts, provided that they were orthogonal. When pooling the effects using meta-analytic techniques, only independent effect sizes per study were included, i.e., those not based on the same participant samples.

Multiple outcomes: For studies that reported effects on more than one mathematics achievement outcome, Panel reviewers decided either to choose one outcome or to average the results from multiple outcomes on a case-by-case basis. Assessments that were overly aligned with an intervention were either not used or noted when used.

Multiple independent samples within a study: In cases where impacts on independent samples within a study were reported, all independent effect sizes were included separately in the pooled analysis.

[19] See http://ies.ed.gov/ncee/wwc/pdf/rating-scheme.pdf for more information on this issue.

Study Identification Procedure for Formative Assessment

Studies that were included in the formative assessment analysis were based primarily on the literature search conducted by the Urban Institute as part of the U.S. Department of Education's Promising Practices Initiative (Olsen, 2006). The Urban Institute study inclusion criteria focused on issues of relevance, appropriate research methods, and adequate reporting of program effects (See Olsen, 2006, for further detail). The Urban Institute research staff identified relevant studies by gathering studies recommended by content experts at the U.S. Department of Education's Center on Instruction, examining reference lists, conducting database searches in Google Scholar, and searching through a dissertation database. Their efforts yielded 92 potentially relevant studies for consideration. Of these, nine studies met the criteria for inclusion in the Urban Institute's meta-analysis. Reasons for exclusion were: qualitative studies, quantitative studies with no measure of program impacts, no relevance to formative assessment, formative assessment unrelated to mathematics. In addition, studies were excluded if they provided insufficient data to calculate effect sizes. The criteria for the search were virtually identical to WWC except that standards for differential attrition, and confounding of intervention with school were not as rigorous (See Methodology report for further discussion). In the search, the keyword string included was ("mathematics" OR "math") ("formative assessment" OR "curriculum-based measurement") ("estimate" OR "coefficient" OR "correlation") ("student achievement" OR "teacher use") ("random" OR "random assignment" OR "randomly assigned" OR "matched"). In order to identify dissertations in the area of formative assessment in mathematics using Proquest's Digital Dissertation database, the keyword string included ("formative assessment" OR "curriculum based measurement" OR "ongoing assessment").

The Urban Institute put "on hold" studies that measured the effect of various enhancements to the formative assessments. The Task Group viewed the "enhancements" as important for understanding best ways for teachers and school districts to use formative assessments.

The National Mathematics Panel retrieved and reviewed all studies that had been excluded by the Urban Institute but were coded as quantitative with a comparison group. As a result, two additional studies were added to the Panel review: one that studied the effect of an enhanced formative assessment program against a control group (Calhoon & Fuchs, 2003), and another where it was possible to estimate a student-level sample size and thus calculate effect sizes and standard errors (Allinder, Bolling, Oats, & Gagnon, 2000).

In addition, after reviewing the nine original studies included by the Urban Institute, the Task Group determined that two of the papers (Spicuzza, Ysseldyke, Lemkuil, Kosciolek, Boys, & Telluchsingh, 2001; Ysseldyke, Spicuzza, Kosciolek, Teelucksingh, Boys, & Lemkuil, 2003) reported usable data based on the same study and sample. As a result, the Panel analysis includes a total of ten studies.

Search Terms Used for Instructional Practices Task Group By Research Question

All of the terms in the lists below were searched with the term *math**

Teacher-Directed and Student-Centered Instruction

active instruction
active teaching
CGI
cognitively guided instruction
constructivist
cumulative review
direct instruction
discovery learning
drill
explicit instruction
guided inquiry
guided learning
learner centered
student directed strategies
student explanations
student feedback
student reasoning
teacher centered instruction
teacher demonstration
teacher-directed instruction
teacher-directed strategies
teacher explanations
teacher feedback
teacher led instruction
teacher modeling

Additional Searches Specifically for Cooperative Learning

classwide peer tutoring
collaboration
cooperative learning
cooperative mastery learning
peer assisted learning
peer tutoring

Real World

aligning everyday and mathematical reasoning
anchored instruction
applications project
applied curricul*
applied problems
Arise
authentic
case-based
complex mathematical tasks
Connected Math
contextual curricul*
contextual problems
Core Plus
effectiveness of real world problem solving
engagement potential
everyday reasoning
Freudenthal Institute
integrated mathematics curricul*
interactive mathematics program
interactive mathematics project
Jasper
math in context
mathematical complexity
mathematical modeling
mathematical reasoning
mathematical word problems
mathematization
Middle School Math*
modeling curricul*
modeling our world
multiple solution paths
PISA
problem-based curricul*
problem-based learning
realistic math*
real-life mathematical problem solving
real world problems
SimCalc
simulations
situated cognition
solution paths
solving word problems
video

Students With Learning Disabilities, Low-Achieving Students, and English Language Learners

academically disadvantaged
anchored instruction
at-risk
classroom practices
cognitive strategy instruction
cooperative learning
curricula adaptation
differentiated instruction
direct instruction
dyscalculia
elementary education
English as a second language
English language learners
heterogeneous group
instructional design
instructional practice
intervention
learning disabil*
limited English proficient
low achiev*
math disability
math dyslexia
peer assisted learning
problem solving strategy
reform curricula
school-based intervention
secondary education
slow learners
teaching methods

Gifted Students

acceleration
developmental placement
differentiated curriculum
differentiated instruction
differentiation
enrichment
exceptional
gifted
grouping
high achiev*
talent*

Technology

artificial intelligence
CAI
calculator*
calculator-based ranger
CampOS
CBL
cellular
computer manipulatives
computer*
computer-assisted instruction
development
education
electronic blackboard
enhanced anchored instruction
graphing calculator
handheld
hypermedia
instruction
instructional tools
interactive whiteboard
interactive*
internet
learning
Logo
PDA
pedagogy
portable
programming
screen projection
screen-based technology
smart board
software
spreadsheet
teaching
technology
turtle graphics
tutor*
virtual manipulatives
visual representation
web-based

BIBLIOGRAPHY

Allinder, R.M., Bolling, R., Oats, R., & Gagnon, W.A. (2000). Effects of teacher self-monitoring on implementation of curriculum-based measurement and mathematics computation achievement of students with disabilities. *Remedial & Special Education*, *21(4)*, 219–226.

Calhoon, M.B., & Fuchs, L. (2003). The effects of peer-assisted learning strategies and curriculum-based measurement on the mathematics performance of secondary students with disabilities. *Remedial and Special Education, 24(4)*, 235–245.

National Reading Panel (2000). *Teaching children to read.* Washington, DC: U.S. Department of Education.

National Research Council (2004). *On Evaluating Curricular Effectiveness.* Washington, DC: National Academy Press.

National Research Council (2002). *Scientific research in education*. Washington, DC: National Academy Press.

National Research Council. (2001). *Adding it up: Helping children learn mathematics*. In J. Kilpatrick, J. Swafford, & B. Findell (Eds.), Mathematics learning study committee, center for education, division of behavioral and social sciences, and education. Washington, DC: National Academy Press.

Spicuzza, R., Ysseldyke, J., Lemkuil, A., Kosciolek, S., Boys, C., & Telluchsingh, E. (2001). Effects of curriculum-based monitoring on classroom instruction and math achievement. *Journal of School Psychology, 39(6)*, 521–542.

Ysseldyke, J., Spicuzza, R., Kosciolek, S., Teelucksingh, E., Boys, C., & Lemkuil, A. (2003). Using a curriculum-based instructional management system to enhance math achievement in urban schools. *Journal of Education for Students Placed at Risk*, 8(2), 247–265.

APPENDIX B: Research Questions

This appendix lists only topics that fall within the areas that we addressed in our literature review. We recognize that there is much needed research on other topics about instructional practice, such as the teaching of specific mathematical topics and content (e.g., fractions).

For future studies on teacher-directed versus student-centered instruction, our major suggestions include:

- Studies that further unpack the underlying variables behind terms such as student-centered, guided inquiry, teacher-centered, direct instruction, and explicit instruction. These studies should entail a strong classroom observational component.

- Studies that describe and evaluate the impact of: 1) various models of teaching, 2) explicit instruction for specific topics, and, 3) the use of visual representations with manipulatives to link abstract concepts (i.e. equations, algorithms).

- Studies that use curricula that are deemed to be mathematically accurate and rigorous and detail teaching practices that enhance understanding.

- Both qualitative and correlational studies of classes with exceptionally high student growth in mathematics to provide deeper insights into the nature of effective practice.

For future studies on the role of technology, our major suggestions include:

- Improved measures and analyses of fidelity of Computer Based Instruction (CBI) to best ascertain the effectiveness of interventions and to reveal "true" effect sizes that are the result of high-quality interventions. Research must also reveal what actions support high-quality large-scale implementation of these interventions.

- Studies that illuminate the particular cognitive and learning processes that different categories of software do or do not support.

- The linking of CBI features to student outcomes so that software engineers and curriculum designers can improve the use of technology in the school setting.

- The role of additional contextual variables (e.g., settings, such as urban, suburban, or rural and student or family characteristics), and implementation variables (e.g., duration, support and availability of resources, funds, and time) should be conscientiously addressed in future research.

- The initiation of longitudinal studies that will assess whether the consistent use of computer-based tools, including computer programming, can benefit learning and improve student skills.

- The implications of technology for the content of mathematics education must be adequately addressed philosophically, theoretically, and empirically.

For future studies on the role of calculators, our major suggestions include:

- The contextualization of advances in technology, curricula, and pedagogical strategies within research that examines the benefits of using calculators. This research should standardize the use of graphing calculators to address education research questions.
- An examination of whether appropriate pedagogical uses reinforce, or at least maintain, students' learning of basic arithmetical facts and properties *while simultaneously* garnering educational advantages. The fidelity of these approaches should be evaluated alongside student outcomes.
- An exploration of the cognitive processes that students use (e.g., in assessment situations) when calculators are available and the ramifications that these findings have for instruction and assessment with calculators.
- An investigation as to why about two-thirds of algebra teachers use graphing calculators infrequently. Are there practical barriers to their use, does curriculum and professional development discourage their use, and do student experiences convince teachers they are not useful? Would the provision of resources and professional development change this situation?

For future studies about instructional practices with low-achieving students and students with learning disabilities:

- Studies of the issues discussed above that focus particularly on impact with students who experience difficulty in mathematics.
- Studies to determine the amount of additional practice with feedback that these students require, the amount of highly systematic instruction needed, and the areas in which this instruction is required needs to be determined.
- Studies that examine how various approaches that are linked to specific mathematical topics are needed.

For future studies about instructional practices with gifted and mathematically precocious students:

- Evaluations of academically rigorous enrichment programs.
- Explorations of the extent to which effective enrichment programs are, in fact, acceleration programs. As students explore the mathematics that underlie their current work, the enrichment activities can develop skills in more advanced areas of mathematics, areas that the student may not cover in a formal sense for several more years.
- Longitudinal studies examining career choices and persistence in mathematics for mathematically gifted students who have participated in various intervention programs.

For future studies about the use of “real-world” problems in mathematics instruction:

- Studies to examine, describe, and clarify the multiple definitions of “real-world” problems and “real-world” problem-based instruction, and to relate those definitions to the interventions and to student learning.

- Development of valid and reliable outcome measures that clearly distinguish what is being assessed (mathematical concepts, mathematical procedures, problem solving, etc.).

- Studies that explore the possibly differential impact of “real-world” approaches to instruction for specific mathematical topics and concepts.

- Studies to examine the nature of the impact of “real-world” problem instructional approaches on student motivation and interest in mathematics, for different student groups.

APPENDIX C: Additional Technology Tables

Table C-1: Results from Prior Meta-Analyses on Drill and Practice

Study	**Pooled effect size**			
	Achievement	*Attitudes*	*Study information*	
Hamilton, 1995	.19			
Burns & Bozeman, 1981	.34			
Hartley, 1978	.34			
Lee, 1990	.35	.23		
Slavin et al., 2007 (one ES only, not "pooled")	.36			
Kuchler, 1999	.51		Junior/Senior students	
Specific education goals	*Computation*	*Concept development*	*Applications*	*Combination of goals*
Burns, 1981, cited in Kuchler, 1999	.38	.18	.14	
Lee, 1990	.45	.19 (concepts & applications)		.28
Contextual variables				
Age/grade level	*Preschool*	*Elementary*	*Middle/Junior*	*Junior/Senior*
Burns & Bozeman, 1981		.35		.24
Lee, 1990		.34	.41	
Hamilton, 1995		.17 (Grades 0–6)		.25 (Grades 9–12)
Ability level	*Low*	*Average*	*Average*	*High*
Burns & Bozeman, 1981	.31	.14	.47 (high/avg)	.32
Lee, 1990	.36	.16		.16
Hamilton, 1995	.12 (very low)	.57 (low/avg)	-.04 (average)	.27
Gender	*Males*	*Females*		
Burns & Bozeman, 1981	.42	.17		
Lee, 1990	.31	-.06		
Hamilton, 1995	.26	.14		
Implementation variables				
Duration	*1–18 weeks*	*19–36 weeks*	*37+ weeks*	
Lee, 1990	.44	.25	.46	
Substitute vs. supplement	*Substitute*	*Supplement*		
Lee, 1990	.57	.33		
Developer	*Experimenter /teacher*		*Commercial*	
Lee, 1990	.42		.34	

Table C-2: Results from Prior Meta-Analyses on Tutorials

Study	**Pooled effect size**			
Overall	*Achievement*	*Attitudes*	*Study information*	
Hamilton, 1995	.20			
Lou et al., 2001	.20			
Kuchler, 1999	.25		Secondary school	
Hartley, 1978	.34			
Kulik, 1994	.38			
Kulik, 2003	.38[b]		Study focused on ILS	
Becker, 1992	.40		Study focused on ILS	
Burns & Bozeman, 1981	.45			
Lee, 1990	.55	.02[a]		
Specific education goals	*Computation*	*Concept & Applications*		*Combination*
Lee, 1990	.42	.63		.62
Topic	*Arithmetic*	*Geometry*	*Algebra*	*General*
Lee, 1990	.47	.47	1.19	.41
Contextual variables				
Age/grade level	*Preschool*	*Elementary*	*Middle/Junior*	*Secondary*
Hartley, 1978		.66 (Grades K–8)		.58 (Grades 9–12)
Lee, 1990		.49	.85	
Hamilton, 1995		.14 (Grades 0–6)		.29 (Grades 9–12)
Burns & Bozeman, 1981		.43		.52
Ability level	*Low*	*Average*	*Average*	*High*
Burns & Bozeman, 1981	.57	.58		.28
Lee, 1990	.62	.20		.18
Hamilton, 1995	.12	.08 (low/avg)	.02	.57
Gender	*Males*	*Females*		
Lee, 1990	.82 [a]	1.58 [a]		
Hamilton, 1995	.58 [c]	.14 [c]		
Implementation variables				
Duration				
	1–18 weeks	*19–36 weeks*	*37+ weeks*	
Lee, 1990	.55	.53	.57	
Substitute vs. supplement	*Substitute*	*Supplement*		
Lee, 1990 (achievement)	.30	.58		
Lee, 1990 (problem solving)	.09	.25		
Developer	*Experimenter / teacher*	*Commercial*	*Both*	
Lee, 1990	.62	.39	.58	
Audience	*Specific*	*General*		
Lee, 1990	.58	.29		

[a] Only two effect sizes were included.

[b] Pooled effect size was .40 when the ILS instruction was in mathematics only, and .17 when it was in both mathematics and reading.

[c] Only three effect sizes were included.

Table C-3: Results from Prior Meta-Analyses on Calculators

Study	**Pooled effect size**			
	Achievement	*Attitudes*	*Study information*	
Ellington, 2003a[a]		0.20[b]		
Hembree, 1984; Hembree & Dessart, 1986		0.190		
Hembree, 1992	0.29			
Smith, 1997		0.3655		
	Calculator type			
	Basic/scientific	*Graphing*	*All*	
Ellington, 2003a (operational skills, testing with calculators)	0.55	0.40	0.25	
Ellington, 2003a (conceptual skills, testing with calculators)	0.13	0.69		
Ellington, 2003a (problem solving skills, testing with calculators)	0.23	0.61 [3]		
	Specific education goals			
	Operational	*Computational*	*Conceptual*	*Problem solving*
Ellington, 2003a[a] (testing without calculators)	0.14[b]	-0.02[b]	-0.05[b]	0.16
Ellington, 2003a, [a] (testing with calculators)	0.32[b]	0.41[b]	0.44[b]	0.22[b]
Hembree, 1984; Hembree & Dessart (1986) (testing with calculators)		0.636		
Hembree, 1984; Hembree & Dessart, 1986 (testing without calculators)			0.018	
Hembree, 1984 (special calculator instruction, testing without calculators)	0.798	0.564	-0.268	0.534
Smith, 1997		0.2054	0.1972	0.1468
Smith, 1997 (Graphing calculators, graphing skills)	-0.523		(concept development)	
Contextual variables				
Age/grade level	*Preschool*	*Elementary*	*Middle/Junior*	*Secondary*
Ellington, 2003)[a] (conceptual skills, testing without calculators)		-0.06	0.52[2]	-0.15[2]
Ellington, 2003a[a] (operational skills, testing with calculators)		0.48[1]	0.57	0.32
Ellington, 2003a[a] (conceptual skills, testing with calculators)		-0.14[2]	0.70	0.43

Continued on p. 6-218

Table C-3, continued

Study	**Pooled effect size**			
Ability level	*Low*	*Average*	*High*	*Mixed*
Ellington, 2003a[a] (operational skills, testing with calculators)			0.69[e]	0.35
Ellington, 2003a[a] (conceptual skills, testing with calculators)			0.84	0.29
Ellington, 2003a[a] (problem solving skills, testing with calculators)	-0.18[c]		0.15[c]	0.43
Hembree, 1984; Hembree & Dessart, 1986 (operational skills, testing without calculators)	-0.107		-0.031	
Hembree, 1984 [Hembree & Dessart, 1986 (operational skills, testing with calculators)]	0.325	0.737		
Hembree, 1984; Hembree & Dessart, 1986 (computation skills, testing without calculators)	-0.009		-0.024	
Hembree, 1984; Hembree & Dessart, 1986 (problem solving composite skills, testing without calculators)	0.005		-0.118	
Hembree, 1984; Hembree & Dessart, 1986 (problem solving composite skills, testing with calculators)	0.436	0.271	0.458	
Implementation variables				
Duration	*0–3 weeks*	*4–8 weeks*	*9+ weeks*	
Ellington, 2003a[a] (operational skills, testing without calculators)	0.31	-0.17[e]	0.24	
Ellington, 2003a[a] (computational skills, testing without calculators)	0.14[e]	-0.25[e]	0.06	
Ellington, 2003a [a] (conceptual skills, testing without calculators)	0.26[d]	-0.29[d]	0.08[e]	
Ellington, 2003a[a] (operational skills, testing with calculators)	0.47	0.34	0.49	
	.285	-.21	.16	

[a] Included both graphing and scientific calculators.
[b] Outliers removed.
[c] Only one study.
[d] One two studies.
[e] Only three studies.

Table C-4: Results from Prior Meta-Analyses on Graphing Calculators Only

Study	Pooled effect size		
	Achievement	*Attitudes*	*Study information*
Khoju, Jaciw, & Miller, 2005	.85		One study was of college students
Ellington, 2006	0.19	0.21[a]	Calculators not allowed in testing
	0.29[a]		Calculators allowed in testing
	Specific education goals		
	Procedural	*Conceptual*	*Combined skills*
Ellington, 2006 (testing without calculators)	–0.21	0.29[a]	0.19
Ellington, 2006 (testing with calculators)	0.32[a]	0.42[a]	0.29[a]

[a] Outliers removed.

Table C-5: Results from Prior Meta-Analyses on Programming

Study	**Pooled effect size**			
Overall	*Achievement*	*Problem solving*	*Attitudes*	*Study information*
Kulik, 1994	.09			All *but* Logo
Lou et al., 2001	.22			
Gordon, 1992	.26	.34		
Kuchler, 1999	.35			Secondary
Lee, 1990	.36	.23	.29	
Khalili, 1994	.45			Logo
Kulik, 1994	.58			Logo
Specific education goals	*Computation*	*Concept & Applications*		*Combination of goals*
Lee, 1990	-.04[a]	.56		.15
Topic	*Arithmetic*	*Geometry*	*Algebra*	*General*
	.40[b]	.68	-.02	.29
Contextual variables				
Age/grade level	*Pre-school*	*Elementary*	*Middle/Junior*	
Lee, 1990 (achievement)		.76	.09	
Lee, 1990 (problem solving)		.29	.27	
Ability level	*Low*	*Middle*	*High*	
Lee, 1990 (achievement)	.22	.11	.37	
Lee, 1990 (problem solving)	.20	-.02	-0.4	
SES	*Low*	*Average*	*High*	
Lee, 1990	.10	.33	.19	
Implementation variables				
Duration	*1–18 weeks*	*19–36 weeks*	*37+ weeks*	
Lee, 1990	.45	.30	.03	
Substitute vs. supplement	*Substitute*	*Supplement*		
Lee, 1990 (achievement)	.40	.34		
Lee, 1990 (problem solving)	.08	.43		
Specific languages	*Logo*	*BASIC*	*Scientific languages*	*Other*
Kuchler, 1999[c]	.78	.34	.42	.47
Khalili, 1994	.45			.33
Lee, 1990	.41	.48		-.15
Kulik, 1994	.58			.09

[a] Only three effect sizes.
[b] Only two effect sizes.
[c] Secondary students.

Table C-6: Results from Prior Meta-Analyses on Tools and Problem Solving Environments

Study	Pooled effect size		
Overall	*Achievement*	*Attitudes*	*Study information*
Lou et al., 2001	.04		Tool and exploratory environments
Kulik & Kulik, 1991	.10		Computer-enhanced instruction
Kuchler, 1999	.24		Problem solving software

Table C-7: Results from Prior Meta-Analyses on Simulation and Games

Study	Pooled effect size			
Overall	*Achievement*	*Attitudes*	*Study information*	
Kulik, 1994	.10			
Kuchler, 1999	.23		Secondary school	
Lee, 1990	.28	.24[a]		
Specific education goals	*Computation*	*Concept & Applications*		*Combination of goals*
Lee, 1990	.61[b]	.24		.63[b]
Topic	*Arithmetic*	*Geometry*	*Algebra*	*General*
Lee, 1990	.61[b]	.24	.15[a]	1.12[a]
Contextual variables				
Age/grade level	*Elementary*	*Junior*		
Lee, 1990	.24	.45		
Gender	*Males*	*Females*		
Lee, 1990	.31	.12		
Implementation variables				
Duration	*1–18 weeks*	*19–36 weeks*		
Lee, 1990	.33	.14		
Substitute vs. supplement	*Substitute*	*Supplement*		
Lee, 1990 (achievement)	.18	.39		
Lee, 1990 (problem solving)	-.83			
Developer	*Experimenter / teacher*	*Commercial*		
Lee, 1990	.26	.29		

[a] Only one effect size included.
[b] Only two effect sizes included.

Table C-8: Subgroup Analysis for Calculator Studies

	Computation				Applications				Concepts		
	N studies/ ES	Hedges g		se	N studies/ ES	Hedges g		se	N studies/ ES	Hedges g	se
Contextual variables											
Grade level											
Elementary	4 / 5	0.367		0.330	2 / 3	0.074		0.371	3 / 3	0.267	0.236
Secondary	3 / 4	0.113		0.169	3 / 4	0.437	**	0.164	1 / 1	0.328	0.489
Mixed	1 / 1	0.855	~	0.512	0 / 0	na		na	0 / 0	na	na
Implementation variables											
Duration		*									
Less than 3 months	5 / 7	0.503	*	0.218	3 / 5	0.295		0.239	2 / 2	0.084	0.307
3 months or greater	3 / 3	-0.134		0.198	2 / 2	0.328		0.249	2 / 2	0.458	0.295

~ p < .10, * p < .05, ** p < .01, *** p < .001

Table C-9: Calculator Effect Sizes not Included in Meta-Analytic Tables

Study	Grade Level	Contrast	Measure	Hedge's g	Standard Error	
Assessments in which calculator group was able to use calculator						
Szetela, 1982	Grade 3	Regular instruction plus calculator-specific materials vs. Regular instructional activities	*Problem-solving*: Posttest 2: 20 researcher-designed items	0.568	~	0.297
Szetela, 1982	Grade 5	Regular instruction plus calculator-specific materials vs. Regular instructional activities	*Problem-solving*: Posttest 2: 20 researcher-designed items	-0.29		0.342
Szetela, 1982	Grade 7	Regular instruction plus calculator-specific materials vs. Regular instructional activities	*Problem-solving*: Posttest 2: 20 researcher-designed items	0.633	*	0.294
Szetela, 1982	Grade 8	Regular instruction plus calculator-specific materials vs. Regular instructional activities	*Problem-solving*: Posttest 2: 20 researcher-designed items	0.589	*	0.274
Alternate interventions/enhancements						
Standifer & Maples, 1981	Grade 3	Programmed feedback calculator vs. No calculator in regular math curriculum	*Computation*: Science Research Associates	0.296		0.395
Standifer & Maples, 1981	Grade 3	Programmed feedback calculator vs. No calculator in regular math curriculum	*Concepts*: Science Research Associates	-0.28		0.395
Standifer & Maples, 1982	Grades 3&4	Programmed feedback calculator vs. General remedial math curriculum	*Computation*: Science Research Associates	0.363		0.326
Standifer & Maples, 1982	Grades 3&4	Programmed feedback calculator vs. General remedial math curriculum	*Concepts*: Science Research Associates	0.064		0.325
Duffy & Thompson, 1980	Grade 4	Calculator plus materials vs. No calculator (classroom-level effect size)	*Computation*: CTBS	-0.01		0.428
Duffy & Thompson, 1980	Grade 4	Calculator plus materials vs. No calculator (classroom-level effect size)	*Problem-solving*: CTBS Applications	-0.15		0.429

Continued on p. 6-223

Table C-9, continued

Study	Grade Level	Contrast	Measure	Hedge's g	Standard Error
Duffy & Thompson, 1980	Grade 5	Calculator plus materials vs. No calculator (classroom-level effect size)	*Concepts*: CTBS	-0.27	0.430
Duffy & Thompson, 1980	Grade 5	Calculator plus materials vs. No calculator (classroom-level effect size)	*Computation*: CTBS	0.222	0.450
Duffy & Thompson, 1980	Grade 5	Calculator plus materials vs. No calculator (classroom-level effect size)	*Problem-solving*: CTBS Applications	-0.090	0.449
Duffy & Thompson, 1980	Grade 6	Calculator plus materials vs. No calculator (classroom-level effect size)	*Concepts*: CTBS	-0.09	0.449
Duffy & Thompson, 1980	Grade 6	Calculator plus materials vs. No calculator (classroom-level effect size)	*Computation*: CTBS	-0.22	0.440
Duffy & Thompson, 1980	Grade 6	Calculator plus materials vs. No calculator (classroom-level effect size)	*Problem-solving*: CTBS Applications	0.191	0.440
Duffy & Thompson 1980	Grade 6	Calculator plus materials vs. No calculator (classroom-level effect size)	*Concepts*: CTBS	-0.21	0.440
Total math achievement (calculator vs. control)					
Standifer & Maples, 1981	Grade 3	Hand-held, four function calculator vs. No calculator in regular math curriculum	*Total achievement*: Science Research Associates	0.309	0.396
Standifer & Maples, 1982	Grades 3&4	Hand-held, four function calculator vs. General remedial math curriculum	*Total achievement*: Science Research Associates	0.341	0.330
Duffy & Thompson, 1980	Grade 4	Calculator only group vs. No calculator (classroom-level effect size)	*Total achievement*: CTBS	0.062	0.428
Duffy & Thompson, 1980	Grade 5	Calculator only group vs. No calculator (classroom-level effect size)	*Total achievement*: CTBS	0.008	0.449
Duffy & Thompson, 1980	Grade 6	Calculator only group vs. No calculator (classroom-level effect size)	*Total achievement*: CTBS	-0.113	0.439
Total math achievement (calculator + enhancement vs. control)					
Standifer & Maples, 1981	Grade 3	Programmed feedback calculator vs. No calculator in regular math curriculum	*Total achievement*: Science Research Associates	-0.047	0.394
Standifer & Maples, 1982	Grades 3&4	Programmed feedback calculator vs. General remedial math curriculum	*Total achievement*: Science Research Associates	0.194	0.325
Duffy & Thompson, 1980	Grade 4	Calculator plus materials vs. No calculator (classroom-level effect size)	*Total achievement*: CTBS	-0.089	0.429
Duffy & Thompson, 1980	Grade 5	Calculator plus materials vs. No calculator (classroom-level effect size)	*Total achievement*: CTBS	0.207	0.450
Duffy & Thompson, 1980	Grade 6	Calculator plus materials vs. No calculator (classroom-level effect size)	*Total achievement*: CTBS	0.021	0.428

Table C-10: Computer Programming Effect Sizes (Comparing Programming to CAI) not Included in Meta-Analytic Tables

Study	Design	Sample	Duration/ Content	Contrast	Measure	Hedge's g	Standard Error
Programming							
Battista & Clements, 1986	RCT	11 fourth-graders in two midwestern middle schools	42 sessions (2 40-min per week)/LOGO	Gr 4: Logo vs. CAI	Problem-solving Combine Test 1&2, Total	0.409	0.611
		26 sixth-graders in two midwestern middle schools		Gr 6: Logo vs. CAI		0.326	0.395
Clements, 1986	RCT	24 first-grade students from a middle-class midwestern school system	44 sessions (22 weeks)/ LOGO	Gr 1: Logo vs. CAI	WRAT Math score	0.486	0.415
		24 third-grade students from a middle-class midwestern school system		Gr 3: Logo vs. CAI		0.473	0.414
Clements, 1987	RCT	16 third-grade students who had received Logo or CAI experience in first-grade	3 months/ LOGO	Gr 3: Logo vs. CAI	CAT - Total	0.452	0.511
Emihovich & Miller, 1988	RCT	36 first-grade students in five classrooms in an elementary school in the southeast	20, 30-min sessions (3 months)/ LOGO	Gr 1: Logo vs. CAI	CTBS - Math	0.214	0.410

~ $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$
[a] Data were adjusted for clustering that occurred within classrooms.

Chapter 7: Report of the Subcommittee on Instructional Materials

Robert S. Siegler, Chair
Bert Fristedt
Vern Williams
Irma Arispe
Daniel B. Berch
Marian Banfield, U.S. Department of Education Staff

CONTENTS

Abbreviations

WWC	What Works Clearinghouse

I. Accuracy of Textbooks

It might be assumed that textbooks for middle school and high school math would be free of errors. When mathematicians have reviewed already published middle and high school textbooks, however, they have identified a nontrivial number of errors, and a large number of ambiguous and confusing statements and problems. Many of these errors and ambiguities arise on word problems that are intended to elicit use of the mathematical concepts and procedures in real-world contexts. The Subcommittee recommends that publishers obtain reviews from mathematicians prior to publication, so that these errors and ambiguities can be identified and corrected. This is especially needed for first editions of textbooks, which tend to have the greatest numbers of errors and ambiguities. Having mathematicians also read textbooks in the formative stages may increase the coherence of the presentation of mathematics between earlier and later grades.

II. Length, Coherence, and Sequencing of Topics

U.S. mathematics textbooks are extremely long. Not counting study guides and answers at the end of the books, middle and high school textbooks typically range from 600 to more than 900 pages. Including the study guides and answers at the end, the books sometimes exceed 1,000 pages. Even elementary school textbooks sometimes exceed 700 pages. The length of math textbooks was much shorter in previous decades and continues to be much shorter in many nations with higher mathematics achievement than the United States.[1] Thus, the great length is not needed for effective instruction.

Textbook publishers emphasize that a major source of the textbooks' length is the need to cover all of the benchmarks encompassed in any state's standards. A topic covered in sixth grade in one state may be covered in seventh grade in another state and in eighth grade in a third state; this leads to the topic being included in all three grades' math textbooks. The large influence of this factor is illustrated by the fact that the state-specific editions of Algebra I textbooks published for California, Texas, and Florida are roughly 25% (more than 200 pages) shorter than the national edition published for the other 47 states. Coverage of all of the states' benchmarks for a given grade is likely to increase length and decrease coherence—this is despite the fact that mathematics is especially amenable to a coherent treatment. Integrating new concepts with previous ones is impossible when textbook writers cannot anticipate which topics students already have encountered. The Subcommittee

[1] Publishers' Testimony. (2006). Testimony of representatives of Harcourt School Publishers, Holt, Rinehart and Winston, Mcdougal Littell/ Houghton Mifflin and Company, Pearson Scott Foresman, Pearson Prentice Hall, and McGraw-Hill Companies at the meeting of the National Mathematics Advisory Panel, Cambridge, MA, September, 2006. (Supporting materials on textbook lengths submitted to the National Mathematics Advisory Panel.)

Schmidt, W.H., McKnight, C.C., & Raizen, S.A. (1997). A splintered vision: An investigation of U.S. science and mathematics education (pp. 1–26). Boston, MA: Kluwer Academic Publishers.

U.S. Department of Education, National Library of Education. (2008). Archived Textbook Collection. Washington, DC.

recommends that states and districts strive for greater agreement regarding which topics will be covered in which grades and that textbook publishers publish editions that include only the material that these states and districts agree to teach in specific grades.

Another indicator and source of lack of coherence of some textbooks is the table of contents. Tables of contents should provide students, teachers, and textbook adoption teams with a sense of the organization of the mathematical topics in the book. In some textbooks, however, tables of contents emphasize not the mathematics but rather specific applications (e.g., Ferris wheels, penny jars). Tables of contents that emphasize the mathematical content seem more likely to help students appreciate the coherence inherent in mathematics.

Other potentially useful ways of decreasing length and increasing coherence are 1) reducing the number of photographs that are not essential to the mathematical content; 2) placing content aimed at providing extra review, enrichment, or motivation in supplements rather than in the main textbook; and 3) excluding applications in which the primary challenge is posed by social studies or science content.

III. The What Works Clearinghouse

The What Works Clearinghouse (WWC) identifies and evaluates studies of the effectiveness of educational interventions (i.e., programs, products, practices, and policies). Using published studies and additional information, such as technical reports, WWC summarizes the strength of evidence about each intervention in terms of established standards. Among these standards is the time frame, design, sample, intervention, outcome, and statistical reporting. Based on the evidence, products are characterized in terms of their effects on student achievement (i.e., positive, potentially positive, negative, potentially negative, mixed, or no discernable effect). WWC also provides information on the extent of evidence, and provides a registry of outcome evaluators (individuals or organizations) who conduct research on the effects of educational interventions. The goal of this registry is to help schools, school districts, and education program developers.

Although many aspects of studies are well characterized within WWC, other aspects are less well specified. The WWC does not evaluate the content of the curriculum, so formal assessments of the length, coherence, and correctness of items within the curricula are not available. In addition, information about methods used to train teachers is often limited or unavailable. This lack of information about teacher training no doubt reflects the level of description within the supporting studies; without such information, however, understanding reasons for the effects or lack of effects of interventions is difficult.

The Subcommittee therefore recommends that WWC report information on 1) curriculum content (including variables related to coherence, length and accuracy) and 2) teacher training and professional development. Such information would help teachers and school districts use the instructional materials. Although this information may not be available from existing studies, requiring it from programs that are evaluated in the future may create a data source that can be used to establish best practices from the perspective of the teacher.

IV. Research Recommendations

A large amount of research has been conducted on instructional materials, but most of it does not meet even moderately stringent methodological criteria. These methodological deficiencies limit the usefulness of the studies in guiding education decisions. The Subcommittee recommends that governmental funding agencies give priority to research that meets stringent methodological criteria, especially randomized controlled designs in which students, classrooms, or schools are randomly assigned to conditions and studied under carefully controlled circumstances. Studies that include large enough samples of students, classrooms, teachers, and schools to identify effects that are present should also be given priority. Such studies are considerably more expensive than studies with small samples, but they provide a much sounder basis for education policy.

BIBLIOGRAPHY

Publishers' Testimony. (2006). Testimony of representatives of Harcourt School Publishers, Holt, Rinehart and Winston, Mcdougal Littell/ Houghton Mifflin and Company, Pearson Scott Foresman, Pearson Prentice Hall, and McGraw-Hill Companies at the meeting of the National Mathematics Advisory Panel, Cambridge, MA, September, 2006. (Supporting materials on textbook lengths submitted to the National Mathematics Advisory Panel.)

Schmidt, W.H., McKnight, C.C., & Raizen, S.A. (1997). *A splintered vision: An investigation of U.S. science and mathematics education* (pp. 1–26). Boston, MA: Kluwer Academic Publishers.

U.S. Department of Education, National Library of Education. (2008). Archived Textbook Collection. Washington, DC.

Chapter 8: Report of the Task Group on Assessment

Camilla P. Benbow, Chair
Susan Embretson
Francis “Skip” Fennell
Bert Fristedt
Tom Loveless
Wilfried Schmid
Sandra Stotsky
Irma Arispe, Ex Officio
Ida Eblinger Kelley, U.S. Department of Education Staff

CONTENTS

Tables

Abbreviations

ACT	American College Testing
AP	Advanced Placement
AYP	Adequate Yearly Progress
CR	Constructed Response
CR-G	Constructed Response Grid In
CR-S	Constructed Response Short Answer
CR-EE	Constructed Response Extended
DIF	Differential Item Functioning
ELL	English Language Learners
ERIC	Education Resources Information Center
IRT	Item Response Theory
MC	Multiple Choice
MSPAP	Maryland State Performance Assessment Program
MTSA	Major Topics of School Algebra
NAEP	National Assessment of Educational Progress
NAGB	National Assessment Governing Board
NAS	National Academy of Science
NCES	National Center for Education Statistics
NCLB	No Child Left Behind
NCTM	National Council of Teachers of Mathematics
NELS	National Education Longitudinal Study
NVS	NAEP Validity Study
PISA	Programme for International Student Assessment
SAT	Scholastic Achievement Test
SAT-M	Math Scholastic Achievement Test
SES	Socioeconomic Status
SSCI	Social Sciences Citation Index
STPI	Institute for Defense Analysis Science and Technology Policy Institute
TIMSS	Trends in International Math and Science Study

Executive Summary

Introduction

Achievement tests are widely used to estimate what students know and can do in specific subject areas. Tests make visible to teachers, parents, and policymakers some of the outcomes of student learning. They also can drive instruction. Due to their important role in education today, especially after enactment of the No Child Left Behind Act, the Panel examined the quality of released items from the mathematics portions of the National Assessment of Educational Progress (NAEP) and six state tests, and reviewed the relevant scientific literature on the appropriate distribution of test content, the setting of performance categories, factors affecting measurement accuracy, and appropriate test design.

Key Questions Addressed by the Task Group

To address the charges in the Executive Order, the Assessment Task Group developed five primary questions:

Part One: Test Content and Performance Categories

1) What should the mathematics content of the NAEP and state tests be at Grades 4 and 8? How does the content of state tests compare with NAEP?
2) How are performance categories determined?

Part Two: Item and Test Design

3) How does item response format affect performance on multiple-choice and various kinds of constructed-response items?
4) What are some nonmathematical sources of difficulty or confusion in mathematics test items that could inappropriately affect performance? How prevalent are they on the NAEP and the six state tests examined?
5) How are calculators used in NAEP and state assessments and how does calculator use affect performance?

These questions are not independent of each other; they overlap because what one tests and how one chooses to test are intertwined. For example, the verbal context of the test items or calculator use could have bearing on what is actually measured.

Test Content

The content strands in most state mathematics tests are similar to the content strands in the NAEP mathematics test. Thus, the Task Group focused its investigation on the NAEP content strands, knowing that any suggestions for the NAEP would have implications for most state mathematics tests as well.

The Task Group presents in Table 1 possible recommendations that could flow from the general principles for organizing the content of the NAEP—that tests should measure what students should be learning. In preparation for algebra, students should become proficient in the critical foundations for algebra as described in the Conceptual Knowledge and Skills Task Group report.

Table 1: Suggested Reorganization of NAEP Content Strands

Grade 4	**Grade 8**
Number: Whole Numbers	Number: Integers
Number: Fractions and Decimals	Number: Fractions, Decimals, and Percents
Geometry and Measurement	Geometry and Measurement
Algebra	Algebra
Data Display	Data Analysis and Probability

The most critical skills to be developed before beginning algebra are extensive work with whole numbers, including whole number operations, and facility with fractions. The NAEP Validity Study (NVS), as well as others, have noted the relative paucity of items assessing fractions in both the fourth- and eighth-grade NAEP.

Moreover, the NVS indicates that half of the data analysis and probability section of the Grade 4 NAEP is probability-related. Given the importance of fractions for the conceptual understanding of probability, the Task Group questions whether probability can be measured appropriately in the fourth grade. Thus, the Task Group suggests that this strand at the fourth-grade level be limited to data analysis and titled as Data Display.

The review of NAEP content also led the Task Group to conclude that there needs to be a more appropriate balance in how algebra is defined and assessed at both the fourth- and eighth-grade levels of the NAEP. At the fourth-grade level, most of the NAEP algebra items relate to patterns or sequences. While the inclusion of patterns in textbooks or as state curriculum expectations may reflect a view of what constitutes algebra, patterns are not emphasized in the curricula of high-achieving countries. In the Major Topics of School Algebra set forth in the Task Group on Conceptual Knowledge and Skills report, patterns are not a topic of major importance. The prominence given to patterns at the preschool through Grade 8 level is not supported by comparative analysis of curricula or by mathematical considerations. Applying the general principle for selecting content for the NAEP and state tests, the Task Group strongly recommends that “algebra” problems involving patterns be greatly reduced in these tests.

It might be useful to note that the Trends in International Math and Science Study (TIMSS) content domains were changed at the time the Task Group was conducting its own work. Adopting the Task Group’s recommendations would bring NAEP into greater alignment with TIMSS (Mullis et al., 2007).

Performance Categories

Once content is selected, decisions must be made about what the performance categories should be and how to assign student scores to them. The Task Group did not investigate what the cut scores for each category should be but, rather, how they should be determined. Although the states and NAEP varied in both process and method for such standard setting (setting cut scores), the six states studied in the NVS report and NAEP employed currently acceptable educational practices to quantify judgments of the standard-setting panelists and to map their judgments onto test scores. Limited research is available on standard-setting methods and processes. The Modified Angoff method requires the most plausible assumptions about raters and tests, but more research is needed comparing the outcomes based on alternative methods.

In the states the Task Group examined, classroom teachers, most of whom are not mathematics specialists, predominate in the standard-setting process. The expertise of mathematicians, as well as of mathematics educators, curriculum specialists, classroom teachers, and the general public, should be consistently used in the standard-setting process. The Task Group also found that the standard-setting panelists often do not take the test as examinees before attempting to set the performance categories, and that they are not consistently informed by international performance data. On the basis of international performance data, there are indications that the NAEP cut scores for performance categories are set too high. This does not mean, however, that the mathematical content of the test is too hard; it is simply a statement about the location of cut scores for qualitative categories such as “proficient” and “beyond proficient.”

Recommendations for Test Content and Performance Categories

1) NAEP and state tests must ensure a focus on the mathematics that students should learn with achievement on critical mathematics content reported and tracked over time. NAEP should ensure that the Conceptual Knowledge and Skills’ Critical Foundations and elements of the Major Topics of School Algebra are integral components of the mathematics assessed. The Task Group proposes reorganization, as well as possible title changes, of NAEP’s current five content strands:

   a. Number Properties and Operations should be renamed and expanded into two separate categories—Grade 4 Number: Whole Numbers; and Fractions and Decimals; and Grade 8 Number: Integers; and Fractions, Decimals, and Percent.

      1. Whole Numbers will include emphasis on place value, comparing and ordering, and whole number operations at Grade 4. This will be expanded to include work with all integers, including operations with negative and positive integers at Grade 8.

      2. Fractions and Decimals will include recognition, representation and comparing and ordering at Grade 4. This will be expanded to include operations involving fractions, decimals, and percent at Grade 8.

b. Geometry and Measurement should be combined into one content area. Topics related to both Measurement and Geometry should serve as important contexts for problems in the Grade 4 and Grade 8 NAEP.

c. Within Algebra, a better balance is needed within the subtopic of patterns, relations, and functions at Grades 4 and 8. That is, there should be far fewer items on patterns.

d. Data Analysis and Probability should be renamed as Data Display at Grade 4 and expanded to include both data interpretation and probability at Grade 8.

2) Procedures should be employed to include a broader base for setting performance categories:

a. The Task Group recommends that standard-setting (setting cut scores) panels include individuals with high levels of expertise, such as mathematicians, mathematics educators, and high-level curriculum specialists, in addition to classroom teachers and the general public.

b. The standard-setting panelists should take the test as examinees before attempting to set the performance categories.

c. The standard setting should be informed by international performance data.

d. Research is needed on the impact of standard-setting procedures and methods (e.g., Bookmark Method, Modified Angoff procedure) in promoting the representation of a broad base of judgments.

Item and Test Design

It is important not only that appropriate content is measured and cut scores for student performance are set appropriately, but also that test scores reliably reflect the competencies intended to be measured. That is, the measurement itself must be carried out in a high-quality and appropriate manner.

Item Response Format

Many educators consider constructed-response items (e.g., short answer) as superior to multiple-choice (MC) items in measuring mathematical competencies and a more authentic measure of mathematical skill. They believe such items also offers the opportunity for students to explain principles and display a range of math skills including verbal explanations. The Task Group examined the literature on the psychometric properties of constructed-response items compared with multiple-choice items. The evidence found in the scientific literature does not support the notion that a constructed-response format, particularly the short-answer type, measures different aspects of mathematics competency compared with a multiple-choice format. While there are skills that may be measured only using a constructed-response format, concern about use of multiple-choice items in these tests at the fourth- and eighth-grade level is not warranted.

Nonmathematical Sources of Difficulty

The NVS panel found many examples of flawed items on NAEP and state assessments that could affect performance of all or some students and trend lines. The Task Group undertook its own examination of released items on state and NAEP tests, looking specifically for nonmathematical sources of difficulty (e.g., particular context portrayed within an item) and found many items on the NAEP and state tests affected by these sources of difficulty, resulting in too many flawed items. The Task Group presents seven types of flawed items illustrating nonmathematical sources of influence that could affect scores. Test developers should be sensitive to the presence of these types of flaws in the test development process.

Careful attention must be paid to exactly what mathematical knowledge is being assessed by a particular item and the extent to which the item is, in fact, focused on the intended mathematics. In other words, significant attention should be devoted to the actual design of individual mathematics items and to the evaluation of items for inclusion. More mathematicians should be involved in the process of designing and evaluating items, as should mathematics educators, curriculum specialists, linguistics experts, and cognitive psychologists.

The frequency of flawed items points to another possible gap in test development procedures that needs to be addressed. Psychometricians are trained to use highly sophisticated statistical models and data analysis methods for measurement, but are not as familiar with issues of item design with respect to measuring mathematical constructs. Item writers and item evaluators often do not have a college degree in the appropriate subject, and apparently do not have the kind of background in task and item design that would lead to a lower percentage of items that are flawed or marginal according to the mathematicians. Moreover, they receive limited feedback from psychometricians on how the items they develop end up functioning for students at varying levels of performance. That is, the feedback mechanism does not provide sufficient information to help pinpoint the sources of item deficiencies.

Calculators

Use of calculators in assessment is another frequently discussed design issue. While findings from the literature revealed that using calculators in assessment has no significant impact on performance overall or in problem solving, the research indicates that calculator use affects performance on computation-related items and also could change the nature of the competencies tested.

Recommendations on Item and Test Design

1) The focus in designing test items should be on the specified mathematical skills and concepts, not item response format. The important issue is how to most efficiently design items to measure content of the designated type and level of cognitive complexity.

2) Much more attention should be paid to what mathematical knowledge is being assessed by a particular item and the extent to which the item addresses that knowledge.

3) Calculators (the use of which constitutes a design feature) should not be used on test items that seek to measure computational skills. In particular, NAEP should not permit calculator use in Grade 4.

4) Mathematicians should be included in greater numbers, along with mathematics educators, and curriculum specialists (not just classroom teachers and the general public), in the standard-setting process and in the review and design of mathematical item content for state, NAEP, and commercial tests.

5) States and NAEP need to develop better quality control and oversight procedures to ensure that test items reflect the best item design features, are of the highest quality, and measure what is intended, with nonmathematical sources of variance in performance minimized.

6) Researchers need to examine whether the language in word problems is suitable for assessing their mathematical objectives before examining their impact in state assessments on student performance, especially the performance of special education students or English language learners.

7) More scientific research is needed on item and test design features.

I. Introduction

Achievement tests are widely used to estimate what students know and can do in specific subject areas. Tests make visible to teachers, parents, and policymakers some of the outcomes of student learning. Tests also can provide an efficient and fair way to assess student achievement. Finally, tests can drive both the content and format of classroom instruction.

Widespread, large-scale testing began in the 1960s after passage in 1965 of the Elementary and Secondary Education Act and the appropriation of Title I funds. Policymakers desired information to gauge the progress of education in the United States as a whole and, thus, the idea for a National Assessment of Educational Progress (NAEP) emerged. NAEP was implemented in 1969, and the long-term trend tests began during the 1972–73 school year. The main NAEP came later, in 1990.

With passage of the No Child Left Behind Act (NCLB) in 2001, the use of testing was expanded beyond its many uses at the time: end-of-course evaluations of learning by teachers; admission tests for college, graduate, or professional programs; loosely structured accountability systems at the school district level; and what had been required under ESEA as reauthorized in 1994 by IASA (the Improving America's Schools Act). In 1994, IASA required states to test students in all schools in reading and mathematics in three grade spans, and to use state assessments. NCLB expanded the number of grades required to be tested. NCLB also mandated, among other things, use of state assessments and other measures to hold schools and districts accountable for increasing all students' achievement, including the achievement of different subgroups of students. A few states had already created content standards on which their state tests, if developed, were based to ensure that students were learning the topics judged important for students to master. And some of these states had made passing state tests a requirement for high school graduation. NCLB, however, required all states to develop standards and state assessments in reading and mathematics in Grades 3–8 and once in high school, and set forth measures of teacher quality, as well. States could choose their own tests and set their own cut scores, but they had to demonstrate annual improvement for all subgroups of students through a measure called Adequate Yearly Progress (AYP).

A provision in NCLB also required all states to participate in NAEP beginning with the 2003 cycle. NAEP was to sample each state every 2 years so that the results on NAEP tests could be compared with the results of state tests. It is intended that NAEP and the state assessments, along with results from international assessments when they are available, inform the public and policymakers on the condition of education in the United States. Given the importance of the NAEP and state tests for measuring the outcomes of education, it is vital that the NAEP and state tests measure appropriately what is deemed important for children to learn in school. For more details on the history of NAEP and the two types of tests it gives, see Appendix A and the NAEP Web site. Descriptions of six state testing programs are provided in Tables 3 and 4.

In this context, the Assessment Task Group of the National Mathematics Advisory Panel (Panel) was formed to address the following charges in the Executive Order:

> (b) the role and appropriate design of standards and assessment in promoting mathematical competence;
>
> (f) the role and appropriate design of systems for delivering instruction in mathematics that combine the different elements of learning processes, curricula, instruction, teacher training and support, and standards, assessments, and accountability (Executive Order No. 13398).

II. Key Questions Addressed by the Task Group

To address the charges in the Executive Order, the Task Group developed five primary questions (divided into two areas):

Part One: Test Content and Performance Categories

1) What should the mathematics content of the NAEP and state tests be at Grades 4 and 8? How does the content of state tests compare with NAEP?
2) How are performance categories determined?

Part Two: Item and Test Design

3) How does item response format affect performance on multiple-choice and various kinds of constructed-response items?
4) What are some nonmathematical sources of difficulty or confusion in mathematics test items that could inappropriately affect performance? How prevalent are they on the NAEP and the six state tests examined?
5) How are calculators used in NAEP and state assessments, and how does calculator use affect performance?

These questions are not independent of each other. They overlap because what one tests and how one chooses to test are intertwined. For example, the verbal context of the test items or calculator use could have bearing on what is actually measured.

III. Background

Two studies were shared with the Task Group that offered a strong foundation for its work and began to answer the key questions. These were: 1) *Validity Study of the NAEP Mathematics Assessment: Grades 4 and 8* (Daro et al., 2007), 2) Response to the Validity Study of the NAEP Mathematics Assessment: Grades 4 and 8 (Schneider, 2007). These two documents, while addressing some of the Task Group's main concerns, led the group to probe some of the reports' findings in deeper and more specific ways.

A. NAEP Validity Study (NVS) Report

The NVS convened an expert panel involving mathematicians, mathematics educators, and an expert on state-based mathematics standards. They compared the NAEP mathematics framework with the standards and frameworks (test blueprints) of six states (California, Massachusetts, Indiana, Texas, Washington, and Georgia), two high-performing nations (Singapore and Japan), and standards outlined by the National Council of Teachers of Mathematics (NCTM) and Achieve, Inc.

The NVS examined the content areas of number properties and operations, algebra, geometry, measurement, and data analysis and probability strands in the 2005 NAEP mathematics framework to determine if NAEP was missing something or overemphasizing topics in a given content area.[1] The reviewers then described what was missing or being overemphasized, and rated the emphasis of each content topic as compared to each of the six states and Singapore and Japan. The panel of mathematicians also examined individual items from the NAEP tests and the six states' tests and found serious problems.

Quoting from the NVS Report:

> *Five percent of NAEP items were found to be seriously flawed mathematically at Grade 4, and 4 percent were designated seriously flawed at Grade 8. The state items were classified as 7 percent seriously flawed in fourth grade and 3 percent seriously flawed in eighth grade. For marginal items, NAEP had 28 percent at Grade 4 and 23 percent at Grade 8, while the state sample had 30 percent at Grade 4 and 26 percent at Grade 8. By this estimation, NAEP is less flawed than some critics have suggested, but it is also less than perfect mathematically. The substantial number of marginal items in NAEP and the states is cause for concern. Marginal items may well be leading to underestimates of achievement, although this study did not produce empirical evidence on this possibility (Daro et al., 2007, p. 79–80).*

The NVS report showed that a high percentages of marginal and flawed items appeared in four major content areas in these tests: Algebra, Geometry, Measurement, and Data Analysis and Probability. The Number Properties and Operations section was better than the other four. As Exhibits IV-3 and IV-4 (pp. 82 and 83) of the NVS report show, for Grade 4, 16 of the 28 NAEP Algebra items and 8 of the 16 state Algebra items were classified as marginal or seriously flawed. As Exhibit IV-5 shows (p. 83) at Grade 8, 16 of 59 NAEP Algebra items were similarly classified. For Grade 4, 15 of 34 NAEP Geometry items, 15 of 42 NAEP Measurement items, and 12 of 19 state Measurement items were so classified. At Grade 8, 11 of 45 NAEP Geometry items, 10 of 37 NAEP Measurement items,

[1] The NVS was asked to address the following questions:
1. Does the NAEP framework offer reasonable content and skill-based coverage compared to the assessments of states (six were selected for study and are described in Table 1 and 2) and other nations?
2. Does the NAEP item pool and assessment design accurately reflect the NAEP framework?
3. Is NAEP mathematically accurate and not unduly oriented to a particular curriculum, philosophy, or pedagogy?
4. Does NAEP properly consider the spread of abilities in the assessable population?
5. Does NAEP provide information that is representative of all students, including students who are unable to demonstrate their achievements on the standard assessment? (Daro et al., 2007, p. i).

and 10 of 25 state Geometry items were so classified. For Grade 4, 5 of 12 state Data Analysis and Probability items were so classified. For Grade 8, 15 of 32 NAEP Data Analysis and Probability items, and 7 of 15 state Data Analysis and Probability items were so classified. In other words, more than one-fourth to more than one-half of the items in these four areas were rated by five mathematicians as not suitable for high-stakes tests.

The NVS report also was concerned about whether the percentage of marginal and flawed items in the NAEP tests might have influenced test results. Quoting from the report:

> *Is it possible or likely that the presence of seriously flawed or marginal items could have altered overall NAEP results? Some of the flaws categorized as "serious" are the mathematical equivalent of grammatical errors: students can still understand the problem situation and answer the questions, so the results are not affected. Still, there is something unacceptable about having such errors on a test. Other types of serious flaws, however, could alter results by creating real obstacles for test takers. The mathematicians also were clear that many of the items they classified as marginal exhibited construct-irrelevant difficulties that could affect performance for some test takers (p. 82).*

Nonetheless, one central finding of the NVS was, "The NAEP mathematics assessment is sufficiently robust to support the main conclusions that have been drawn about United States and state progress in mathematics since 1990" (p. ii and 119). They noted, however, that while the framework was reasonable, the specifications communicated to test developers were not detailed enough. In addition, while they thought item quality was typical of large-scale assessments, it could be improved.

The NVS made recommendations for improving the NAEP that flowed from its study. Two of these recommendations are of particular importance to the Task Group. First was its recommendation to sharpen the focus of the current NAEP framework. Specifically, it recommended, "[F]ocus: don't worry about leaving things out; worry about targeting the most important things…Explicitly address high priority issues that cut across content areas" (p. vi).

The second recommendation in the NVS report of importance to the Task Group involved item quality and the provision of exemplars of good items for future NAEP tests. NVS recommended improved quality assurance, with particular attention focused on the following: mathematical quality of the items, quality of the situated mathematics problems (e.g., word problems), measurement of complexity, non-construct relevant sources of item difficulty (i.e., nonmathematical sources of difficulty; e.g., verbiage; complex graphical displays; vocabulary), item performance and construction (e.g., response format such as multiple choice versus constructed response), and the range of item difficulty and curricular reach.

B. NCES Response to the NVS Report

In its response to the NVS report, the National Center for Education Statistics (NCES) claimed that student gains on the main NAEP have not been underestimated by the number of poor-quality items. The NCES statistical analysis of marginal and flawed items revealed that their mean discrimination indices (mean r-biserial for each category) did not differ from the items judged as adequate (Schneider, 2007). NCES was also less critical of NAEP's quality control process for item development, but NCES concurred with the NVS recommendation about the importance of item quality and the provision of exemplars of good items for future NAEP tests.

IV. Methodology

The Task Group determined a strategy to probe the quality of the state tests given their particular charge. The Task Group needed to determine if the recommendations for the NAEP also applied to the state tests they could inspect and if there potentially were other issues. Thus, the Task Group undertook a review of released test items from six state tests and NAEP. Moreover, it wanted to explore more deeply the validity of the process for setting performance categories, especially given the recent NCES report, *Mapping 2005 State Proficiency Standards onto the NAEP Scales* (National Center for Education Statistics, 2007). In that study, NCES mapped state performance categories in reading and mathematics onto the appropriate NAEP scale using data from fourth and eighth grades in the 2004–05 school year. Finally, the Task Group took some of the recommendations for the NAEP a step further, explored appropriate content, and posed additional questions, such as the impact of calculator usage.

With the assistance of Abt Associates Inc. (Abt), the Task Group conducted searches of the scientific literature with respect to the questions posed. The Institute for Defense Analyses Science and Technology Policy Institute (henceforth STPI) also assisted, in particular, with the review of released test items from the NAEP and state assessments. See Appendix B for more detail on the Task Group's methodology.

The Task Group conducted its work during a three-month time period. Consequently, it was limited in its ability to collect, examine, and analyze an extensive amount of information. For example, the identification of relevant literature was limited to what could be identified and reviewed in that time period. Furthermore, there was insufficient time to field a survey. The analysis was, thus, based on information readily available on state and NAEP Web sites, in publications, and in prior analysis and research.

V. Part I: Test Content and Performance Categories

A. Question 1: What Should the Mathematics Content of NAEP and State Tests Be at Grades 4 and 8? How Does the Content of State Tests Compare with NAEP?

1. Background

Currently, NAEP assesses mathematics organized by the following five content strands:

- Number Properties and Operations
- Geometry
- Algebra
- Measurement
- Data Analysis and Probability

These strands were developed to meet the requirement that large-scale achievement tests of this nature must measure competencies reflecting all areas of mathematics taught and considered most important at a given developmental level. To assess the appropriateness of the content strands, the Task Group examined five sources: 1) the Critical Foundations, the skills and knowledge essential for success in algebra as described in the Conceptual Knowledge and Skills Task Group report; 2) the NAEP Validity Study (Daro et al., 2007); 3) findings from the Task Group's literature review; 4) the Task Group-generated review of released items from NAEP and selected state tests; and 5) the Panel's National Survey of Algebra Teachers (Hoffer, Venkataraman, Hedberg, & Shagle, 2007).[2] In the Panel's survey, teachers identified particular aspects of mathematical content areas (e.g., fractions and success with word problems) as both critically important to the preparation for algebra and insufficiently acquired by students in introductory algebra courses. These findings added to the Task Group's basis for reviewing these topics. In addition, the Task Group's review of some state tests and their released items provided further information on how one might reset the focus of test content frameworks.

2. Reorganizing the Content Strands of NAEP and Implications for State Assessments

Based on the review of the five sources described earlier in this section, the Task Group proposes several principles for reorganization, as well as possible title changes, for the five content strands of the NAEP and, potentially, for state tests. The suggested reorganization is presented in Table 1 and represents the possible outcome of employing the principles for organizing the content of the NAEP. This possible reorganization has implications for state mathematics tests, as well.

[2] The order in which the sources are listed bears no significance of the importance of each source.

Table 1: Suggested Reorganization of NAEP Content Strands

Grade 4	**Grade 8**
Number: Whole Numbers	Number: Integers
Number: Fractions and Decimals	Number: Fractions, Decimals, and Percent
Geometry and Measurement	Geometry and Measurement
Algebra	Algebra
Data Display	Data Analysis and Probability

The suggested principles are as follows:

1) NAEP and state tests must ensure a focus on the mathematics that students should learn with achievement on critical mathematics content reported and tracked over time. NAEP should ensure that the Conceptual Knowledge and Skills' Critical Foundations and elements of the Major Topics of School Algebra are integral components of the mathematics assessed. The Task Group proposes for reorganization, as well as possible title changes, for NAEP's current five content strands:

   a. Number Properties and Operations should be renamed and expanded into two separate categories—Grade 4 Number: Whole Numbers; and Fractions and Decimals; and Grade 8 Number: Integers; and Fractions, Decimals, and Percent.

      1. Whole Numbers will include emphasis on place value, comparing and ordering, and whole number operations at Grade 4. This will be expanded to include work with all integers, including operations with negative and positive integers at Grade 8.

      2. Fractions and Decimals will include recognition, representation and comparing and ordering at Grade 4. This will be expanded to include operations involving fractions, decimals, and percent at Grade 8.

   b. Geometry and Measurement should be combined into one content area. Topics related to both Measurement and Geometry should serve as important contexts for problems in the Grade 4 and Grade 8 NAEP.

   c. Within Algebra, a better balance is needed within the subtopic of patterns, relations, and functions at Grades 4 and 8. That is, there should be many fewer items on patterns.

   d. Data Analysis and Probability should be renamed as Data Display at Grade 4 and expanded to include both data interpretation and probability at Grade 8.

   These principles and their possible implications are now explained in greater detail.

Number Properties and Operations

The Task Group suggests that Number Properties and Operations be expanded and renamed as Number. It should include a focus on whole numbers, including place value, comparing and ordering, and whole number operations (i.e., addition, subtraction, multiplication, division—arithmetic) at Grade 4 and then be expanded to include extensive work with all integers (negative and positive) at Grade 8. A proposed additional content area involving number would focus on fractions. At the Grade 4 level, this would involve beginning work with fractions and decimals including, recognition, representation, and comparing and ordering. This would be expanded to include operations with fractions, decimals, and percent at Grade 8. Similarly, the focus of work with whole numbers and fractions on state tests should expand as concepts and operations are developed from year to year, particularly at Grades 5, 6, and 7, which are grade levels when the NAEP test is not offered.

A review of the eighth-grade NAEP Number Properties and Operations content area (Daro et al., 2007) found an emphasis on topics from number theory—factorization, multiples, and divisibility. Their review suggests, however, the need to ensure that eighth-grade students have developed proficiency with whole numbers, positive and negative integers, fractions, decimals, and percent given their importance as prerequisites for algebra. Because this content area stood out in the NVS review as under-sampling grade-level content, "It is possible that students are making gains in this content area that are not being detected by NAEP" (p. 123). In the Panel's judgment, it is also possible that students are losing ground that goes undetected. Indeed, because the NAEP minimizes this area, this could be a driving force for reduced attention to it within the school curriculum.

One of the Task Group's greatest concerns is that fractions (defined here as fractions, decimals, and related percent) are underrepresented on NAEP. The NVS, as well as others, have noted the relative paucity of items assessing fractions in both the fourth- and eighth-grade NAEP. The validity study identified fewer than 20% of items as involving fractions and decimals in Grade 8. It also was noted that, while Number Properties and Operations should be the most emphasized content area at the fourth-grade level, the NAEP provides a very limited assessment of fractions at this level. Implementation of the Task Group's recommendations would result in a more appropriate balance of content and address the issue of underrepresentation of fractions on the NAEP.

Geometry and Measurement

As seen in Table 1, the Task Group also suggests that Geometry and Measurement be combined into one content area, which would make the Grade 4 and 8 test frameworks consistent with that of Grade 12 NAEP (2005).The proposed merging of these content areas also would address the concern that there is a "need for a close look at how the NAEP measurement objectives compare to the treatment of measurement elsewhere" (Daro et al., 2007, p. 9). Such an examination was deemed important given that measurement is second only to number properties and operations in the fourth-grade NAEP in terms of the number of items assessed in a particular content area. The larger number of measurement items within NAEP however, is "not well leveraged to include fractions or decimals used in realistic situations" (p. 126). It is also noted that while there is considerable overlap in the

NAEP and Trends in International Math and Science Study (TIMSS) assessments involving measurement, there is greater emphasis in NAEP on using measurement instruments and units of measurement. As a result, NAEP may be underestimating achievement in measurement. TIMSS includes a higher percentage of items on estimating, calculating, or comparing perimeter, areas, and surface area.

The review of the Geometry items indicated wide variation across six states. For example, the NAEP at Grade 8 includes more geometry than the comparison states or nations. Also, the eighth-grade NAEP Geometry items assess symmetry and transformations more than those of the states and emphasize parallel lines and angles less than the comparison states. Finally, it should be noted that topics related to both measurement and geometry (e.g., perimeter, area, and circumference) could serve as important contexts for problems within the Grade 4 and Grade 8 NAEP. This constitutes another principle for organizing the content of the NAEP not previously noted.

Algebra

Algebra is the most heavily weighted content topic on the eighth-grade NAEP, with 30% of the assessment targeting algebra objectives (NAEP 2007 framework). Fifteen percent of the fourth-grade NAEP is dedicated to algebra. At the fourth-grade level most of the NAEP algebra items relate to patterns or sequences (Daro et al., 2007). Chazan et al. (2007) also note that released Grade 4 NAEP items place a heavy emphasis on pattern completion at the expense of other types of algebraic reasoning. This is cause for concern. While states' inclusion of patterns in textbooks or as curriculum expectations may reflect their views of what constitutes algebra, patterns are not emphasized in high-achieving countries (Schmidt, 2007). The NVS (Daro et al., 2007) recommended better item balance within the algebra subtopic of patterns, relations, and functions at the Grade 4 level. In the Conceptual Knowledge and Skills Task Group's Major Topics of School Algebra (MTSA), patterns are not a topic of major importance. The prominence given to patterns at the preschool to eighth-grade level is not supported by comparative analysis of curricula or by mathematical considerations (Wu, 2007). In addition, this has broad implications for interpreting student performance. For example, although student performance on the eighth-grade NAEP Algebra strand has increased, reviewers note the underrepresentation of high-complexity items in algebra (Daro et al., 2007). Thus, one cannot be clear on what this increased performance means.

Data Analysis and Probability

While recognizing that data analysis provides the context for many interesting problems in mathematics, the Task Group notes that the work of the NVS indicates that half of the Data Analysis and Probability section of the Grade 4 NAEP is probability related, whereas TIMSS has a greater proportion of items than NAEP that emphasize reading, interpreting, and making predictions from tables and graphs, and data representation, especially at the fourth-grade level. Given the importance of fractions for the conceptual understanding of probability, the Task Group questions whether probability can be taught and measured appropriately at the fourth-grade level. As students begin work with fractions, probability becomes a more viable mathematics topic and thus should come later in elementary and middle school. The Task Group, therefore, suggests that the Grade 4 content area be renamed as Data Display, consistent with TIMSS 2007, and at Grade 8, as Data Analysis and Probability. The focus at the eighth-grade level would be expanded to include both data interpretation and probability.

Comparison to TIMSS

It is useful to note here that the TIMSS content domains were changed (Mullis et al., 2007) at the time the Task Group was conducting its own study. The Grade 4 content domains are now identified as Number, Geometric Shapes, and Measures and Data Displays. At this level, TIMSS has merged Geometry and Measurement, as the Task Group also suggests, and deleted the domain formerly called Patterns, Equations, and Relationships, again consistent with the concerns the Task Group raises about patterns and algebra at the fourth-grade level. The Grade 8 content domains are Number, Algebra, Geometry, and Data and Chance. At this level, TIMSS has infused Measurement within Geometry and expanded Data to include Probability. The Task Group's suggested principles for reorganizing the NAEP would bring it into greater alignment with TIMSS.

3. A Comparison of State Test Content with NAEP Content

How does the content of State Tests Compare with the Content of the NAEP Tests?

A comparison by strand of the percentage of items on state tests for Grades 4 and 8 with the percentage of items on the Grade 4 and Grade 8 NAEP test in 2007 yields the following results in Table 2:

Table 2: A Comparison of State Test Content with NAEP Content

Grade 4			**Grade 8**		
	NAEP (2007)	*State Tests*		*NAEP (2007)*	*State Tests*
Numbers	40%	15–48%	Numbers	20%	10–26%
Measurement and Geometry (combined)	35%	18–34%	Measurement and Geometry (combined)	35%	20–28%
Algebra	15%	12–28%	Algebra	30%	20–28%
Data Analysis, Statistics and Probability	10%	6–20%	Data Analysis, Statistics and Probability	15%	12–20%

Source: Daro et al., 2007.

What do these comparisons indicate? First, they show considerable differences in content distribution across these six states for most strands at both Grades 4 and 8, as well as differences from the weight given a strand on the corresponding NAEP test. The percentages or content emphasis for each strand at each grade level for each of the six states can be seen in Table 3. Table 4 shows the percentages for each strand on the 2003 and 2007 NAEP tests. In Grade 4, NAEP has a greater emphasis on Numbers and in Measurement and Geometry combined than all six states but a much lower percentage in Algebra and Data Analysis, Statistics, and Probability. In Grade 8, NAEP still has a greater emphasis on Measurement and Geometry combined than all six states, but (because of a change in the weight assigned Algebra from 2003 to 2007) now has a higher percentage of its 2007 test in Algebra than all six states. The NAEP tests tend to be lower in the percentage of items in Data Analysis, Statistics, and Probability at both grade levels than most of the six states.

More information from a literature review offers additional support for organizing NAEP and state assessments in the areas of content frameworks can be found in Appendix C.

Table 3: Summary of State Mathematics Test Content, Grade 4 and 8

State	Grade	Content Strand Weights - Number of Questions or Points Possible (% of total)						Total Questions or Points	Item Formats	Use of testing aids		
		Number	Measurement	Geometry	Algebra and Functions	Statistics, Data Analysis, and Probability	Other		Constructed Response	Calculators	Manipulatives	Formula Sheets
California	4th	31 (48%)	12 (18%)		18 (28%)	4 (6%)		65 (100%)	Multiple Choice	No	—	—
	8th	100%						65	Multiple Choice	No	No	No
Georgia (QCC)	4th	— (20%)	— (23%)		— (12%)	— (10%)	— (35%)	— (100%)	Multiple Choice	No	—	—
	8th	— (17%)	— (22%)		— (23%)	— (12%)	— (26%)	— (100%)		No	—	—
Indiana	5th	13* (15%)	13* (15%)	11* (13%)	14* (17%)	10* (12%)	23* (27%)	84* (100%)	Multiple Choice, Constructed Response	No	—	No
	9th	9* (10%)	16* (17%)	10* (11%)	25* (27%)	11* (12%)	22* (24%)	93* (100%)		Yes	—	Yes
Massachusetts	4th	19* (35%)	13* (25%)		11* (20%)	11* (20%)		54* (100%)	Multiple Choice, Short Answer, Open Response	No	Yes	No
	8th	14* (26%)	14* (26%)		15* (28%)	11* (20%)		54* (100%)		by section	No	Yes
Texas	4th	11 (26%)	6 (14%)	6 (14%)	7 (17%)	4 (10%)	8 (19%)	42 (100%)	Multiple Choice, some Gridded	No	ruler	measurement conversions
	8th	10 (20%)	5 (10%)	7 (14%)	10 (20%)	8 (16%)	10 (20%)	50 (100%)		No	ruler	measurement conversions
Washington	4th	3–6 (8–17%)	3–6 (8–17%)	3–6 (8–17%)	3–6 (8–17%)	3–6 (8–17%)		35 (100%)	Multiple Choice, Short Answer, Extended Response	by section	by section	High School only
	8th	4–7 (9–20%)	4–7 (9–20%)	4–7 (9–20%)	4–7 (9–20%)	4–7 (9–20%)		50 (100%)		by section	by section	High School only

Notes:
* Content strand weight based on number or points possible instead of number of items in strand.
— Information not available

California: The expectation for 8th grade is that students will take CST Algebra 1 test. However, only about half the cohort takes that test. The others take a general math test as they are not ready for algebra.

Indiana's ISTEP+ is administered in the fall of each academic year and draws from the curricula of all previous grades.

Other strands are Computation and Problem Solving (Georgia and Indiana) and Mathematical Processes and Tools (Texas).

Source: This table was created for the Task Group by STPI using publicly available data from state Web sites. Data on California from S. Valenzuela (personal communications, February 1, 2008).

Table 4: Summary of 2003 and 2005 NAEP Mathematics Test Content, Grade 4 and 8

<table>
<tr><th rowspan="2">Year</th><th rowspan="2">Grade</th><th colspan="5">Content Strand Weights - (% of total)</th><th colspan="3" rowspan="2">Cognitive Dimension</th><th rowspan="2">Item Formats</th><th colspan="3">Use of testing aids</th></tr>
<tr><th>Number</th><th>Measurement</th><th>Geometry</th><th>Algebra and Functions</th><th>Statistics, Data Analysis, and Probability</th><th>Calculators</th><th>Manipulatives</th><th>Formula Sheets</th></tr>
<tr><td rowspan="2">2003</td><td>4th</td><td>40%</td><td>20%</td><td>15%</td><td>15%</td><td>10%</td><td rowspan="2">Conceptual: at least 1/3 of items</td><td rowspan="2">Procedural: at least 1/3 of items</td><td rowspan="2">Problem Solving: at least 1/3 of items</td><td rowspan="2">Multiple Choice (50%), Short and Extended Constructed Response (50%)</td><td>four function calculators provided for approximately 1/3 of items</td><td>students are provided rulers</td><td rowspan="4">Selected formulas and conversion factors (ones students are not necessarily expected to have memorized) are given on a per-item basis (e.g., volume of a cylinder, number of feet in a mile).</td></tr>
<tr><td>8th</td><td>25%</td><td>15%</td><td>20%</td><td>25%</td><td>15%</td><td>scientific calculators provided for approximately 1/3 of items</td><td>students are provided rulers and protractors</td></tr>
<tr><td rowspan="2">2005</td><td>4th</td><td>40%</td><td>20%</td><td>15%</td><td>15%</td><td>10%</td><td rowspan="2">Low Complexity 25% of score</td><td rowspan="2">Moderate Complexity 50% of score</td><td rowspan="2">High Complexity: 25% of score</td><td>Multiple Choice (64%), Short Constructed Response (32%), Extended Constructed Response (4%)</td><td>four function calculators provided for approximately 1/3 of items</td><td>students are provided rulers</td></tr>
<tr><td>8th</td><td>20%</td><td>15%</td><td>20%</td><td>30%</td><td>15%</td><td>Multiple Choice (69%), Short Constructed Response (28%), Extended Constructed Response (4%)</td><td>scientific calculators provided for approximately 1/3 of items</td><td>students are provided rulers and protractors</td></tr>
</table>

Notes:
Various populations, rather than individual students, are the targets of the NAEP assessments. In particular, the assessment administered to any given student does not follow all the strict NAEP guidelines for mathematics assessment composition. Instead, the guidelines apply to the entire set of items for a given year and grade. The entire set of items consists of ten 25-minute blocks. The booklets administered to students participating in the mathematics assessment contain only two 25-minute blocks, in part to minimize the burden on students participating in the assessment. In effect, each student takes one-fifth of an assessment.

Assessments in 2003 and earlier classified the "cognitive dimension" of an item according to the "mathematical ability" required of a student responding to the item (conceptual understanding, procedural knowledge, and problem solving). The 2005 assessment changed the focus to the item itself; it classified the cognitive dimension of an item according to its complexity (low, moderate, high). On the 2003 assessments, a single item may be assigned to more than one mathematical ability level. Thus, this rule means that at least one-third of the items must have a major element of conceptual understanding. For 2005 Item Format Percentages see http://www.ed.state.nh.us/education/doe/organization/Curriculum/NAEP/2005/NAEPReport4MathWCoverRecoveredCorrect.pdf.

Source: STPI compiled this table using information from 1) National Assessment Governing Board, U.S. Department of Education, Mathematics Framework for the 2005 National Assessment of Educational Progress, September 2004, retrieved on October 1, 2007 from http://www.nagb.org/pubs/m_framework_05/toc.html and 2) National Assessment Governing Board, U.S. Department of Education, Mathematics Framework for the 2003 National Assessment of Educational Progress, September 2002, retrieved on October 1, 2007 from http://www.nagb.org/pubs/math_framework/toc.html.

B. Question 2: How Are Performance Categories Determined?

Question 1 concerned the nature and the weighting of the content that should appear on the assessment of mathematics. Question 2 examines how students' scores on mathematics tests are assigned to a particular performance category, e.g., Basic or Proficient. Of foremost concern is the minimum performance level on a test required for a student to be placed in a certain category. Performance level categories appear on both NAEP and the state tests, but the labels and underlying procedures may differ.

1. Background

Establishing performance categories involves a set of procedures currently known in educational measurement as standard setting (or setting cut scores). Judgments about performance categories are made by a panel of persons selected for their expertise or educational perspective. The exact procedures for determining performance categories can range from the panel's judgment about the test as a whole (i.e., the minimum percent of items passed at the various levels) to quantified judgments of individual items with respect to expected performance of students in the categories.

Several procedures and methods for combining judgments in standard setting have been developed. These procedures typically involve training panelists on the definitions of the standards and the nature of performance within the categories, soliciting judgments about the relationship of the test to the performance categories, and providing successive feedback to the panelists about their judgments. Various methods to combine judgments have been developed. Variants of the Bookmark method and the Modified Angoff method involve panelists judging how students at varying levels of competency will respond to representative test items. In these two methods, the cut score for competency classifications is determined by linking the judgments to empirical indices of item difficulty. In contrast, the Body of Work method requires the panelist to classify representative students into competency categories by examining their full pattern of item responses. While the methods are all scientifically acceptable, they may differ in effectiveness. The Bookmark method may involve the most assumptions about the data, while the Body of Work method may demand the highest level of rater judgment. While more research is needed in this area, the Modified Angoff method performs well against several criteria for psychometric adequacy (Reckase, 2006).

The Task Group was interested in determining the nature of the performance categories and the standard-setting procedures and methods for NAEP and the six states.

2. A Review of State Assessments and NAEP

The standard-setting procedures of NAEP and six states were examined with respect to the following seven questions. Not all information was fully available on all questions for each state.

1) What are the performance standards of NAEP and the states?
2) How were the NAEP and state performance standards established?
3) Are they based on procedures in which experts inspect actual item content or on global definitions?
4) Are empirical procedures used to combine individual expert opinions?
5) What is the background of the experts?
6) What descriptions or instructions are given, if any, about the nature of performance at different levels?
7) Do the experts receive the items in an examination under the same conditions as the students?

To answer these questions, documents available from Web sites of NAEP (National Assessment Governing Board) and six states (California, Georgia, Indiana, Massachusetts, Texas and Washington) were retrieved by STPI and provided to the Task Group. These documents were reviewed for relevant data by the Task Group members.

For the first question, NAEP employs a three-category system, Basic, Proficient and Advanced. The six states employed similar categories, although some made more distinctions than others. California's performance categories are labeled as Far Below Basic, Below Basic, Basic, Proficient, Advanced; Georgia, Does Not Meet, Meets, Exceeds Standard; Indiana, Did Not Pass, Pass, Pass+; Massachusetts, Warning, Needs Improvement, Proficient, Advanced; Texas, Basic, Proficient, Advanced; and Washington, Basic, Proficient, Advanced. For NAEP and all six states, global definitions of the performance categories are available. The definitions are all characterized as "global" because the definitions were fairly abstract characterizations of behavior that would require high degrees of judgment to determine the categorization of student performance.

For the other six questions, we draw on data in Table 5. First, although there is variability in the methods, all states use a contemporary method for standard setting or setting cut scores. Second, the Bookmark method was most frequently applied in standard setting. Second, individual item content is judged in NAEP and in all states except Massachusetts, where whole tests from students at various performance levels are examined. Third, empirical combination of judgments is implemented in all states. Fourth, the background of the experts varies within panels and probably somewhat across states. For example, Georgia uses primarily classroom teachers as experts while Texas represents broader contingencies, including curriculum experts from higher education and non-educators. However, in both NAEP and the six states, classroom teachers generally predominate as standard-setting panelists. Fourth, all six states train the panelists prior to eliciting their ratings. Fifth, although all six states applied some training procedures for panelists, the Task Group cannot judge the quality without having access to exact content. Sixth, only two states have the panelists experience the items in the same way as the test-takers.

Table 5: Information on Features of Standard-Setting Procedures (Setting Cut Scores) for NAEP and the Six States

	1. How Established?	2. Item Content Judgments?	3. Combination Procedures	4. Background of Experts	5. Instructions & Definitions	6. Test Taken?
NAEP	Modified Angoff Method	Yes	Empirical with successive feedback.	55% teachers, 15% non-teacher educators, and 30% members of the general public. Panelists should be knowledgeable in mathematics. Panelists should be familiar with students at the target grade levels. Panelists should be representative of the nation's population in terms of gender, race and ethnicity, and region.	Yes	N/A
California	Bookmark Method	Yes	N/A	N/A	Yes	Yes
Georgia	Modified Angoff Method	Yes	Empirical with successive feedback.	Primarily the panelists selected were educators currently teaching in the grade and content area for which they were selected to participate.	Yes	Yes
Indiana	Bookmark Method	Yes	Empirical preliminary followed by feedback & consensus.	Not specifically given, but appears to be classroom teachers.	Yes	None specified. Probably first viewed in panel setting.
Massachusetts	Expert Opinion – Body of Work Method	No	Empirical aggregation of first judgments. Details not available about feedback & consensus.	The panel consists primary of classroom teachers, school administrators or college and university faculty, but also non-educators including scientists, engineers, writers, attorneys, and government officials.	Yes	None specified. Probably first viewed in panel setting.
Texas	Item-mapping	Yes	Empirical preliminary followed by feedback & consensus.	The majority of the panelists on each committee were active educators—either classroom teachers at or adjacent to the grade level for which the standards were being set, or campus or district administrative staff. All panels included representatives of the community "at large."	Yes	None specified. Items probably first viewed in panel setting.
Washington	Bookmark Method	Yes	Empirical preliminary followed by feedback & consensus.	The majority of the panelists on each committee were active educators—either classroom teachers with some representation of higher education.	Yes	Yes

Source: This table was created by the Task Group using publicly available information from state Web sites. Data on California from S. Valenzuela (personal communications, February 1, 2008).

3. Conclusion

Although NAEP and the six states the Task Group examined varied in both process and method for standard setting or setting cut scores, NAEP and all states for which information was available employed currently acceptable educational practices. The methods may differ in effectiveness; however, scant evidence on their efforts is available. The Bookmark method may involve the most assumptions about the data while the Body of Work method may demand the highest level of judgment from the raters. The Modified Angoff method is preferred (Reckase, 2006) because the assumptions of the Bookmark method (e.g., unidimensionality) are probably not met in practice. The Body of Work method is less often applied to year-end tests because it requires higher levels of judgments from the experts. More research is needed on standard-setting methods and processes.

It was found that classroom teachers, most of whom are not mathematics specialists, predominate in the standard-setting process. Higher levels of expertise, including the expertise of mathematicians, as well as mathematics educators, curriculum specialists, classroom teachers and the general public, should be consistently used in the standard-setting process. The Task Group also found that the standard-setting panelists often do not take the complete test as examinees before attempting to set the performance categories, and that they are not consistently informed by international performance data. On the basis of international performance data, there are indications that the NAEP cut scores for performance categories are set too high. This does not mean that the test content is too hard; it is simply a statement about the location of cut scores for qualitative categories such as "proficient" and "beyond proficient." Additional information on this literature review can be found in Appendix D.

C. Part I: Recommendations on Test Content and Performance Categories

1) NAEP and state tests must ensure a focus on the mathematics that students should learn with achievement on critical mathematics content reported and tracked over time. NAEP should ensure that the Conceptual Knowledge and Skills' Critical Foundations and elements of the Major Topics of School Algebra are integral components of the mathematics assessed. The Task Group proposes reorganization, as well as possible title changes, of NAEP's current five content strands:

    a. Number Properties and Operations should be renamed and expanded into two separate categories—Grade 4 Number: Whole Numbers; and Fractions and Decimals; and Grade 8 Number: Integers; and Fractions, Decimals, and Percent.

        1. Whole Numbers will include emphasis on place value, comparing and ordering, and whole number operations at Grade 4. This will be expanded to include work with all integers, including operations with negative and positive integers at Grade 8.

2. Fractions and Decimals will include recognition, representation and comparing and ordering at Grade 4. This will be expanded to include operations involving fractions, decimals, and percent at Grade 8.

b. Geometry and Measurement should be combined into one content area. Topics related to both Measurement and Geometry should serve as important contexts for problems within the Grade 4 and Grade 8 NAEP.

c. Within Algebra, a better balance is needed within the algebra subtopic of patterns, relations, and functions at this level. That is, there should be many fewer items on patterns.

d. Data Analysis and Probability should be renamed as Data Display at Grade 4 and expanded to include both data interpretation and probability at Grade 8.

2) Procedures should be employed to include a broader base for setting performance level categories:

a. The Task Group recommends that standard-setting (setting cut scores) panels include high levels of expertise, such as mathematicians, mathematics educators, and high-level curriculum specialists, in addition to classroom teachers and the general public.

b. The standard-setting panelists should take the complete test as examinees before attempting to set the performance categories.

c. The standard setting should be informed by international performance data.

d. Research is needed on the impact of standard-setting procedures and methods (e.g., Bookmark Method, Modified Angoff procedure) in promoting the representation of a broad base of judgments.

VI. Part II: Item and Test Design

It is important not only that appropriate content is measured and cut scores for student performance are set appropriately but also that test scores accurately reflect the competencies intended to be measured. That is, the measurement itself must be carried out in a high-quality and appropriate manner. Test specifications that dictate the content of mathematics are not sufficient to ensure that valid assessments will be obtained. Thus, the Task Group reviewed the area of item and test design.

A. Question 3. How Does Item Response Format Affect Performance on Multiple-Choice and Various Kinds of Constructed-Response Items?

1. Background

Constructed-response formats, in which the examinee must produce a response rather than select one, are increasingly utilized in standardized tests. One motivation to use the constructed-response format arises from its presumed ecological validity, the belief that it reflects tasks in academic and work settings, and stresses the importance of “real-world” tasks. Constructed-response formats also are believed to have potential to assess dynamic cognitive processes (Bennett, Ward, Rock, & Lahart, 1990), and systematic problem solving and reasoning at a deeper level of understanding (Webb, 2001), as well as to diagnose the sources of mathematics difficulties (Birenbaum & Tatsuoka, 1987). Finally, constructed-response formats also may encourage classroom activities that involve skills in problem solving, graphing, and verbal explanations of principles (Pollack, Rock, & Jenkins, 1992). However, these purported advantages can incur a cost. The more extended constructed-response formats require raters to score them. Hence, they are more expensive and create delays in test reporting, as well as possibly introducing subjectivity in scoring.

In contrast, multiple-choice items have been the traditional type used on standardized tests of achievement and ability for over a century. Multiple-choice items can be inexpensively and reliably scored by machines or computers; they may require relatively little testing time and they have a successful history for psychometric adequacy.

Constructed-response (CR) items vary substantially in the amount of material that an examinee must produce. There are three basic types of CR items:

- The *grid-in* constructed-response format (CR-G) requires the examinee to obtain the answer to the item stem and then translate the answer to the grid by filling in the appropriate bubble for each digit.
- The *short answer* constructed-response format (CR-S) varies somewhat. The examinee may be required to write down just a numerical answer or the examinee may need to produce a couple of words to indicate relationships in the problem. The

CR-S format potentially can be scored by machine or computer, given a computerized algorithm that accurately recognizes the varying forms of numerical and verbal answers. Further, an intelligent algorithm also may provide for alternative answers (e.g., slightly misspelled words, synonyms).

- The *extended* constructed-response format (CR-EE) requires the examinee to provide the reasoning behind the problem solution. Thus, the CR-EE format would include worked problems or explanations. This format readily permits partial credit scoring; however, human raters are usually required. Use of human raters, however, can lead to problems with consistency and reliability of scoring.

The stems of CR-G and CR-S items and multiple-choice (MC) items can be identical, especially if the correct answer is a number. It is not clear how the stems of CR-EE can be identical to MC items, although this possibility cannot be excluded.

The NMP Assessment Task Group examined the literature on the psychometric properties of constructed-response items as compared to multiple-choice items. The original focus was to address the following three questions:

1) Do the contrasting item types (e.g., multiple choice, constructed response) capture the same skills in these tests equally well?
2) What does the scientific literature reveal?
3) What are the implications for NAEP and state tests?

2. A Review of the Literature

Impact of response format on mathematical skills, knowledge, and strategies. Potentially the most pressing issue about response format is the extent to which the same skills, knowledge, and strategies can be measured by the MC and CR item formats. The research generally does not support major differences in the nature of the construct that is measured by CR and MC items, nor in the strategies that are applied. However, much more data on this issue are potentially available because many state accountability, graduation, and year-end tests employ both item formats.

Impact of response format on psychometric properties. The evidence about the psychometric properties of constructed-response items as compared to multiple-choice items is inconsistent and depends on the source and the design of the comparison. If the studies utilize operational test data, comparisons of MC and CR items have indicated greater omit rates and greater difficulty for the CR items. This pattern is probably repeated on many state tests and would be a strong finding if such data were available for study by the methods employed in the Task Group’s study. It should be noted, however, that studies on operational test items were not designed to isolate the impact of format by controlling or measuring other properties of items. If the studies utilized stem-equivalent versions of MC and CR items, the difference in psychometric properties depended on other design features of the items, such as the nature of the distractors and the use of grid-in responses. For example, some studies have found the CR format to be more difficult, which is consistent with the operational test studies. Other studies, however, have found the MC items to be more difficult when the distractors are constructed to represent common error patterns. Moreover, little evidence from any design is available to support differences between MC and CR items on item discrimination levels, differential item functioning (DIF), strategy use, and the nature of the construct that is measured.

Impact of response format on differences between groups. The results on the interaction of the magnitude of gender-related differences in performance and item format are inconsistent and depend on the design of the specific study. However, the evidence suggests either no impact of response format on gender-related differences or that the relatively lower scores of girls than boys on mathematics items may be lessened in the constructed-response format.

The interaction of racial-ethnic differences with item format also has been examined in several studies. The research provides some evidence that Black-White differences in performance in mathematics are lessened on CR item format as compared to the MC item format.

Other results on item format that are potentially interesting include Hastedt and Sibberns (2005) finding on TIMSS data that scores based on MC versus CR items produced only slight differences in the relative ranking of the various participating countries. And, DeMars (2000) found that the difference between MC and CR items in difficulty depended on the test context. The two item formats differed less in difficulty on high-stakes tests than on low-stakes tests.

Additional information from this literature review can be found in Appendix E.

3. Conclusion

The available evidence on comparing the psychometric properties of MC items and CR items must be interpreted in the context of several factors. These factors include the following limitations: 1) the limited scope of the available scientific literature, 2) the uncontrolled design features for comparisons based on operational tests, 3) the design strategy in available controlled comparisons of MC and CR items, 4) the limited scope of the controlled comparisons, and 5) the impact of test context on the relative performance on MC and CR items. These limitations and the methodology for this review are discussed in more length in Appendix E.

Given the limitations of the research, there is little or weak evidence to support the CR format as providing much different information than the MC format. For example, the available evidence provides little or no support for the claim that different constructs are measured by the two formats or that item discrimination varies across formats. Although some evidence suggests that CR items are more difficult, especially for the more extended CR formats, there is some contrary evidence that indicates that more difficult MC items can be constructed for their stem-equivalent CR items. Finally, the impact data do not support much difference between the two item formats. That is, the impact of response format on gender differences is inconsistent, while the impact on racial-ethnic differences is weak. Suggestions to guide the evaluation of assessment item design are listed in Appendix F.

B. Question 4: What are Some Nonmathematical Sources of Difficulty or Confusion in Mathematics Test Items That Could Negatively Affect Performance?

Because flawed and marginal items on NAEP and state assessments could affect performance of students and could affect trend lines, the Task Group probed this issue.

1. Background

A crucial skill in learning mathematics is gaining the ability to understand what mathematical relationships and operations are intended by the language of word problems. Word problems are very common in most if not all state assessments, as well as in school curriculum materials. Nevertheless, it is clear that several nonmathematical aspects of word problems can adversely influence performance on tests of mathematical competence. These include misleading language and confusing visual displays. Problems also can emerge when reading, writing, and other skills that overlap with mathematical competence have an undue influence on performance.

The chapter of the NVS report on “Language that is unclear, inconsiderate, or misleading” provides only three examples of test items that show language or wording defects, even though many test items exhibited difficulties of this nature, as suggested by the comments on pages 94 and 95 (Daro et al., 1997). In addition, only two of the examples in this section were mathematical story problems, or, as they may also be labeled, situated mathematics problems.

How prevalent are poorly worded problems on high-stakes assessments? The Task Group wanted to find out if there was evidence on the frequency of language or wording issues from other analyses of test items on state, NAEP, or commercial mathematics assessments. Have any researchers systematically analyzed state, national, or commercial tests to determine the number of problems with poorly chosen, or developmentally inappropriate, unnecessary, or misleading language? Have any researchers found empirical evidence on the difficulties that students in general or various subgroups of students have with items that could be judged as linguistically defective? Are there research-based recommendations on language or wording issues to avoid, not only in abstract mathematics problems (problems not contextualized in real life) but especially in applied, or situated, problems that typically use everyday language to describe the givens of a mathematical problem?

2. A Review of the Literature

The Task Group undertook a review of the literature by examining 28 studies that met the Panel’s criteria for quantitative empirical studies. The methodology for this review can be found in Appendix F. The Task Group was able to group most of these 28 studies into three general areas of interest in mathematics assessments. Seven looked at gender-related issues, two of which (Sappington, Larsen, Martin, & Murphy, 1991; McLarty, Noble, & Huntley, 1989) examined whether gender-related wording in mathematics word problems could lead to a difference in scores between boys and girls. A third (Low & Over, 1993) reported on whether girls more often incorporated irrelevant information into constructed responses than did boys. The other four examined the differences in boys' and girls' scores on mathematics word problems with respect to their format, i.e., whether they provided MC options for response or required CR (Reiss & Zhang, 2006; Pomplun & Capps, 1999; Ryan & Fan, 1996; Wilson & Zhang, 1998).

The studies that examined differential responses by gender to the format of a test item found that, as also noted in the section comparing MC and CR formats, that for the most part, females do better on CR formats, while boys do better on MC formats. However, complexities appear when these differences are explored in greater depth. Reiss and Zhang (2006) found, when they controlled for language skills, that girls did less well than boys on both types of formats, more so on MC than on CR. The researchers observed that "the advantage females have in reading and writing improves their mathematics scores" while "males' lower reading and writing scores negatively impact their mathematics performance" (p. 13). In no way, however, do the researchers suggest that reading and writing skills are not important in mathematics; at issue was the role these skills play in mathematics assessments. It was not clear to the researchers whether raters rated responses by females more favorably because their responses were more complete (and of a higher quality) or because the females wrote more words, a dilemma in interpretation that has been found in writing assessments.

Another four studies examined, in differing ways, mathematical problem-solving difficulties for students who have learning disabilities, mathematics disabilities, or low reading skills (Fuchs & Fuchs, 2002; Moyer, Moyer, Sowder, & Threadgill-Sowder, 1984; Moyer, Sowder, Threadgill-Sowder, & Moyer, 1984; Larsen, Parker, & Trenholme, 1978). These students have difficulty reading and understanding how to solve mathematics word problems, sometimes because of the syntactic complexity of the language. Indeed, three other studies (Ketterlin-Geller, Yovanoff, & Tindal, 2007; Bolt & Thurlow, 2006; Johnstone, Bottsford-Miller, & Thompson, 2006) used a "read-aloud" method to explore what these kinds of students find difficult, in part to determine how items might be altered to remove what these students verbalized as difficulties. But in none of these studies did the researchers explore what the tests were actually measuring or whether the test items were defective or abnormal in any way. As a result, they did not explore the effects of erroneous, misstated, or poorly constructed items on student performance.

Five other studies examined assessment issues for English language learners (ELLs) (Brown, 2005; Butler & Stevens, 1997; Abedi & Lord, 2001; Abedi, Lord, Hofstetter, & Baker, 2000; Abedi, 2003). The researchers were interested in the effects of these students' English language limitations on test performance, in ways to accommodate their English language limitations on test items, or in the effects of accommodations in test items on them. In none of these studies, however, did the researchers examine how appropriate the language in test items was for assessing their mathematical objectives, whether the studies examined the effects of the original language or of altered language.

The remaining studies examined a variety of other issues, ranging from correlations of socioeconomic status (SES), race, and ethnicity with achievement (Lubienski, 2001; Lubienski, 2002; Lubienski & Shelley, 2003), the influence of scoring quality on assessment reliability (Myerberg, 1999), and various issues in mathematics education (Romberg, 1992), to the features of mathematics and language arts tests constructed in various ways (Perkhounkova & Dunbar, 1999), a processing model to predict difficulty (Kintsch & Greeno, 1985), and the effects of personalized and reworded mathematics word problems on problem-solving skill (DeCorte, Verschaffel, & De Win, 1985; Davis-Dorsey, Ross, & Morrison, 1991). However, none of these studies addressed the relationship between flawed items and individual or group differences in performance. That is, this research did not examine the suitability of the language in a word problem for assessing a mathematical objective.

The Task Group examined other research that focused on content validity, reliability, and item performance. They reviewed the bibliography of the NVS for references to studies addressing the language or wording of test items. The five-page NVS bibliography revealed no published studies on language-related factors in test items influencing mathematical performance. The Task Group also examined the table of contents of a newly published volume on mathematics assessment. Not one of the titles of the 22 chapters in *Assessing Mathematical Proficiency,* edited by Alan Schoenfeld, published by Cambridge University Press in May 2007, hints at a discussion of language or wording issues in test items. Nor does a recent article by Lane and Stone (2006). None of these works addressed the Task Group's specific interest in language or wording issues.

The Task Group's review of research on content frameworks also noted no studies examining mathematical item quality, aside from the NVS itself. Most of the studies describe content validity, and examine scope of content and depth of treatment within content areas in relation to national and international tests. Only a few comment extensively on item difficulty and complexity, which may be considered aspects of item quality or test content or both, as seems to be the case in a 2004 analysis of the contents of six state exit tests by Achieve, Inc.

In sum, while the Task Group found many studies on other aspects of mathematics assessments, including item performance and item difficulty, they did not locate any studies that examined how suitable the wording of a test item may be for its mathematical objectives or the effects of wording-related issues in test items on student performance. Therefore, the Task Group proceeded to examine an array of test items from NAEP and state tests to see what kinds of language or wording flaws could be found.

3. Seven Types of Flaws in Released Items from State Assessments and NAEP

The Task Group determined first the extent to which quality is an issue as it relates to the language or wording of an item. The Task Group did an initial cursory reading of the word problems in the 2005 NAEP assessments for Grade 4 and Grade 8, and in assessments for Grade 4 and Grade 8 from six states: California, Georgia, Indiana, Massachusetts, Texas, and Washington. No special significance should be attached to these particular states, except that they were included in the NVS report. The Task Group simply wanted a sufficiently varied pool of items. In all cases, it used only released items that were supplied to them.

The Task Group did not look for all kinds of defects in test items. It focused only on defects in the language or visual displays for word problems. It did not try to determine 1) correctness of answers, 2) appropriateness of constructed-response items, 3) the quality of the rubrics given for grading constructed responses, or 4) the quality of wrong options in multiple-choice items. A test item was judged unsatisfactory if its language or visual display seemed to be distracting, confusing, or misleading, or if its wording or context made the test item too difficult for some students with grade-level mathematics skills. That is, the Task Group focused on the wording or the context for the problem that might, in its judgment, lead some students to give wrong responses independent of their mathematical skills.

A sufficient number of unsatisfactory items were found to warrant a detailed review of released items, with the goal of pinpointing various types of flaws. An array of released test items from NAEP and state tests were then examined. The Task Group stresses that examples could have equally been selected from other states to illustrate the types of flaws found. Enough flawed items were found to support a recommendation that, in future test development, careful attention should be paid to exactly what mathematical knowledge is being assessed by a particular item and the extent to which the item is, in fact, focused on that mathematics.

Below are seven types of flaws that the Task Group identified. Some of the graphics below have been reduced in size for ease of presentation. The Task Group also found many examples of satisfactory word problems in which the nonmathematical knowledge is minimal and for which the student is expected, as appropriate for a mathematics test, to convert relationships described verbally into mathematical symbolism or calculations. See Appendix G for examples of satisfactory word problems.

1) Use of nonmathematical knowledge in a word problem that might not be equally available to all students, or use of terms whose meaning might not be equally available to all students.

   For example: Grade 8: Block 8, M12-Item 11 on NAEP 2005

   Ms. Thierry and 3 friends ate dinner at a restaurant. The bill was $67. In addition, they left a $13 tip. Approximately what percent of the total bill did they leave as a tip?

   A) 10 % *B) 13 %* *C) 15 %* *D) 20 %* *E) 25 %*

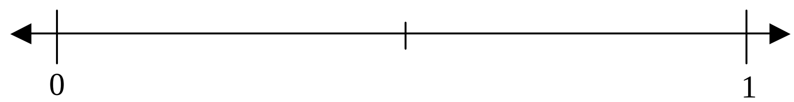
   Comment: This problem assesses conversion of a relationship described verbally into appropriate mathematical symbolism. But there are terminology issues that might trip up some students who would otherwise be able to understand the relationship described. They might not know what a tip is. More importantly, the use of ‘bill’ in one place and ‘total bill’ in another place clouds the relationship: Which is correct: 13/67 or 13/80? Some students will have the nonmathematical knowledge needed for this problem. For others, it will be unfamiliar or vague. It is this feature that makes this question flawed.

2) Use of a "real-world" setting for an essentially mathematical problem, a use that seems to serve only as a distraction because there is no apparent mathematical purpose for that setting.

For example: Grade 4, Massachusetts, 2006
Question 16: Multiple Choice
Reporting Category: Number Sense and Operations
Standard(s): 4.N.4 (No calculator permitted)

The picture below shows four fractions and a number line. Wilson's homework is to place a point on the number line for the location of each of the fractions.

$\frac{1}{6}$ $\frac{1}{3}$ $\frac{1}{12}$ $\frac{1}{4}$



If Wilson places the fractions correctly, which fraction will be closest to 0 *on the number line?*

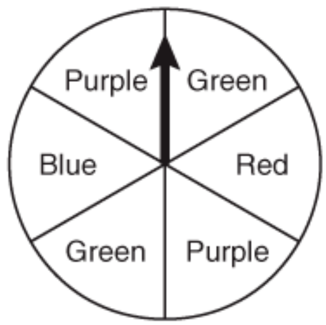
A. $\frac{1}{6}$ *B.* $\frac{1}{3}$ *C.* $\frac{1}{12}$ *D.* $\frac{1}{4}$

Comment: The content of this problem is strictly mathematical. The test-taker must identify which of four given fractions is closest to 0 on the number line. Students who know that $\frac{1}{12}$ is the smallest of the four fractions and understand the relationship between smallness and closeness to 0 should choose the correct answer. But some fourth-graders might be confused by seeing the fractions listed twice or be distracted by the story about Wilson. Straightforward mathematical questions should not be turned into questions about what someone else might do.

3) A focus on logical reasoning in what is essentially a nonmathematical problem.

For example: Grade 4, Texas, Problem 32 in probability and statistics

Kyle will spin the arrow on a spinner like the one shown below.



If Kyle spins the arrow twice, which of these is NOT a possible outcome?

F. Green, green *G. Purple, green* *H. Blue, blue* *J. Red, orange*

Comment: There is no mathematics involved in selecting Option J, the right answer. The student who did not choose Option J did not read the problem with care. While careful reading is part of solving a mathematics problem, a problem involving logical reasoning on a broadly given mathematics assessment also should have a mathematical component.

4) An unnatural sequence of sentences in the word problem, probably created to make the problem "suitable" for assessing mathematical reasoning.

For example: Grade 8, Washington, Problem 33

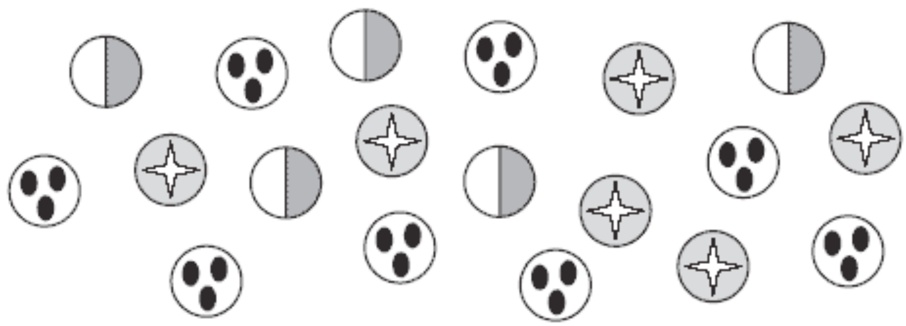
Barb's class is conducting a walkathon. Her mother pledges $15.00. Her father pledges $3.50 per mile. Barb says she can determine the amount of money she will earn using the equation p = *3.5* m + *15. Explain the meaning of* m *in the equation. Explain the meaning of* p *in the equation.*

Comment: The natural question has been convoluted so as to permit the test-maker to ask for explanations.

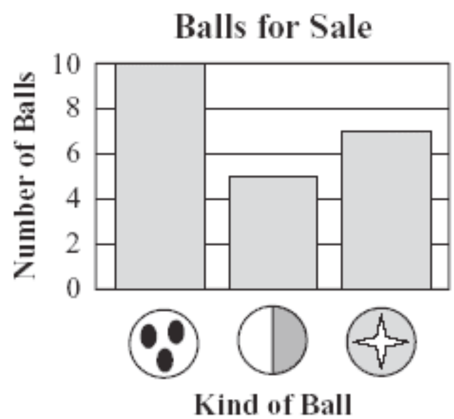
5) Use of a visual display having little connection with mathematics.

For example**:** Grade 4, Massachusetts, 2006, Question 20: Multiple Choice
Reporting Category: Data Analysis, Statistics, and Probability
Standard(s): 4.D.1 (No calculator permitted)

The picture below shows the balls that are for sale at a store.



Which of the following graphs shows the correct number of each kind of ball?

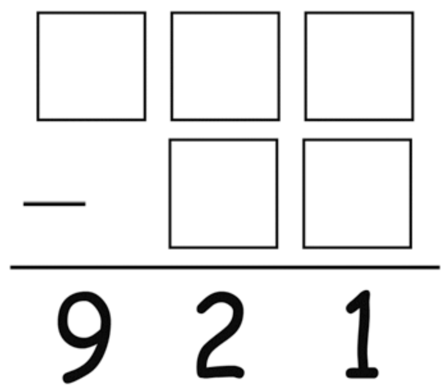


Comment: Solving this problem requires good eyesight as well as the ability to point and count with one hand while covering already counted items with the other.

6) A reliance on general understanding or ingenuity beyond the level of the actual mathematics involved.

   For example: Grade 4, NAEP 2005, Problem 11, Page 3–16

   Audrey used only the number tiles with the digits 2, 3, 4, 6, and 9. She placed one tile in each box below so that the difference was 921. Write the numbers in the boxes below to show where Audrey placed the tiles.

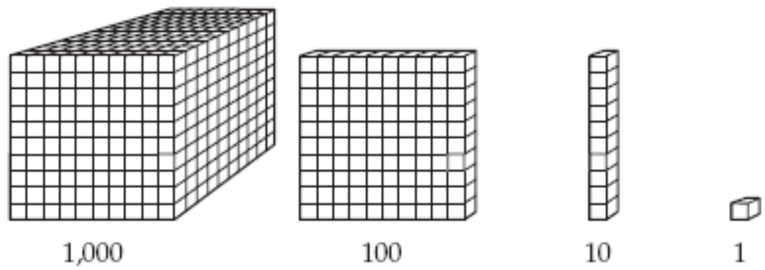


   Comment: The mathematics involved is an understanding of the subtraction algorithm. But some students who are proficient with the subtraction algorithm might get this problem wrong because of its puzzle format. While the skills for doing this puzzle can be taught, they are not critical skills in mathematics, and large-scale assessments should not, in effect, be saying that it is important that every teacher teach these skills. Although it might be argued that this problem also involves mathematical reasoning and that only students who can reason at this level will do the problem correctly, this particular type of mathematical reasoning is not central to mathematics.

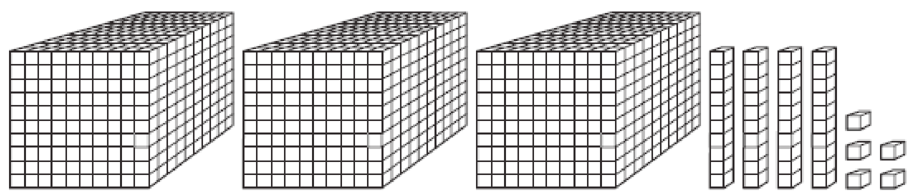
7) A required understanding of a pedagogical technique or tool that might be used for teaching mathematics but is not a part of its content.

For example: Grade 4, Indiana, Problem 3, Page 2–3

Look at the place-value blocks below.



What number does the following place-value model represent?



Answer _______________

Comment: Place-value blocks are a tool for teaching, but one should not expect all students to be familiar with them. A student could figure out how to do the problem without ever having heard of a place-value block, but this makes the item more difficult for such a student than for a student who used one in class.

4. Discussion

A crucial skill in mathematics is the ability to understand what mathematical relationships and operations are intended by the language of word problems. But, flawed items that contain misleading language or confusing visual displays could affect performance of students and could affect trend lines and comparisons from NAEP and state assessments. Because the NVS report indicated that NAEP and state assessments include many items with misleading language and confusing visual displays, the Task Group searched the literature for relevant studies.

No directly relevant studies were identified on how suitable the wording of a test item may be for its mathematical objectives or the effects of wording-related issues on student performance. Thus, research on how aspects of mathematical problems and their item context (e.g., item format, problem scenario, wording, visual displays) are related to the construct that is measured, psychometric properties, and adverse impact should be supported, conducted, and reported. Furthermore, positive examples of well-designed items are needed to guide test development.

To begin this process, the Task Group examined state tests to provide examples of both undesirable and desirable content in mathematics word problems. In approaching this task, the Task Group's premise was that the major purpose of word problems on broadly given assessments should be to assess skill in converting relationships described verbally into mathematical symbolism or calculations. Many flawed items were found on the state tests in sufficient quantity to raise further concerns about item quality. The examples given above illustrate seven types of flaws that were found. Our findings, when combined with NVS findings on the large percentage of flawed and marginal items, point to possible gaps in test development procedures that need to be addressed. Developers of NAEP and state tests use sophisticated psychometric models and methods to select items and yet, according to NCES, these statistics are unable to detect the type of flaws noted in the NVS study.

Several aspects of the item and test development process may contribute to the large numbers of undetected flawed and marginal items.

First, there is a gap in the educational background of psychometricians and item writers. Psychometricians are trained to use highly sophisticated statistical models and data analysis methods for measurement but are not as familiar with issues of item design with respect to measuring mathematical constructs. Typical item writers and item evaluators often do not have a college degree in the appropriate subject and typically have little or no training in task and item design.

Second, item writers receive limited feedback from psychometricians on how the items they develop end up functioning for students at varying levels of performance. That is, the feedback mechanism does not provide sufficient information to help pinpoint the sources of item deficiencies.

Third, traditional psychometric indices of item quality are not sufficient indicators of item quality. According to the NCES report, the flawed and marginal items differed little from the adequate items in the average biserial correlations with total score, which is a classical test theory indicator of item quality. On other achievement tests, such as the state tests, the statistical criteria for evaluating item quality may be set much lower than the indices reported by NCES for NAEP. The lowered statistical criteria may be necessary to accommodate the inherent heterogeneity of educational achievement tests. Requiring high item discrimination may counter efforts to broadly represent an item domain. But an unintended consequence of broad representation is that it can allow even more items with marginal features to meet the low standard.

Fourth, it is increasingly maintained in some educational circles that ensuring that test items fulfill blueprints, along with traditional psychometric indices of item quality, provides sufficient evidence for test validity (e.g., Lissitz & Samuelson, 2007). As the findings of NVS suggest, these criteria do not provide the necessary assurance that students are responding to the items in the manner assumed by the test developers. Further, relying only on content specifications contrasts sharply with current standards for constructing tests (Myerberg, 1999), which expect multiple kinds of evidence for the construct validity of any test. While content specifications are part of the required evidence to support educational test validity, other kinds of evidence are also needed, including evidence based on theory, logical analysis, and scientific research (Embretson, 2007). Specifically, they include the current theory of the domain structure (e.g., the Conceptual Knowledge and Skills Task Group's view of how content "strands" relate to performance in algebra) and item design features. For the latter, the Task Group cannot assume without empirical evidence that students do indeed apply the knowledge, processes, and strategies that are intended for an item classified in a blueprint.

These several factors, taken together, work against ensuring that the items used to assess mathematical competencies are of the highest quality. Better procedures in item development, quality control, and oversight appear needed to counter this problem.

5. Conclusion

The Task Group examined state tests to provide examples of desirable content in mathematics word problems. In approaching this task, the Task Group's premise was that the major purpose of word problems on broadly given assessments should be to assess skill in converting relationships described verbally into mathematical symbolism or calculations. However, word problems also should satisfy the following conditions:

a) be written in a way that reflects natural and well-written English prose at the grade level assessed;
b) assess mathematics knowledge and skills for the grade level of the assessment, as judged by agreed-upon benchmarks, while restricting nonmathematical knowledge to what would be general knowledge for most students.

c) clearly assess skill in converting relationships described verbally into mathematical symbolism or calculations or, if a "real-world" setting is used, the problem uses a setting that aids in solving a problem that would be more cumbersome to state in strictly mathematical language.

Appendix G offers examples of word problems that follow these guidelines.

C. Question 5: How Are Calculators Used in NAEP and State Assessments and How Does Calculator Use Affect Performance?

1. Background

Tests that assess achievement in mathematics are administered under a variety of conditions, and using a variety of procedures, instructions, and technologies. For example, some tests are administered in large groups using paper and pencil booklets while other tests are administered in small groups in which each student is seated at a computer. These conditions may affect performance. Taken together, the diverse conditions under which tests are administered constitute an area of test design.

A very salient aspect of test design is the use of calculators. On some tests, calculators are made available for all items while, on other tests, calculators are available for only some items or for no items at all. Calculator use may affect performance in several ways, including total time on test, the strategies that students apply, the skills that are measured, and might result in differences between diverse groups.

Abt Associates Inc. conducted a review of the scientific literature on the effects of calculator usage on mathematics achievement test scores, using selection criteria described in the Assessment Task Group methodology statement. Below is a description of the studies identified, followed by a synthesis of the results of that literature search.

2. A Review of the Literature

Loyd (1991) noted, in a study involving eighth-graders completing a summer enrichment program (45% of the 160 students), that there was no evidence that use of calculators increased or decreased the speed with which examinees performed on four different types of items on a test. Calculator use was found to be advantageous with some item types (computation-based items), but less so with others.

Loveless (2004a) investigated the extent to which the use of calculators on NAEP computation items at the fourth-grade level produced significantly different results compared to student performance when calculators were not used. He also analyzed the impact of using calculators on performance gaps among black, white, and Hispanic students. Findings indicated that large differences in performance on computation items occurred when students used calculators on the fourth-grade NAEP. In 1999, students averaged 85% correct on whole number computation items when using calculators. On the same items, students who did not have access to calculators averaged only 57% correct on whole-number computation items.

Deeper analyses showed differences in achievement within the whole number operations of addition, subtraction, multiplication, and division. Interestingly, when comparing white, black, and Hispanic students, the gaps relative to achievement in computation narrow when black and Hispanic students have access to and use calculators. A conclusion drawn from this work is that, when young children have access to calculators on test items that focus on computations involving whole numbers, the results will not be indicative of their computational fluency with the operations assessed. In support of Loveless (2004a), Carpenter et al. (1981) also found increased performance on the long-term NAEP computation items for all three age groups if calculators were allowed but not on the problem-solving items.

Dye (1981) assessed eighth-graders using one of the forms used in a prior state mathematics assessment. One student group had access to calculators, one group was told that they could bring calculators and use them if they wished, and one group did not use calculators. The results indicated that the use of a calculator did not make any significant difference on final test scores; however, it was found that, if a mathematics test included many computation items, using a calculator would increase scores. It needs to be noted, however, that some design problems in this study lessened the Task Group's confidence in the conclusions drawn.

Hanson et al. (2001) studied 50 eighth-grade students completing a set of NAEP problems and a set of computation tests with their own calculator and comparable sets of problems with a scientific calculator provided to them. The researchers found no performance advantages associated with calculator type, nor was there an advantage related to student background characteristics (gender, race, math ability, socioeconomic status). Hanson et al. did find that calculator preference depended on the complexity of the student's own calculator relative to the standard one provided. The researchers concluded that there was no compelling reason to prohibit students from bringing their own calculators to a testing situation. On the other hand, the work of Chazan et al. (2007) seems to indicate that experience with calculators matters. They discovered, on the 2003 eighth-grade NAEP, that students who use calculators on a regular basis in their schooling scored higher on algebra and functions items than students who reported little use of calculators. Among all eighth-graders, regardless of socioeconomic status, the average scale scores of students who reported that they used calculators was 6 to 11 points higher on algebra and functions items than those who reported that they did not use calculators.

Brooks et al. (2003) analyzed calculator use on the Stanford Series Achievement Tests. They found that the score differences between calculator users and nonusers on the Stanford 10, which is the latest edition of the Stanford Achievement Series, were not large enough to warrant development of separate score conversion tables. This decision is consistent with findings on recent prior editions of the Stanford Series. The American College Testing Program (ACT) conducted a study in 1996 to assess effects of using a calculator on ACT's mathematics tests. The main purpose was to determine the effect of calculator use on the ACT's PLAN-ACT score scale. This study found that calculator use did not affect scores on either the PLAN or ACT tests. Additionally, the study revealed no differences related to gender and ethnicity with regard to calculator use on the PLAN and ACT tests. On the College Board Scholastic Assessment Test (SAT), however, Lawrence and Dorans (1994) did find that, while most of the items on experimental versions of a pretest taken by thousands of students were unaffected, calculator usage affected item difficulties of those test items that had a heavy computational load. Such items became less difficult with calculator usage.

Long et al. (1989) looked at the role of calculators on the performance on the Missouri state test in 1987 for the tested 8th- and 10th-graders. In these two grades, the use of calculators was allowed but calculators were not provided. Students who used calculators did perform better but the advantage decreased as problem complexity increased. Ansley et al. (1989) and Forsyth and Ansley (1982) reported similar results with two samples of Iowa high school students. Use of calculators did not affect scores on the quantitative test of the Iowa Test of Educational Development, which is purported to be a test of problem solving. In the Ansley et al. (1989) study, all students in the 10th grade of a single high school were tested and randomly assigned to calculator or no-calculator condition. In the Forsyth and Ansley (1982) study several high schools that were matched on student characteristics participated, with each school being randomly assigned to calculator or no-calculator condition.

3. Conclusion

Based on the literature review conducted by the Task Group, it does not appear that using a calculator has a significant impact on test scores overall. However, the use of a calculator does seem to increase scores on computation-related items. Tables 3 and 4 capture key features of the NAEP and the six state tests, including information on calculator and tool use in assessment. Calculators are permitted for use in solving 35–40% of the fourth-grade NAEP test items. This is not the case for the six state tests reviewed. (One of the six states allowed calculators but only on certain sections.)

Thus, care must be taken to ensure that computational proficiency is not assessed using calculators. Additionally, the Task Group highlights one more aspect of this issue. It appears as if students who are comfortable with the calculator may have an advantage in knowing how and when the calculator may be used profitably in problem solving. If there are differences, therefore, in comfort level, the use of calculators might add nonmathematical sources of difficulty to test scores. This should be avoided.

D. Part II: Recommendations on Item and Test Design

1) The focus in designing test items should be on the specified mathematical skills and concepts, not item response format. The important issue is how to most efficiently design items to measure content of the designated type and level of cognitive complexity.

2) Much more attention should be paid to what mathematical knowledge is being assessed by a particular item and the extent to which the item addresses that knowledge.

3) Calculators (the use of which constitutes a design feature) should not be used on test items that seek to measure computation skills. In particular, NAEP should not permit calculator use in Grade 4.

4) Mathematicians should be included in greater numbers, along with mathematics educators, and curriculum specialists (not just classroom teachers and the general public), in the standard-setting process and in the review and design of mathematical item content for state, NAEP, and commercial tests.

5) States and NAEP need to develop better quality control and oversight procedures to ensure that test items reflect the best item design features, are of the highest quality, and measure what is intended, with nonmathematical sources of variance in performance minimized.

6) Researchers first need to examine whether the language in word problems is suitable for assessing their mathematical objectives before examining their impact in state assessments on student performance, especially the performance of special education students or English language learners.

7) More scientific research is needed on item and test design features.

BIBLIOGRAPHY

Abedi, J. (2003). *Impact of student language background on content-based performance: Analyses of extant data*. CSE Report, CSE-R-603.

Abedi, J., & Lord, C. (2001). The language factor in mathematics tests. *Applied Measurement in Education, 14*(3), 219–234.

Abedi, J., Lord, C., Hofstetter, C., & Baker, E. (2000). Impact of accommodation strategies on English language learners' test performance. *Educational Measurement: Issues and Practice, 19*(3), 16–26.

Achieve, Inc. (2004). *Do graduation tests measure up?: A closer look at state high school exit exams*. Washington, DC: Achieve, Inc.

American College Testing Program. (1996). *Using calculators on the PLAN and ACT assessment mathematics tests.* Iowa City, IA: American College Testing Program.

Ansley, T.N., Spratt, K.F., & Forsyth, R.A. (1989). The effects of using calculators to reduce the computational burden on a standardized test of mathematics problem solving. *Educational and Psychological Measurement, 49*(1), 277–286.

Behuniak, P., Rogers, J.B., & Dirir, M.A. (1996). Item function characteristics and dimensionality for alternative response formats in mathematics. *Applied Measurement in Education, 9*(3), 257–275.

Bennett, R.E., Ward, W.C., Rock, D.A., & Lahart, C. (1990). Toward a framework for constructed-response items (ETS Research Report No. 90-7). Princeton, NJ: Educational Testing Service.

Birenbaum, M., & Tatsuoka, K.K. (1987). Open-ended versus multiple-choice response formats—It does make a difference for diagnostic purposes. *Applied Psychological Measurement, 11*(4), 385–395.

Birenbaum, M., Tatsuoka, K.K., & Gurtvirtz, Y. (1992). Effects of response format on diagnostic assessment of scholastic achievement. *Applied Psychological Measurement, 16*(4), 353–363.

Bolt, S.E., & Thurlow, M.L. (2006). *Item-level effects of the read-aloud accommodation for students with reading disabilities*. (Synthesis Report 65). Minneapolis, MN: University of Minnesota.

Brooks, T.E., Case, B.J., Cerrillo, T., Severance, N., Wall, N., & Young, M.J. (2003). *Calculator use on Stanford series mathematics tests*. Austin, TX: Harcourt Assessment, Inc.

Brown, C.L. (2005). Equity of literacy-based math performance assessments for English language learners. *Bilingual Research Journal. 29*(2), 337–363.

Burton, N.W. (1996). How have changes in the SAT affected women's math scores? *Educational Measurement: Issues and Practice, 15*(4), 5–9.

Butler, F.A., & Stevens, R. (1997). *Accommodation strategies for English language learners on large-scale assessments: Student characteristics and other considerations*. No. 448 CSE Technical Report. Los Angeles, CA: Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Carpenter, T.P., Corbitt, M.K., Kepner, H.S., Lindquist, M.M., & Reys, R.E. (1981). Calculators in testing situations: Results and implications from national assessment. *Arithmetic Teacher*, *28*(5), 34–37.

Chazan, D., Leavy, A.M., Birky, G., Clark, K., Lueke, M., McCoy, W., et al. (2007). What NAEP can (and cannot) tell us about performance in algebra. In P. Kloosterman & F. Lester, Jr. (Eds.), *Results and interpretations of the 2003 mathematics assessment of the national assessment of educational progress* (pp.186–188). Reston, VA: National Council of Teachers of Mathematics.

Daro, P., Stancavage, F., Ortega, M., DeStefano, L., & Linn, R. (2007). *Validity study of the NAEP mathematics assessment: Grades 4 and 8.* (Chapters 2 and 3). Washington, DC: National Center for Education Statistics and American Institutes for Research.

Davis-Dorsey, J., Ross, S.M., & Morrison, G.R. (1991). The role of rewording and context personalization in the solving of mathematical word problems. *Journal of Educational Psychology, 83,* 61–68.

De Corte, E., Verschaffel, L., & De Win, L. (1985). Influence of rewording verbal problems on children's problem representations and solutions. *Journal of Educational Psychology*, *77,* 460–470.

DeMars, C.E. (1998). Gender differences in mathematics and science on a high school proficiency exam: The role of response format. *Applied Measurement in Education, 11*(3), 279–299.

DeMars, C.E. (2000). Test stakes and item format interactions. *Applied Measurement in Education*, *13*(1), 55–77.

Dossey, J.A., Mullis, I.V.S., & Jones, C.O. (1993). *Can students do mathematical problem solving? Results from constructed-response questions in NAEP's 1992 mathematics assessment* (No. 23-FR01). Princeton, NJ: Educational Testing Service.

Dye, D.L. (1981). *The use and non-use of calculators on assessment testing*. St. Paul, MN: Minnesota State Department of Education.

Embretson, S.E. (2007). Construct validity: A universal validity system or just another test evaluation procedure? *Educational Researcher, 36*, 8, 449–456.

Executive Order No. 13398, National Mathematics Advisory Panel, *Federal Register*, Vol. 71, No. 77, p. 20519, April 21, 2006.

Forsyth, R.A., & Ansley, T. N. (1982). The importance of computational skill for answering items in a mathematics problem-solving test: Implications for construct validity. *Educational and Psychological Measurement*, *42*(1), 257–263.

Fuchs, L.S., & Fuchs, D. (2002). Curriculum-based measurement: Describing competence, enhancing outcomes, evaluating treatment effects, and identifying treatment nonresponders. *Peabody Journal of Education*, *77*, 64–84.

Fuchs, L.S., & Fuchs, D. (2002). Mathematical problem solving profiles of students with mathematics disabilities with and without comorbid reading disabilities. *Journal of Learning Disabilities*, *35*(6), 563–573.

Gallagher, A. (1992). *Strategy use on multiple-choice and free-response items: An examination of differences among high scoring examinees on the SAT-M* (No. ETS Research Report No. 92-54). Princeton, NJ: Educational Testing Service.

Garner, M., & Engelhard Jr., G. (1999). Gender differences in performance on multiple-choice and constructed response mathematics items. *Applied Measurement in Education*, *12*(1), 29–51.

Hanson, K., Brown, B., Levine, R., & Garcia, T. (2001). Should standard calculators be provided in testing situations?: An investigation of performance and preference differences. *Applied Measurement in Education*, *14*(1), 59–72.

Hastedt, D., & Sibberns, H. (2005). Differences between multiple choice items and constructed response items in the IEA TIMSS Surveys. *Studies in Educational Evaluation*, *31*, 145–161.

Hoffer, T., Venkataraman, L., Hedberg, E.C., & Shagle, S. (2007). *Final report on the national survey of algebra teachers for the National Math Panel*. Chicago: University of Chicago, National Opinion Research Center.

Hombo, C.M., Pashley, K., & Jenkins, F. (2001). *Are grid-in response format items usable in secondary classrooms?* Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.

Johnstone, C.J., Bottsford-Miller, N.A., & Thompson, S.J. (2006). *Using the think aloud method (cognitive labs) to evaluate test design for students with disabilities and English language learners (Technical Report 44)*. Minneapolis, MN: National Center on Educational Outcomes.

Karantonis, A., & Sireci, S.G. (2006). The bookmark standard-setting method: A literature review. *Educational Measurement: Issues and Practice*, *25*(1) pp. 4–12.

Katz, I.R., Bennett, R.E., & Berger, A.E. (2000). Effects of response format on difficulty of SAT-Mathematics items: It's not the strategy. *Journal of Educational Measurement, 37*(1), 39–57.

Kenney, P. (2000). Families of items in the NAEP mathematics assessment. In N.S. Raju, J.W. Pellegrino, M.W. Bertenthal, K.J. Mitchell, & L.R. Jones (Eds.), *Grading the nation's report card: Research from the evaluation of NAEP*. Washington, DC: National Academy Press.

Kenney, P., Silver, E., Alacaci, C., & Zawojewski, J. (1998). *Content analysis project: State and NAEP mathematics assessments. Report of results from the Maryland-NAEP Study.* Conducted under contract with the National Assessment Governing Board. Retrieved on September 1, 2007 from http://www.marces.org/mdarch/pdf/M031915.pdf.

Ketterlin-Geller, L., Yovanoff, P., & Tindall, G. (2007). Developing a new paradigm for conducting research on accommodations in mathematics testing. *Exceptional Children, 73*(3), 331–347.

Kintsch, W., & Greeno, G. (1985). Understanding and solving word arithmetic problems. *Psychological Review, 92*(1), 109–129.

Koretz, D., Lewis, E., Skewes-Cox, T., & Burstein, L. (1993). *Omitted and not-reached items in mathematics in the 1990 National Assessment of Educational Progress* (No. CSE Technical Report 357). Los Angeles, CA: Center for Research on Evaluation, Standards and Student Testing (CRESST) University of California Los Angeles.

Lane, S., & Stone, C.A. (2006). Performance assessments. In B. Brennan (Ed.), *Educational measurement*. Westport, CT: Praeger.

Larsen, S., Parker, R., & Trenholme, B. (1978). The effects of syntactic complexity upon arithmetic performance. *Learning Disability Quarterly*, *1*(4), 80–85.

Lawrence, I.M., & Dorans, N.J. (1994). *Optional use of calculators on a mathematical test: Effect on item difficulty and score equating.* Princeton, NJ: Educational Testing Service.

Lewis, D.M., Mitzel, H.C., & Green, D.R. (1996). Standard setting: A bookmark approach. In D. R. Green (Chair), *IRT-based standard-setting procedures utilizing behavioral anchoring.* Symposium conducted at the Council of Chief State School Officers National Conference on Large-scale Assessment, Phoenix, AZ.

Linn, R., & Kiplinger, V. (1994). *Linking statewide tests to the national assessment of educational progress: Stability of results.* National Center for Research on Evaluation, Standards, and Student Testing Technical Report 375. Los Angeles, CA: University of California–Los Angeles.

Lissitz, R.W., & Samuelson, K. (2007). Suggested change in terminology and emphasis regarding validity and education. *Educational Researcher*, *36*, 8, 437–448.

Long, V.M., Reys, B., & Osterlind, S.J. (1989). Using calculators on achievement tests. *Mathematics Teacher, 82*, 318–325.

Loveless, T. (2004a). *Computation skills, calculators, and achievement gaps: An analysis of NAEP items.* American Educational Research Association Conference, San Diego, CA.

Loveless, T. (2004b). *The 2004 Brown Center Report on American education: How well are American students learning?* Washington, DC: The Brookings Institution Press.

Low, R., & Over, R. (1993). Gender differences in solution of algebraic word problems containing irrelevant information. *Journal of Educational Psychology*, *85*(2), 331–339.

Loyd, B.H. (1991). Mathematics test performance: The effects of item type and calculator use. *Applied Measurement in Education, 4*(1), 11–22.

Lubienski, S.T. (2001). *A second look at mathematics achievement gaps: Intersections of race, class, and gender in NAEP data*. Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA.

Lubienski, S.T. (2002). A closer look at Black-White mathematics gaps: Intersection of race and SES in NAEP achievement and instructional practices data. *The Journal of Negro Education*, *71*(4), 269–288.

Lubienski, S.T., & Shelley II, M.C. (2003, April). *A closer look at U.S. mathematics instruction and achievement: Examinations of race and SES in a decade of NAEP data.* Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

McLarty, J.R., Noble, A.C., & Huntley, R.M. (1989). Effects of item wording on sex bias. *Journal of Educational Measurement 26*(3), 285–293.

Moyer, J.C., Sowder, L., Threadgill-Sowder, J., & Moyer, M.B. (1984). Story problem formats: Verbal versus telegraphic. *Journal for Research in Mathematics Education, 15*(1) 64–68.

Moyer, J.C., Moyer, M.B., Sowder, L., & Threadgill-Sowder, J. (1984). Story problem formats: Drawn versus verbal versus telegraphic. *Journal for Research in Mathematics Education, 15*(5) 342–351.

Mullis, I. V.S., Martin, M.O., Ruddock, G.J., O'Sullivan, C.Y., Arora, A., & Erberber, E. (2007). TIMSS 2007 assessment frameworks. Boston, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

Myerberg, N.J. (1999, April). *The relationship between scoring quality and assessment reliability*. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada.

National Center for Education Statistics (2007). *Mapping 2005 state proficiency standards onto the NAEP scales* (NCES 2007-482). U.S. Department of Education. Washington, DC: Author.

National Mathematics Advisory Panel (2008). Reports of the task groups and subcommittees. Washington, DC: Author.

Neidorf, T., Binkley, M., Gattis, K., & Nohora, D. (2006). *Comparing mathematics content in the National Assessment of Educational Progress (NAEP), Trends in International Mathematics and Science Study (TIMSS), and Program for International Student Assessment (PISA) 2003 assessments* (NCES 2006-029). Washington, DC: U.S. Department of Education, National Center for Education Statistics.

O'Neil Jr., H.F., & Brown, R.S. (1998). Differential effects of question formats in math assessment on metacognition and affect. *Applied Measurement in Education, 11*(4), 331–351.

Perkhounkova, Y., & Dunbar, S.B. (1999, April). *Influences of item content and format on the dimensionality of tests combining multiple-choice and open response items: An application of the Poly-DIMTEST procedure.* Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada.

Pollack, J.M., Rock, D.A., & Jenkins, F. (1992, April). *Advantages and disadvantages of constructed-response item formats in large-scale surveys*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Pollack, J.M., & Rock, D.A. (1997). *Constructed response tests in the NELS: 88 high school effectiveness study. National Education Longitudinal Study of 198, second follow up* (No. NCES 97-804). Chicago, IL and Princeton, NJ: National Opinion Research Center and Educational Testing Service.

Pomplun, M., & Capps, L. (1999). Gender differences for constructed-response mathematics items. *Educational and Psychological Measurement*, *59*(4), 597–614.

Reckase, M.D. (2006). A conceptual framework for a psychometric theory for standard setting with examples of its use for evaluating the functioning of two standard setting methods. *Educational Measurement: Issues and Practice, 25*(2), pp. 4–18.

Reiss, P.P., & Zhang, S. (2006). Why girls do better in mathematics in Hawai'i: A causal model of gender differences on selected and constructed response items. Paper presented at the Annual Meeting of the American Educational Research Association, San Diego.

Romberg, T.A. (1992). Mathematics assessment and evaluation: Imperatives for mathematics education. New York: SUNY Press.

Romberg, T.A. (1992). Problematic features of the school mathematics curriculum. In P.W. Jackson (Ed.). Handbook of Research on Curriculum. New York: Macmillan.

Romberg, T.A. (1992). Perspectives on scholarship and research methods. In D.A. Grouws (Ed.) Handbook of Research on Mathematics Teaching and Learning. New York: Macmillan.

Ryan, K.E., & Fan, M. (1996). Examining gender DIF on a multiple-choice test of mathematics: a confirmatory approach. *Educational Measurement: Issues and Practices*, *15*(4), 15–20.

Sappington, J., Larsen, C., Martin, J., & Murphy, K. (1991). Sex differences in math problem solving as a function of gender-specific item content. *Educational and Psychological Measurement*, *51*(4), 1041–1048.

Schmidt, W., & Houang, R. (2007). Lack of focus in mathematics: Symptom or cause? In T. Loveless (Ed.), *Lessons learned: What international assessments tell us about math achievement.* Washington, DC: Brookings Institution Press.

Schneider, M. (2007). *Response to the "Validity study of the NAEP mathematics assessment: Grades 4 and 8.*" U.S. Department of Education, National Center for Education Statistics. Retrieved November 26, 2007, from http://nces.ed.gov/whatsnew/commissioner/remarks2007/11_23_2007.asp.

Schoenfeld, A. (Ed.). (2007). Assessing mathematical proficiency. New York: Cambridge University Press.

Silver, E., & Kenney, P. (1993). An examination of relationships between the 1990 NAEP mathematics items for grade 8 and selected themes from the NCTM standards. *Journal for Research in Mathematics Education, 24*(2), 159–167.

Traub, R.E., & Fisher, C.W. (1977). On the equivalence of constructed-response and multiple-choice tests. *Applied Psychological Measurement, 1*(3), 355–369.

U.S. Department of Education (2002). *National Assessment Governing Board. Mathematics framework for the 2003 national assessment of educational progress*. Retrieved on October 1, 2007 from http://www.nagb.org/pubs/math_framework/toc.html.

U.S. Department of Education (2004). *National Assessment Governing Board. Mathematics framework for the 2005 national assessment of educational progress*. Retrieved on October 1, 2007 from http://www.nagb.org/pubs/m_framework_05/toc.html.

Webb, D.C. (2001, August). *Classroom assessment strategies to help all students master rigorous mathematics*. Co-presenter for a four-day workshop for District of Columbia Public School mathematics teachers, American Association for the Advancement of Science, Washington, DC.

Wilson, L.D., & Zhang, L. (1998, April). *A cognitive analysis of gender differences on constructed-response and multiple-choice assessments in mathematics*. Paper presented at the Annual Meeting of the American Educational Research Association, San Diego.

Wu, H. (2007). *Fractions, decimals, and rational numbers.* Retrieved on November 1, 2007 from http://math.berkeley.edu/~wu/.

Zieleskiewicz, J. (2000). Subject-matter experts' perceptions of the relevance of the NAEP long-term trend items in science and math. In N. S. Raju, J.W. Pellegrino, M.W. Bertenthal, K.J. Mitchell, and L.R. Jones (Eds.), *Grading the nation's report card: Research from the evaluation of NAEP*. Washington, DC: National Academy Press.

APPENDIX A: National Assessment of Educational Progress (NAEP)

Background

Since 1969 the National Assessment of Educational Progress (NAEP) has been regularly conducting assessments of samples of the nation's students attending public and private schools at the elementary, junior high, and high school levels. NAEP's goal, since its inception, has been to make available reliable information about the academic performance of U.S. students in various learning areas. To this end, NAEP has produced more than 200 reports in 11 instructional areas.

Teachers, administrators, and researchers from across the United States have helped propel NAEP into the valuable informational source it is today. As a result, members of the educational community are able to make use of NAEP's findings on students' learning experiences to inform policymakers and to improve students' educational experiences.

NAEP is an indicator of what students know and can do. Only group statistics are reported, no individual student or teacher data are ever released.

NAEP is conducted under congressional mandate and is directed by National Center for Educational Statistics (NCES) of the U.S. Department of Education. NCES currently contracts with the Educational Testing Service (ETS) to design instruments, and conduct data analysis and reporting; Westat, Inc., to conduct sampling and data collection activities; and National Computer Systems to manage materials distribution, scoring, and data processing.

Who Is Sampled?

Every 2 years, NAEP assesses nationally representative samples of more than 120,000 students in public and private schools in Grades 4, 8, and 12. The NAEP state assessment samples also include students from both public and private schools to be representative of schools in the participating state. Scientific sampling procedures are used to ensure reliable national, regional, and state samples.

Schools

Schools are randomly selected for NAEP based on demographic variables representative of the nation's schools. Trained NAEP staff members administer the assessment. In NAEP state assessments, the participating schools work with a coordinator designated by the respective state department of education to collect information on a statewide level.

Students

Students are selected randomly; their names are not collected. Confidentiality of all participants is ensured and their names do not leave the school.

What Subjects Are Assessed?

The academic subject areas assessed vary from year to year. According to the law authorizing NAEP, all subjects listed under National Educational Goal 3 are to be tested periodically in the national assessment. Reading, writing, mathematics, and science are the most frequently assessed subjects. To minimize the burden on students and schools, no student takes the entire assessment. Instead, assessment sessions are limited to $1\frac{1}{2}$ to 2 hours. Questionnaires are also given to students, teachers, and principals to obtain current information about school and instructional practices that may influence learning and student performance.

When Do Assessments Take Place?

Assessments occur throughout the school year; however, most are conducted January through March. State assessments occur in February.

Source: http://www.nagb.org/about/abt_naep.html.

APPENDIX B: Methodology of the Assessment Task Group

The Assessment Task Group addressed several different kinds of questions related to the influence of different item types and test administration procedures on student responses; the content validity, item types, and item difficulties of the National Assessment of Educational Progress (NAEP) and state tests; and the way NAEP and state performance categories are established. Several different sources of information contributed to the resulting report, each involving somewhat different methodological considerations. These included a review of relevant research literature, elaboration of findings from a recently completed report by the NAEP Validity Studies (NVS) Panel, and an analysis of the content and performance categories of NAEP and selected state mathematics achievement tests.

Literature Review

Literature searches were conducted by Abt Associates Inc. (Abt) to locate studies on mathematics assessment that included the content validity of NAEP, the effect of test administration procedures, the influence of item wording, and the skills and concepts captured by various item types. The criteria for selecting relevant studies required that they a) be published between 1970 and 2007 in a journal, government or national report, book, or book chapter; b) involve K–12 mathematics assessments; c) be available in English; and d) use quantitative methods for analyzing data. Because of the diversity of pertinent topics and associated forms of research, no other general methodological criteria were imposed but, rather, the Task Group made individual judgments about the appropriateness and quality of each candidate study located in the search.

Electronic searches were made in Education Resource Information Center (ERIC), PsycInfo, and the Social Sciences Citation Index (SSCI) using the search terms identified by the Assessment Group, shown below.

Assessment or testing *and* math *and* each of the following:

item language	verbiage	short answer
item wording	excessive language	true-false
question language	validity	administration procedure
question wording	reliability	manipulatives
item wordiness	item type	calculators
language bias	item structure	formulas
linguistic simplification	multiple choice	accommodations
language density	constructed response	Bloom’s taxonomy
essential language	open response	bias

Additional studies were identified through manual searches of relevant journals, Internet search engines, and reference lists and recommendations from experts. Abstracts from these searches were screened for relevance to research questions and appropriate study design. For studies deemed relevant, the full study report was obtained. Citations from those articles and research reviews were also examined to identify additional relevant studies. Abt extracted summary information from the qualifying studies and provided it to the Task Group along with the complete articles. The Task Group then further screened the studies to narrow those included in their reviews to the most relevant and highest quality available.

The Task Group drew on the studies located and screened through this procedure for its reviews of the content validity of NAEP mathematics assessment, the influence of item format on test performance, and calculator use during mathematics achievement assessments.

Analysis of NAEP and State Test Mathematics Items

The Task Group's assessment of the mathematics items used in NAEP and state achievement tests initially drew on a report of a validity study released in the early fall of 2007 (Daro et al., 2007). The Task Group was briefed on the NAEP Validity Study by the authors, given access to an embargoed early version of the report, and shown the response of National Center for Education Statistics (NCES) to that report. The Task Group then conducted its own further analysis of the items in the six state tests represented in the NVS sample.

The IDA STPI (Institute for Defense Analyses Science and Technology Policy Institute) provided the Task Group with information on the test frameworks, testing procedures, and test items for the six state mathematics tests used in the NVS report: California, Georgia, Indiana, Massachusetts, Texas, and Washington. STPI collected the state assessment information it provided to the Task Group from each state's Department of Education Web site and the NAEP information from the U.S. Department of Education's National Center for Education Statistics Web site. STPI assembled the relevant material it located in response to the Task Groups request but did not conduct any analyses of that material.

One of the mathematicians on the Task Group then analyzed the released items for the Grade 4 and Grade 8 state tests and NAEP provided in the STPI material. The results of that analysis were then further reviewed by the other Task Group members and incorporated in the Assessment Report to supplement the analysis by five mathematicians that was reported by the NVS.

Analysis of the Content and Performance Categories of NAEP and State Mathematics Tests

The material provided to the Task Group by STPI on the test frameworks, testing procedures, and test items for the six state mathematics tests used in the NVS report included descriptions of the content of each test, the performance categories, and the procedures for establishing the performance categories. STPI collected this information from each state's department of education Web site and the NAEP information from the NCES Web site. The reports and descriptive summaries they provided were reviewed by the members of the Task Group and used, along with studies from the literature review, as the basis for their analysis of these topics.

APPENDIX C:
Test Content Frameworks and Items: A Review

Ten studies that assessed, in various ways, the content validity of the mathematics portion of the National Assessment of Educational Progress (NAEP) for Grades 4 and 8 were reviewed for this portion of the report, most of which are reviewed in this section. Two of the studies (Daro et al., 2007; Kenney et al., 1998) compared the NAEP framework with the framework of other mathematics assessments (among other topics). Five of the studies (Kenney et al., 1998; Loveless, 2004b; Neidorf et al., 2006; Silver & Kenney, 1993; Zieleskiewicz, 2000) compare NAEP items to frameworks from other mathematics assessments. Three studies (Daro et al., 2007; Silver & Kenney, 1994; Silver et al., 1992) compared NAEP items to the NAEP framework. The remaining three studies used a variety of methods to explore NAEP content and items. Kenney (2000) discussed the rationale for creating families of items and demonstrates the creation of families with released NAEP items. Linn and Kipplinger (1994) tested whether an equating function could be developed to equate standardized achievement test scores to NAEP scores.

Test Content Frameworks

Daro et al. (2007) convened an expert panel involving mathematicians, mathematics educators, and an expert on state-based mathematics standards. They compared the NAEP mathematics framework with the standards and frameworks (test blueprints) of six states (California, Massachusetts, Indiana, Texas, Washington, and Georgia), two high-performing nations (Singapore and Japan), and standards outlined by the National Council of Teachers of Mathematics (NCTM) and Achieve, Inc. In examining the content areas of Number Properties and Operations, Algebra, Geometry, Measurement, and Data Analysis and Probability in the 2005 NAEP mathematics framework, the reviewers attempted to determine if NAEP was missing something or overemphasizing topics in a given content area. The reviewers then described what was being overemphasized and rated the emphasis of each content topic as compared to each of the six states and Singapore.

Item Comparisons within Content Frameworks

Daro et al. (2007) indicated that, at the fourth-grade level, the only area where NAEP has a higher percentage of items than the other frameworks was Measurement. It also was noted that, while Number Properties and Operations is the most emphasized content area at the fourth-grade level, the NAEP provides a very limited assessment of fractions at this level. The NAEP Geometry items assess symmetry and transformations more than the other states, and emphasize parallel lines and angles less than the comparison states. Moreover, the fourth-grade NAEP Algebra content area appears to be especially problematic. The pattern items overemphasize sequences of numbers that grow in a regular way; and, this type of pattern is used in NAEP more than in the other frameworks. Mathematics reviewers suggested that NAEP consider pattern items based on the relationship between two quantities. The review panel recommended better item balance within the algebra subtopic of patterns, relations, and functions at this level.

The review of the eighth-grade NAEP Number Properties and Operations content area found an emphasis on topics from number theory—factorization, multiples, and divisibility. Given this review, a focus dedicated to ensuring that eighth-grade students have developed proficiency with whole numbers, negative integers, and fractions, decimals, and percent may be considered given their importance as prerequisites for algebra. Because this content area stood out in the review as undersampling grade level content, “It is possible that students are making gains in this content area that are not being detected by NAEP” (p. 123). In the Panel’s judgment, it is also possible that students are losing ground that goes undetected. Indeed, because the NAEP minimizes this area, this could be a driving force for reduced attention to it within the school curriculum.

The eighth-grade Measurement content area appears to be assessing lower-level concepts and skills and, as a result, NAEP may be underestimating achievement. It also was noted that the larger number of measurement items is “not well leveraged to include fractions or decimals used in realistic situations” (p. 126). The review of the Geometry items indicated wide variation across the six states. The NAEP at Grade 8 includes more geometry than the comparison states or nations. A consensus does not appear to exist on what is important in geometry at Grade 8.

Loveless (2004b) found that the majority of the fourth- and eighth-grade NAEP items assessing problem solving, algebra and numbers sense involve whole numbers. While this is understandable at the fourth-grade level, it is cause for serious concern at the eighth-grade level. Fractions, decimals, and percent are under-assessed. In items assessing problem solving, whole numbers make up approximately 72% of the fourth-grade items and approximately 70% of the eighth-grade items. The possible overemphasis regarding whole numbers continues for the eighth-grade NAEP algebra items as well. This suggests again raising the level of arithmetic to include more direct assessment of fractions, decimals, and percents within the number and algebra content areas and that the confinement of arithmetic to whole numbers is largely responsible for the low grade-level demands of many of the items Loveless also questions the identification of some of the items as algebra.

Neidorf et al. (2006) compared the mathematics content in NAEP, Trends in International Mathematics and Science Study (TIMSS), and the Program for International Student Assessment (PISA). They note that the NAEP and TIMSS content frameworks are quite consistent with regard to their basic organization of mathematics content. They both have five main content areas: Number, Measurement, Geometry, Data, and Algebra. They did note different emphases within topics and subtopics and in some grade level expectations. PISA differs from both NAEP and TIMSS in that it samples 15-year-olds rather than specific grades and that it focuses on problem-solving using what are called, in the world of K–12 education, “real-world” problems, rather than curriculum content areas. However, the mathematics content assessed by PISA is consistent with the NAEP eighth-grade mathematics framework.

Neidorf et al. (2006) noted that PISA has more items classified as Data Analysis and fewer as Algebra and Number Sense than the other two assessments. The NAEP and TIMSS comparison indicates that there is a greater emphasis by NAEP on applications. Moreover, TIMSS includes a higher proportion of items involving ratio and proportion and, thus, has a more appropriate balance for assessing number using fractions, decimals, and percent. While there is considerable overlap in the NAEP and TIMSS assessments involving measurement, there is greater emphasis in NAEP on using measurement instruments and units of measurement. TIMSS included a higher percentage of items involving estimation, calculation, or comparing perimeter, area, volume, and surface area. With regard to data, NAEP has a greater proportion of probability items, whereas TIMSS has a greater proportion of items than NAEP that emphasize reading, interpreting, and making predictions from tables and graphs and data representation, especially at the fourth-grade level. Finally, in NAEP, "mathematical reasoning" is included in making conjectures and other related subtopics. This is not the case in TIMSS.

The Task Group notes that the TIMSS content domains were recently changed (Mullis et al., 2007). The Grade 4 content domains are now identified as Number, Geometric Shapes, and Measures and Data Displays. At this level, TIMSS has merged Geometry and Measurement and deleted the domain formerly called Patterns, Equations, and Relationships. The Grade 8 content domains are Number, Algebra, Geometry, and Data and Chance. At this level, TIMSS has infused Measurement within Geometry and expanded Data to include Probability.

Kenney et al. (1998) compared the mathematics portion of the 1996 NAEP and Maryland State Performance Assessment Program (MSPAP) at the eighth-grade level as part of the Content Analysis Project supported by National Assessment Governing Board (NAGB). It should be noted that the MSPAP is no longer used as Maryland's eighth-grade assessment due to a host of problems. Nonetheless, based on a comparison of the content frameworks, there was moderate congruence regarding the content characteristics of the MSPAP and NAEP Grade 8 tests. Content areas and topics were similar; however, the similarity was more evident in some content areas than in others. For instance, the Measurement items were nearly identical. The differences between the two tests are not sufficient to account for the magnitude of the difference between proficient performance on the MSPAP (48%), a high-stakes assessment, and on the NAEP (24%). This is likely the result of different performance categories.

Zieleskiewicz (2000) completed a study that involved 30 raters who were selected to evaluate math items on the long-term trend NAEP and the main NAEP. The reviewers felt that both the long-term trend and main NAEP frameworks assess important mathematics, with little variation across the types of raters, which included classroom teachers and mathematics specialists (e.g., university professors, leaders in professional organizations, assessment specialists).

Linn and Kiplinger (1994), moreover, in their work linking statewide tests to NAEP, found substantial content differences between the state tests and NAEP, with the majority of the statewide test items falling into one of the NAEP content areas—Number and Operations. They note that, if linking state assessments and NAEP is a goal, tests should be developed with a common framework.

Finally, Kenney (2000) reviewed the rationale for creating a family of items about a specific topic. She suggests that the ideal method for creating an item family for the NAEP would be to begin with the topic (e.g., algebra) and information based on research about students' understanding of the topic. A family of items would be built based on theoretical grounds and validated by examining results from tests. It was proposed that item families would increase NAEP's potential to provide important information about the depth of students' knowledge in a particular content strand or across content strands. It was suggested that research could support creating item families on fractions, decimals, probability, with vertical item families assessing depth in these content areas. Proportionality in measurement, geometry, and number would be a horizontal item family that would assess an important concept (proportionality) across content areas.

These studies guided the Task Group's thinking when developing the principles for organizing the content of the NAEP and state tests. Together, they form the rationale for any recommendations drawn from the general principles.

APPENDIX D: Establishing Performance Categories

Establishing performance categories involves a set of procedures currently known in educational measurement as standard setting (or setting cut scores). Judgments about performance categories are made by a panel of persons selected for their expertise or educational perspective. The exact procedures to classify students' test scores into performance categories can range from a panel consensus global judgment about the test as a whole (i.e., the minimum percentage of items passed at the various levels) to quantified judgments of individual items with respect to expected performance of students in the categories.

Several procedures and methods for combining judgments in standard setting have been developed. These procedures typically involve training panelists on the definitions of the standards and the nature of performance within the categories, soliciting judgments about the relationship of the test to the performance categories and providing successive feedback to the panelists about their judgments. Various methods to combine judgments have been developed. Variants of the Bookmark method and the Modified Angoff method involve panelists judging how students at varying levels of competency will respond to representative test items. In these two methods, the cut score for competency classifications is determined by linking the judgments to empirical indices of item difficulty. In contrast, the Body of Work method requires the panelist to classify representative students into competency categories by examining their full pattern of item responses. While the methods were all scientifically acceptable, they may differ in effectiveness. The Bookmark method may involve the most assumptions about the data, while the Body of Work method may demand the highest level of rater judgment. While more research is needed in this area, the Modified Angoff method performs well against several criteria for psychometric adequacy (Reckase, 2006).

The Task Group was interested in the following questions about standard setting in NAEP and the six states:

1) What are the performance categories of NAEP and the states?
2) How were the NAEP and state performance categories established?
3) Are they based on procedures in which experts inspect actual item content or on global definitions? (Definitions are characterized as "global" when fairly abstract characterizations of behavior necessitate high degrees of judgment to determine the categorization of student performance.)
4) Are empirical procedures used to combine individual expert opinions?
5) What is the background of the experts?
6) What descriptions or instructions are given, if any, about the nature of performance at different levels?
7) Do the experts receive the items in an examination under the same conditions as the students?

Method

To answer these questions, documents available from Web sites of NAEP (National Assessment Governing Board) and six states (California, Georgia, Indiana, Massachusetts, Texas and Washington) were retrieved by Institute for Defense Analyses Science and Technology Policy Institute (STPI) and provided to the Task Group. These documents were reviewed for relevant data by the Task Group members.

Results

Table D-1 shows the performance categories and definitions given by the NAEP and six states that were studied. Information was not fully available on all questions for each state.

Table D-1: Standard-Setting Procedures of NAEP and Six States

	Performance Categories	**Definitions**
NAEP	Basic, Proficient, Advanced	Global
California	Far Below Basic, Below Basic, Basic, Proficient, Advanced	Global and by Area
Georgia	Does Not Meet, Meets, Exceeds Standard	Global
Indiana	Did Not Pass, Pass, Pass+	Global, brief
Massachusetts	Warning, Needs Improvement, Proficient, Advanced	Global
Texas	Basic, Proficient, Advanced	Global
Washington	Basic, Proficient, Advanced	Global

The first question that was examined was the definitions of performance categories on NAEP and the six states. NAEP and all six states employed a three category system, although the labels varied somewhat. NAEP's performance categories are Basic, Proficient, Advanced. California's performance categories are labeled as Below Basic, Basic, Proficient, Advanced; Georgia, Does Not Meet, Meets, Exceeds Standard; Indiana, Did Not Pass, Pass, Pass+; Massachusetts, Warning, Needs Improvement, Proficient, Advanced; Texas, Basic, Proficient, Advanced; and Washington, Basic, Proficient, Advanced. For NAEP and all states, global definitions of the performance categories are available. Data on the NAEP and six states are tabulated in Table D-2.

For question number 2, several different standard-setting (or setting cut scores) methods have been developed over the last decade. The most widely used methods involve a generally similar standard-setting process. That is, the standard-setting process begins with a training session for the panelists, focusing on the definitions of the standards and the relevant behaviors. Then, the panelist was asked to rate, categorize, or set cutlines, depending on the exact standard-setting method. The process is iterative, with feedback about the results given and opportunities to revise judgments.

Three different standard-setting methods were employed in the states for which information was available. Historically, the Bookmark method (Lewis, Mitzel, & Green, 1996) is the most widely used method. Prior to the standard-setting process, items are ordered by their empirical difficulty in the item response theory metric. Then, the panelist sets marks in the ordered set of items to designate the points at which the minimally competent student in a category (e.g., Basic, Proficient, Advanced) is more likely to pass than to fail the item (often defined as a probability of .67). Although panelists are instructed to examine all items, items near the marks probably receive the most scrutiny. Item mapping is a modified Bookmark method, which differs somewhat in the standard-setting process as compared to the standard Bookmark method. The Modified Angoff method requires each panelist to consider each test item and to estimate for each item what percentage of students who minimally qualify for a category (e.g., "meets standards") would answer the item correctly (this is also referred to as assigning a p-value). This method involves empirically aggregating ratings and giving feedback to panelists, followed by opportunities to revise ratings. Because each item must be rated, close scrutiny of each item is required. The Body of Work method is more holistic, because panelists examine test protocols for students at varying score levels. Their material includes item content, actual item responses, and the scoring rubrics. The panelist's task is to determine which students fall in the various categories.

As summarized by Karantonis and Sireci (2006), scant research is available on how the popular Bookmark method compares to other methods. Thus, insufficient empirical evidence is available to recommend it over the other methods. Further research should be conducted, and variables such as reliability across panelists, exact item content, domain multidimensionality, as well as resulting levels set for the standards, should be examined.

For question number 3, the standard-setting (or setting cut scores) methods reviewed by the Task Group all involve the actual inspection of item content by the panelists. However, some methods involved more intensive consideration of item content than others. In particular, the Modified Angoff method requires judgments for each item. The Bookmark method involves discussion of items, but the quantified judgments are for the category distinctions. Items with more extreme difficulties may be not considered extensively. In the Body of Work method, items are given but they are not judged.

For question number 4, judgments that are elicited from the panelist may be combined empirically. In practice with the various methods, judgments are often taken repeatedly and combined, thus allowing feedback and possible revision of judgments.

For question number 5, the background of the experts used to set standards varies within panels and possibly between states. Classroom teachers may be predominantly represented, but other experts, such as curriculum experts from higher education, may be present. Further, community representatives also may be panelists.

For question number 6, the standard-setting process for the methods described above typically involve extensive instruction about the definitions of the standards and the procedures used to set standards. Such instructions are expected to have substantial impact on the judgments. This question was scored separately because states may deviate from typical procedures or methods.

For question number 7, the experience of actually taking test items not only serves to establish the panelist's understanding of the subject area test items but also to have the experience of the students who take the tests. Judgments of items that are viewed under operational conditions are based more on individual information than on panel consensus.

All states for which information was available applied one of the current standard-setting (or setting cut scores) methods. The data in Table D-2 can be summarized as follows. First, although there is variability in the methods, all states use a contemporary method for standard setting. The Bookmark method was most frequently applied in standard setting. Second, item content is judged in all states except Massachusetts. Third, empirical combination of judgments is implemented in all states. Fourth, the background of the experts varies within panels and probably somewhat across states. For example, Georgia uses primarily classroom teachers as experts while Texas represents broader contingencies, includes curriculum experts from higher education and non-educators. Fifth, all states train the panelists prior to eliciting their ratings. Finally, only two states have the panelists experience the items in the same way as the test-takers.

Table D-2: Information on Features of Standard-Setting Procedures (Setting Cut Scores) for NAEP and the Six States

	1. How Established?	2. Item Content Judgments?	3. Combination Procedures	4. Background of Experts	5. Instructions & Definitions	6. Test Taken?
NAEP	Modified Angoff Method	Yes	Empirical with successive feedback.	55% teachers, 15% non-teacher educators, and 30% members of the general public. Panelists should be knowledgeable in mathematics. Panelists should be familiar with students at the target grade levels. Panelists should be representative of the nation's population in terms of gender, race and ethnicity, and region.	Yes	N/A
California	Bookmark Method	Yes	N/A	N/A	Yes	Yes
Georgia	Modified Angoff Method	Yes	Empirical with successive feedback.	Primarily the panelists selected were educators currently teaching in the grade and content area for which they were selected to participate.	Yes	Yes
Indiana	Bookmark Method	Yes	Empirical preliminary followed by feedback & consensus.	Not specifically given, but appears to be classroom teachers.	Yes	None specified. Probably first viewed in panel setting.
Massachusetts	Expert Opinion – Body of Work Method	No	Empirical aggregation of first judgments. Details not available about feedback & consensus.	The panel consists primary of classroom teachers, school administrators, or college and university faculty, but also non-educators including scientists, engineers, writers, attorneys, and government officials.	Yes	None specified. Probably first viewed in panel setting.

Continued on p. 8-57

Table D-2, continued

	1. How Established?	2. Item Content Judgments?	3. Combination Procedures	4. Background of Experts	5. Instructions & Definitions	6. Test Taken?
Texas	Item-mapping	Yes	Empirical preliminary followed by feedback & consensus.	The majority of the panelists on each committee were active educators—either classroom teachers at or adjacent to the grade level for which the standards were being set, or campus or district administrative staff. All panels included representatives of the community "at large."	Yes	None specified. Items probably first viewed in panel setting.
Washington	Bookmark Method	Yes	Empirical preliminary followed by feedback & consensus.	The majority of the panelists on each committee were active educators—either classroom teachers with some representation of higher education.	Yes	Yes

Source: This table was created by the Task Group using publicly available data from state Web sites. Data on California is from S. Valenzuela (personal communication, February 1, 2008).

Discussion

Although the NAEP and states varied in both process and method for standard setting (or setting cut scores), all states for which information was available employed currently acceptable educational practice. The methods may differ in effectiveness; however, scant evidence is available. The Bookmark method may involve the most assumptions about the data, while the Body of Work method may demand the highest level of judgment from the raters. The Modified Angoff method is preferred (Reckase, 2006) because the assumptions of the Bookmark method (e.g., unidimensionality) are probably not met in practice. The Body of Work method is less often applied to year-end tests because it requires higher levels of judgments from the experts. More research is needed on the standard-setting process.

It was found that classroom teachers, most of whom are not mathematics specialists, predominate in the standard-setting process. Higher levels of expertise, including the expertise of mathematicians, as well as mathematics educators, high-level curriculum specialists, classroom teachers and the general public, should be consistently used in the standard-setting process. The Task Group also found that the standard-setting panelists often do not take the complete test as examinees before attempting to set the performance categories, and that they are not consistently informed by international performance data. On the basis of international performance data, there are indications that the NAEP cut score for performance categories are set too high. This does not mean that the test content is too hard (sufficient mathematical item complexity).

APPENDIX E: Item Response Format and Performance on Multiple-Choice and Various Kinds of Constructed-Response Items

Introduction

Constructed-response (CR) item formats, in which the examinee must produce a response rather than select one, are increasingly utilized in standardized tests. One motivation to use the CR format arises from its presumed ecological validity by more faithfully reflecting tasks in academic and work settings, and stressing the importance of “real-world” tasks. CR formats also are believed to have potential to assess dynamic cognitive processes (Bennett, Ward, Rock & Lahart, 1990) and principled problem solving and reasoning at a deeper level of understanding (Webb, 2001), as well as to diagnose the sources of mathematics difficulties (Birenbaum & Tatsuoka, 1987). Finally, CR formats also may encourage classroom activities that involve skills in demonstrating problem-solving methods, graphing, and verbal explanations of principles (Pollack, Rock & Jenkins, 1992). However, these purported advantages can incur a cost. The more extended CR formats require raters to score them. Hence, they are more expensive and create delays in test reporting.

In contrast, multiple-choice (MC) items have been the traditional type used on standardized tests of achievement and ability for over a century. MC items can be inexpensively and reliably scored by machines or computers, they may require relatively little testing time, and they have a successful history for psychometric adequacy.

The Assessment Task Group examined the literature on the psychometric properties of constructed-response items as compared to multiple-choice items. The original focus was to address the following three questions:

1) Do the contrasting item types (e.g., MC, CR) capture the same skills in these tests equally well?
2) What does the scientific literature reveal?
3) What are the implications for National Assessment of Educational Progress (NAEP) and state tests?

Methodology

Constructed-response formats. CR items vary substantially in the amount of material that an examinee must produce. There are three basic types of CR items:

- The *grid-in* constructed-response format (CR-G) requires the examinee to obtain the answer to the item stem and then translate the answer to the grid by filling in the appropriate bubble for each digit.
- The *short answer* constructed-response format (CR-S) varies somewhat. The examinee may be required to write down just a numerical answer or the examinee may need to produce a couple of words to indicate relationships in the problem. The CR-S format potentially can be scored by machine or computer, given a computerized algorithm that accurately recognizes the varying forms of numerical and verbal answers. Further, an intelligent algorithm also may provide for alternative answers (e.g., slightly misspelled words, synonyms).
- The *extended* constructed-response format (CR-EE) requires the examinee to provide the reasoning behind the problem solution. Thus, the CR-EE format would include worked problems or explanations. This format readily permits partial credit scoring; however, human raters are usually required. Use of human raters, however, can lead to problems with consistency and reliability of scoring.

The stems of CR-G and CR-S items and MC items can be identical, especially if the correct answer is a number. It is not clear how the stems of CR-EE can be identical to MC items, although this possibility cannot be excluded.

Coding. With the Task Group's guidance, Abt Associates Inc. (Abt), a research group hired to assist the National Mathematics Advisory Panel, developed a list of variables to code for identified studies on this topic. The coding scheme is as follows:

Table E-1: List of Variables for Coding Studies

Category	Description
Citation	Reference citation.
Purpose of Study	Brief summary of the purpose/focus of the publication.
Content area	Studies that were not about math were excluded. Give more information on type of math if available.
Assessment	Assessment being investigated, if available.
Grade level	Grade being investigated. If not provided, age or other grouping characteristic (e.g., high school).
Item type(s) investigated	CR-G = constructed response, grid format CR-SR = constructed response, short response CR-EE = constructed response, extended essay CR-O = constructed response, other—provide details MC = multiple choice Other = other—provide details
General description of design	Brief summary of study design.
Number of items	Number of items for each type included in the study.
Description of Sample	Sample size, sampling technique (e.g., random, matrix, stratified, purposive), source of sample (e.g., national, region, locale, school district), specific characteristics (e.g., college-bound, general population, special population).
Appropriateness of sample	Enter Y if sample is representative of test taking population. Enter N if sample is not representative of test taking population and describe gap.
Source of items	Operational test, specially constructed for study, etc.
Summary of findings	Brief summary of findings. Possible reference to more detail in original text.
Subgroup performance	English language learners, gender, race/ethnicity, etc.
Psychometric properties	Item/test reliability, item/test difficulty (p-value), differential item functioning (DIF) results.
Information on scoring	Information on scoring of test (e.g., rubrics, criterion-referenced, norm-referenced).
Design flaws	Describe any obvious design flaws.

Search procedures. Abt used key words to search the literature and identify a broad band of potentially relevant research to all research questions addressed by the Task Group. Abt identified 161 articles that were potentially relevant for the specific research question on item format. They were then screened for several criteria: 1) inclusion of comparisons based on mathematics items, 2) presentation of empirical evidence, 3) published as a document other than a conference paper, 4) relevancy to the research question, 5) had the appropriate grade level or assessment [i.e., nothing higher than Advanced Placement (AP) or SAT, and 6] availability of the article.

Abt then extracted information from the 31 articles that remained and provided it to the Assessment Task Group. The full articles also were provided. An examination of the 31 articles by the Task Group led to further restriction of the set for the following reasons: 1) technical reports that were superseded by a published version, 2) irrelevant purpose for this specific research question, 3) inappropriate sample, and 4) inclusion of only MC or CR items, but not both. Ten additional articles were excluded; thus, 21 relevant articles were available to address the question on item format. To analyze the results, the 21 articles were examined in detail by the Task Group for relevant data on the several issues of concern.

Results

Impact of response format on mathematical skills, knowledge, and strategies. Potentially the most pressing issue about response format is the extent to which the same skills, knowledge, and strategies can be measured by the MC and CR item formats. Traub and Fisher (1977) found that stem-equivalent MC and CR mathematics items measured the same construct on a national math achievement test. That is, items from the two formats loaded on the same factor. Further, the skills and abilities measured by separate tests (i.e., ability to follow directions, recall and recognition memory, and risk-taking) had similar correlations with mathematics scores based on the two formats. Behuniak et al. (1996) also found that items in MC and CR format loaded on the same factor, but CR items were significantly more difficult. In contrast, Birenbaum et al. (1992) found a format effect in which there were larger performance differences on stem-equivalent MC and CR items than on parallel (i.e., superficially different) items in the same task format. Pollock and Rock (1997) examined National Education Longitudinal Study (NELS) data and found that MC items loaded on a different factor than CR items, although the factors correlated highly ($r = .86$). DeMars (1998) found that competencies on a state achievement test that were calculated from MC items versus CR items correlated in the .90s when the measures were statistically corrected for unreliability.

Katz, Bennett, and Berger (2000) studied strategy choice for stem-equivalent MC and CR items by analyzing verbal reports of problem solving processes during item solving, using “talk-aloud” procedures. Katz et al. (2000) found that the “plug-in strategy,” which is usually associated with MC items, was used nearly as commonly with CR items. For MC items, examinees “plug-in” numbers from the response alternatives to identify the key. Students were observed adapting the plug-in strategy to CR items by estimating potential solutions and plugging-in numbers. O’Neil and Brown (1998) administered a questionnaire following the administration of a state standardized achievement test that contained both CR and MC items. Students reported greater use of systematic cognitive strategies for CR items than for MC items. However, they reported greater self-checking activity for MC items.

Thus, these studies generally do not support major differences in the nature of the construct that is measured by CR and MC items, nor in the strategies that are applied. However, much more data on this issue is potentially available because many state accountability, graduation, and year-end tests employ both item formats.

Impact of response format on psychometric properties. Several studies have results that are relevant to the psychometric properties of CR and MC items. Specifically, the psychometric properties that have been examined include item difficulty, item discrimination, omission rates, and differential item functioning (DIF) by diverse subgroups. The results vary somewhat over the exact type of CR format that is used.

The psychometric properties of MC items to their equivalent CR-G format were compared in three studies. Behuniak, Rogers, and Dirir (1996) found a moderate effect size (eta^2 of .057), which indicated that the CR-G items were harder. However, item format was unrelated to item discriminations and to gender-related DIF. Burton (1996) was interested primarily in the impact of item format on gender differences with mathematics items. She reports only DIF as item-level statistics and found that MC items exhibited no gender-related DIF, while the CR-G format had very small and inconsistent DIF. Hombo, Pashley, and Jenkins (2001) found that CR-G items were more difficult than the MC stem-equivalent items for most items. They also found sizable differences between the grid-in responses and their accompanying written responses, suggesting that examinees have difficulty translating their answer to the grid. Hombo et al. (2001) did not report results on item discrimination or DIF.

In summary, these studies are consistent in supporting the conclusion that the CR-G format leads to higher levels of item difficulty as compared to stem-equivalent MC items. Yet, the results suggest that some, but not all, of the increased difficulty may be attributable to examinee difficulties in translating answers to grids. These studies do not provide support for format differences in item discrimination or gender-related DIF. Again, it should be noted that the number of studies yielded by the search procedures is very small.

Other studies comparing the psychometric properties of MC and CR items use the CR-S response format, which may be either in written or verbal format. Birenbaum, Tatsuoka, and Gutvirtz (1992) and Birenbaum and Tatsuoka (1987) found that stem-equivalent MC items were more difficult and less discriminating than CR-S items. However, because the MC items were constructed from previously administered CR-S items, they were able to contain common CR errors as distractors. This item design presumably minimizes the feedback received by examinees when their calculated answer does not appear as an alternative. Katz, Bennett, and Berger (2000) find MC items somewhat more difficult (proportion correct of .78 and .75 versus .66 and .74, with difference between .78 and .66 being statistically significant) than their stem-equivalent CR-S items. However, they note that large differences between item formats occurred when the MC stem-equivalent item did not have distractors representing common errors. Thus, MC items may be substantially easier than their CR-S counterparts because feedback about incorrect answers may have been provided.

In other studies, operational test data are used to examine the psychometric properties of MC and CR items. These comparisons do not involve specially constructed items (e.g., no stem-equivalent items) to control for other design differences. The nature of the MC items and the CR items that were compared is typically less well specified. In fact, the CR items may have been expressly constructed to represent other aspects of performance. DeMars (1998) found CR items on a low-stakes form of a state high school proficiency test more difficult than MC items. DeMars (2000) also found a similar format effect in a study that included both low-stakes and high-stakes high school proficiency tests. Koretz, Lewis, Skewes-Cox, and Burstein (1993) found that the omit rates are higher for CR items than for MC items on the National Assessment of Educational Progress (NAEP), which can lead to greater apparent difficulty for the CR items. Dossey et al. (1993) reported that, although NAEP CR-S items have proportions correct in the target range of .40 to .60, CR-EE items are very difficult. Garner and Engelhard (1999) also reported that the CR items on a state achievement test are more difficult than MC items, but the CR items exhibited less gender-related DIF.

In summary, the evidence about the psychometric properties of CR items as compared to MC items is inconsistent and depends on the source and the design of the comparison. If the studies utilize operational test data, comparisons of MC and CR items have indicated greater omit rates and greater difficulty for the CR items. This pattern is probably repeated on many state tests and would be a strong finding if such data were available for study by the methods employed in this study. It should be noted, however, that studies on operational test items were not designed to isolate the impact of format by controlling or measuring other properties of items. If the studies utilized stem-equivalent versions of MC and CR items, the difference in psychometric properties depended on other design features of the items, such as the nature of the distractors and the use of grid-in responses. For example, some studies have found the CR format to be more difficult, which is consistent with the operational test studies. Other studies, however, have found the MC items to be more difficult when the distractors are constructed to represent common error patterns. Moreover, little evidence from any design is available to support differences between MC and CR items on item discrimination levels, DIF, strategy use and the nature of the construct that is measured.

Impact of response format on differences between groups. Finally, the impact of response format on differences between diverse groups has been examined in several studies. The most information is available on how item format might differentially affect mathematics performance of males and females. Historically, boys score higher than girls on many tests of mathematical competency (see the Learning Processes Task Group report). Hastedt and Sibberns' (2005) analysis of Trends in International Math and Science Study (TIMSS) data indicated that the magnitude of gender-related differences depended on item format, with girls scoring relatively higher on the CR item format. DeMars (1998) found that the interaction of gender with response format differed between two forms of a low-stakes state high school proficiency test, which included both CR-S and CR-EE as well as MC items. On one form, no significant interaction was found, while on the other form a small significant interaction was observed, indicating that girls scored relatively higher on the CR items, but still not as high as boys. DeMars (2000) examined both low-stakes and high-stakes high school proficiency tests and found that gender interacted with item format; namely, girls scored relatively higher on the CR format while the MC format favored boys. Although Gallagher (1992) also found that gender differences were greater on CR items (i.e., boys performing better), her comparison was based on high-ability students only. Garner and Engelhard (2000), moreover, examined a state high school graduation test. They found small gender-related differences on MC items, favoring boys, and smaller gender-related differences on CR items, but again favoring boys. Pollock and Rock's (1996) analysis of NELS data found that performance of males and females did not vary as a function of MC versus CR items. Thus, while girls scored lower, this was not due to item format. Burton (1996), moreover, found that the changes in item content on math section of the Math Scholastic Aptitude Test (SAT-M), among which was the inclusion of CR-G items, did not impact gender-related differences in quantitative scores, which traditionally has favored boys. Finally, Koretz, Lewis, Skewes-Cox, and Burstein (1993) report that gender-related differences in omitting either MC items or CR items were infrequent on NAEP.

Thus, the results on the interaction of the magnitude of gender-related differences in performance and item format are inconsistent and depended on the design of the specific study. However, the evidence suggests either no impact of response format on gender-related differences or that the relatively lower scores of girls than boys on mathematics items may be lessened in the constructed-response format.

The interaction of racial-ethnic differences with item format also has also been examined in several studies. Historically, minority groups score lower on tests of mathematical performance (see the Learning Processes Task Group report). In DeMars's (2000) study, proportion minority interacted with format, indicating that black-white differences were greater on MC items. Pollock and Rock's (1996) analyses of NELS data, moreover, indicated that demographic variables based on race-ethnicity (black versus white, Hispanic versus white) correlated more highly with the MC item factor than the CR item factor, indicating greater adverse impact on MC items. Finally, Koretz, Lewis, Skewes-Cox, and Burstein (1993) found that racial-ethnic groups differed on the relative rate of omitted items for MC and CR items on NAEP. Thus, as a set, these studies provide some evidence that black-white differences in performance in mathematics are lessened on CR item format as compared to the MC item format.

Other results on item format that are potentially interesting include Hastedt and Sibberns (2005) finding on TIMSS data that scores based on MC versus CR items produced only slight differences in the relative ranking of the various participating countries. And, DeMars (2000) found that the difference between MC and CR items in difficulty depended on the test context. The two item formats differed less in difficulty on high-stakes tests than on low-stakes tests.

Discussion

The available evidence on comparing the psychometric properties of MC items and CR items must be interpreted in the context of several factors. These factors include the following limitations: 1) the limited scope of the available scientific literature, 2) the uncontrolled design features for comparisons based on operational tests, 3) the design strategy in available controlled comparisons of MC and CR items, 4) the limited scope of the controlled comparisons, and 5) the impact of test context on the relative performance on MC and CR items.

First, the literature that could be retrieved by the methods in this study did not yield many journal articles and widely circulated technical reports. Yet, analyzing the psychometric properties of test items is routine test development procedure for many state and national tests, many of which contain both MC and CR item formats and most of which contain demographic information on examinees. It is unclear if these results are unavailable due to lack of appropriate publication outlets, lack of incentives to provide results, or some combination of these features. Despite some limitations in these comparisons (discussed further in this section), it would be useful to know if the CR items on current tests measured the same construct, had greater difficulty but equal discrimination and DIF, and result in lessened adverse impact on some groups of test takers as compared to MC items. Given the lack of evidence from the wider sphere of operational tests, the best conclusion about these issues from the studies that are available is that the evidence is weak or inconsistent.

Second, even if comparison data from operational tests were more available, the evidence is limited by the design of the items that appear on operational tests. That is, the goal of operational tests is to assess mathematical competency broadly, not to compare the MC and CR item formats. Thus, MC and CR items differ on a number of features, only one of which is format. Thus, more carefully controlled comparisons are desirable to isolate the impact of response format.

Third, however, the available comparisons between MC and CR items under controlled designs (i.e., stem-equivalent items) have yielded inconsistent results. Moreover, there is another design issue that has emerged—namely, the strategy for constructing stem-equivalent items. In some studies, CR items are created by removing the distractors from operational MC items. Evidence from these studies suggests that CR items are more difficult. In other studies, MC items are created to correspond to CR items by using common errors as distractors. Evidence from these studies suggests that MC items are more difficult.

Fourth, the controlled designs for comparing MC items to CR items have had limited scope. The items that were compared involved short numerical responses. Items at a higher level of complexity, those that involve understanding principles or showing steps, have not been compared between formats. Perhaps MC items that involve higher levels of complexity cannot be constructed; but then again, maybe they can but appropriate studies have not been undertaken.

Fifth, test context may interact with comparisons of MC and CR item formats. For example, one study found that a high-stakes context versus a low-stakes context of the same operational test was associated with decreased differences between the item formats. One interpretation is that the high-stakes test context may evoke sufficient levels of motivation in the examinees to complete the more time-consuming constructed-response formats. Moreover, tests with mixed item formats may lead to reduced format differences, due to item-solving strategies carrying over from one format to another. That is, in one study reviewed in this section, the plug-in strategy that can be effectively applied to the MC format may be extended to the CR format if the MC format precedes it.

Given the limitations described in this section, there is little or weak evidence to support the CR format as providing much different information than the MC format. For example, the available evidence provides little or no support for the possible claim that different constructs are measured by the two formats or that item discrimination varies across formats. Although some evidence suggests that CR items are more difficult, especially for the more extended CR formats, there is some contrary evidence that indicates that more difficult MC items can be constructed for their stem-equivalent CR items. Finally, the impact data does not support much difference between the two item formats. That is, the impact of response format on gender differences is inconsistent, while the impact on racial-ethnic differences is weak.

Item response format is one of the several design features that may impact item complexity. The evidence found in the scientific literature did not support the notion, however, that CR format, particularly the short answer type, measures different aspects of mathematics competency compared to MC. The impact of item format may interact with other design features, such as test context or strategy for developing controlled comparisons items. Thus, the important issue is not whether to select MC versus CR format, but rather how to most efficiently design items to measure content of the designated type and level of cognitive complexity.

APPENDIX F: Factors to Evaluate the Quality of Item Design Principles

Ensuring that high-stakes tests, such as the NAEP and various state tests, are of the highest quality psychometrically is critical. Measurement instruments need to be accurate and unbiased. The Task Group presents suggestions, or factors, for how quality control might be carried out. These factors, posed in the form of questions to be addressed, are relevant to principles such as item format and problem context (word problems), as well as item administration methods, such as including calculators and manipulatives.

1) Are the items generated from the principles appropriate for the targeted construct, or are they more likely to have nonmathematical sources of difficulty? What additional factors can reduce this vulnerability?

    a. For example, do items created in constructed-response (CR) formats rely more heavily on verbal skills than mathematical skills? Which CR formats or rubrics are less likely to involve these skills?

    b. Are items generated with "real-world" context more likely to contain confusing or irrelevant verbal material, visual displays, or practical knowledge? What mechanisms reduce this confounding?

2) Are the items generated from the principles generally appropriate to provide maximal information for the competency levels targeted by the assessment?

    a. Several factors reduce information, including unreliability in the scoring mechanisms, inappropriate item difficulty for the targeted levels, low item discrimination, and high vulnerability to guessing.

    b. What mechanisms reduce this source of confounding (e.g., machine scoring of CR)?

3) Are the items generated from the principles more likely to be vulnerable to differential item functioning (DIF)?

4) Are the items generated from the principles appropriate for model-based approaches to measurement?

    a. Current state and national assessment typically apply item response theory (IRT) approaches to scaling items. These approaches allow equating of tests across forms and time, which is necessary to examine trend and maintain comparable standards.

Are there special mechanisms to adapt diverse item design principles to IRT models? IRT easily accommodates binary and polytomous formats, including partial credit scoring. Formats known to produce local dependence (a violation of IRT assumptions) can sometimes be accommodated by special mechanisms, such as testlet scoring.

APPENDIX G: Descriptors Used in the Literature Search and Exemplars of Satisfactory Word Problems

To help the Assessment Task Group locate any previous work that might have been done with respect to language or wording defects in test items used in mathematics assessments, Abt Associates Inc. (Abt) conducted an extensive search of the research literature and related items. The descriptors used by Abt appear in the following chart.

Table G-1: Descriptors Used in Literature Search

Math* and	
Assessment*	
Test*	
Testing*	
	Item language
	Item wording
	Question language
	Question wording
	Item wordiness
	Language bias
	Linguistic simplification
	Language density
	Essential language
	Non-essential language

***Note:** All of the terms in the list were searched simultaneously with math, and assessment or test or testing.

Examples of Satisfactory Word Problems

The Task Group examined state tests to provide examples of desirable content in mathematics word problems. The major purpose of word problems on broadly given assessments should be to assess skill in converting relationships described verbally into mathematical symbolism or calculations. Moreover, they should satisfy the following conditions:

a) be written in a way that reflects natural and well-written English prose at the grade-level assessed;

b) assess mathematics knowledge and skills for the grade level of the assessment as judged by international benchmarks while restricting nonmathematical knowledge to what would be general knowledge for most students;

c) clearly assess skill in converting relationships described verbally into mathematical symbolism or calculations, or, if they use a "real-world" setting, they use a setting that aids in solving a problem that would be more cumbersome to state in strictly mathematical language.

The following five items illustrate some features that are relevant for quality word problems.

1) The following is an appropriate example of a satisfactory item for which the nonmathematical knowledge is minimal and the student is expected (as appropriate in a mathematics test) to convert relationships described verbally into mathematical symbolism or calculations:

   For example: Grade 4, Georgia, No. 1, Page 2–5

   The local park is having a game day. There are 5 teams, with 3 boys and 4 girls on each team. How many children are there in all?

   ☐ *12* ☐ *15* ☐ *20* ☐ *35*

   Comment: This problem assesses whether the student can convert a relationship described verbally into appropriate mathematical symbolism. Moreover, the nonmathematical nouns are at an appropriately low level of vocabulary.

2) Here is an appropriate word problem in which the real-world setting is an aid to the student in solving a problem that could have been expressed in strictly mathematical language:

   For example: Grade 8, Massachusetts, N.12 on Page 2–20

   Mona counted a total of 56 ducks on the pond in Town Park. The ratio of female ducks to male ducks that Mona counted was 5:3. What was the total number of female ducks Mona counted on the pond?

   A. 15 *B. 19* *C. 21* *D. 35*

   Comment**:** A student has to decide which fractions are relevant. Moreover, any statement of a ratio problem similar to this problem becomes harder to read if the context is removed. [However, the problem could be improved by making the total number of ducks equal to 120, so that the total could be divided by any of 3, 5, and 8 without giving a remainder.]

3) Here is an appropriate example of a test item requiring logical reasoning with appropriate mathematical aspects:

   For example: Grade 8, NAEP 2005, Problem 5, page 5–28

   Ravi has more tapes than magazines. He has fewer tapes than books. Which of the following lists these items from the greatest in number to the least in number?

   A) Books, magazines, tapes *B) Books, tapes, magazines*
   C) Magazines, books, tapes *D) Tapes, magazines, books*

   Comment: Although this problem could also be appropriate for a language arts assessment, it is also appropriate for a mathematics assessment because the language of inequalities is so closely related to the terminology in the item.

4) Here is an appropriate item with a natural sequence of sentences in which disciplined mathematical reasoning is the cornerstone:

   For example: Grade 8, Washington, No. 20

   Mrs. Bartiletta's class has 7 girls and 3 boys. She will randomly choose two students to do a problem in front of the class. What is the probability that she will choose 2 boys?

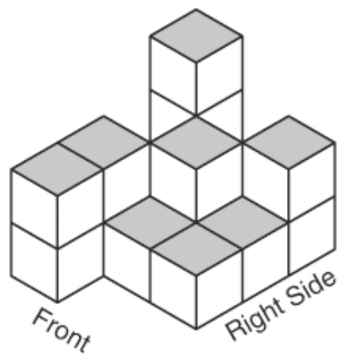
   A. $\frac{1}{15}$ *B.* $\frac{2}{5}$ *C.* $\frac{3}{7}$ *D.* $\frac{5}{19}$

   Comment: The student must first realize that this is a "without replacement" problem. Then the student is free to choose either permutations or combinations for the denominator. After that, however, the student must be consistent for the numerator. Finally, a fraction reduction is needed to match Option A. [This problem could reasonably be viewed as beyond the level required for performance at Grade 8.]

5) Here is an appropriate problem focused on one of the three methods of making planar pictures to represent 3-dimensional objects:

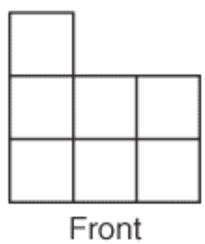
For example: Grade 8, Texas, No. 50

Melody made a solid figure by stacking cubes. The solid figure is shown below.

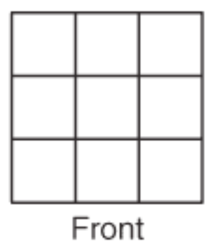


Which drawing best represents a front view of this solid figure?

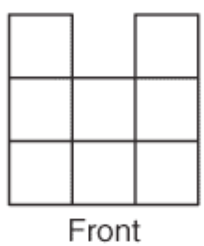
F



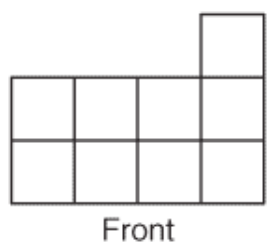
H



G



J



Comment: The problem is stated nicely. In particular, the phrase “by stacking” and the attribution to Melody make it clear to the student that he/she is being faced with a static situation.

Chapter 9: Report of the Subcommittee on the National Survey of Algebra I Teachers

Tom Loveless, Chair
Francis “Skip” Fennell
Vern Williams
Deborah Loewenberg Ball
Marian Banfield, U.S. Department of Education Staff

Final Report on the National
Survey of Algebra Teachers
For the National Mathematics Advisory Panel Subcommittee

Conducted by:
National Opinion Research Center (NORC)
at the University of Chicago
Thomas B. Hoffer
Lekha Venkataraman
E.C. Hedberg
Shobha Shagle

CONTENTS

Tables

Figures

Acknowledgments

This report draws on the data collected from 743 Algebra I teachers in public schools across the United States in spring 2007 by the National Survey of Algebra Teachers (NSAT). We first want to express our appreciation to these teachers for taking time from their busy schedules to complete and return the questionnaires. We also want to thank the principals of the schools for supporting the project and providing lists of names of the Algebra I teachers, and to the officials at the district offices who granted access to the principals and teachers.

The content of the questionnaires primarily reflects the expertise and diligence of the National Mathematics Advisory Panel members who served on the subcommittee directing the NSAT: Tom Loveless, Chair; Francis “Skip” Fennel; Vern Williams; and Deborah Loewenberg Ball. The NSAT project was staffed by Marian Banfield of the U.S. Department of Education, and the authors are particularly indebted to her for guidance and encouragement throughout the project.

The questionnaire development process benefited greatly from reviews by Jenny Nagaoka of the Consortium on Chicago School Research at the University of Chicago and by the mathematics faculty at the University of Chicago Laboratory Schools.

At the National Opinion Research Center (NORC), Tom Hoffer directed the project and Lekha Venkataraman managed the day-to-day activities of instrument development and pretesting, sample recruitment, mailouts, and data collection. Shobha Shagle developed reports for monitoring sample recruitment and data collection, and combined, documented, and cleaned the data files. Michael Yang provided expert guidance on the sample design and selection, as well as sample weighting. E. C. Hedberg built the sample frame files and conducted the data analysis for this report. Hector Flores managed the mail-out activities at NORC’s Production Center. We were ably assisted throughout the sample recruitment and data collection stages by research assistant Kirstin Millius of the University of Chicago. Imelda Demus and Isabel Guzman-Barron provided expert assistance with the report formatting. The project is indebted to Harrison Greene, Senior Vice President for Education and Child Development Studies, of NORC, for additional support of staffing and for teacher incentive stipends for survey participation.

Executive Summary

This report presents findings from a study by the National Opinion Research Center (NORC) of a nationally representative sample of public school Algebra I teachers, the National Survey of Algebra Teachers (NSAT). A sample of 310 schools was selected from a comprehensive list of public schools that included the eighth grade or higher. Of the 310 schools selected, 258 agreed to provide rosters of their Algebra I teachers. A total of 1,026 teachers were identified on this basis, and 743 (72%) returned completed questionnaires by the July 1, 2007, close of data collection. The report begins with a demographic and professional profile of the public school Algebra I teachers, and then presents findings related to the research questions identified by the National Mathematics Advisory Panel (National Math Panel, NMP, Panel) to guide the study.

Teacher Background

The Algebra I teachers who completed the survey were predominately female (66%), white (91%), and had a median age of 41 years old. The median years of teaching experience was nine years, and these teachers had taught algebra for a median of six years.

In terms of education, all had at least a baccalaureate degree and 51% had an M.A. or M.S., or other advanced degree. About 44% majored in mathematics and another 24% minored in mathematics during college; 8% earned an advanced degree in mathematics.

About 28% of the Algebra I teachers were teaching at the middle or junior high school level, while almost all of the other 72% were teaching in high schools (less than 5% were in combined middle-high schools).

Student Preparation

Research Question #1: How do the teachers rate the preparation of students coming into their Algebra I classes? Are there widespread problems, or are problems confined to individual students?

The teachers generally rated their students' background preparation for Algebra I as weak. The three skill areas in which teachers reported their students have the poorest preparation are rational numbers, word problems, and study habits (Table 7).

The teachers' ratings of student preparation generally did not vary much by school demographic. The main point of difference was that teachers of classes that primarily enroll seventh- or eighth-graders rated their students' backgrounds more highly, by 0.87 standard deviations ($p < .001$). The grade level of the class is likely to be a proxy for the ability level of the class, with eighth grade being the advanced group, ninth grade the average group, and 10th and higher the lower groups.

Research Question #2: To the degree that the teachers believe students need to be better prepared, what are the major shortcomings?

The teachers were asked to rate the importance of a “solid foundation” in the each the 15 skill and knowledge areas asked about with respect to their target class students’ background preparation. Since the same background skills and knowledge for which the teachers rated student background as inadequate were also rated as important, the following areas emerge as the major shortcomings: rational numbers, word problems, and study habits.

Research Question #3: Given their experience with incoming students, would teachers change the level of emphasis placed on mathematics topics at the elementary level? If so, how would they change it?

- Would they put more or less emphasis on basic understandings or arithmetic and whole number, fraction and decimals operations?
- Would they put more or less emphasis on helping students master basic concepts?

These questions are covered to some extent in the open-ended survey question #2 in section 3 (item 3.2), “Please provide a brief description of any changes you would like to see in the curriculum leading up to Algebra I in your district.” Of the 743 teachers who returned completed questionnaires, 578 provided verbatim responses to this item.

The most frequent type of suggestion among the 578 respondents was a greater focus in primary education placed on mastery of basic mathematical concepts and skills.

Curriculum and Instruction

Research Question #4: How do teachers rate their state and local district curricular expectations in algebra for PreK–12? How do they rate the state or local school district mathematics standards and math tests that they currently use?

- The modal response (67%) from teachers is that they feel that local expectations for student proficiency in Algebra I are “about right,” while about equal numbers rated them as “too high” (8%) or “too low” (11%) (see Figure 3).
- The teachers were also generally favorable about content standards for Algebra I in their state or local district. A majority (53%) of teachers feel that the content standards are good and 16% rate them as excellent. Only about 5% rated their content standards as poor (see Figure 4).
- Teachers were less positive about state and local assessment standards, but the modal response (43%) was still that they were “good.” About 9% rated them as excellent and 15% rated them as poor (see Figure 5).

Research Question #5: How do they rate their textbook (or textbooks in general) regarding algebra instruction?

The questionnaire included several items asking for the teacher's evaluation of the textbook they use in the target class (survey items 1.8a–i). For the most part, teachers were satisfied with their texts' topics (Figure 7). The teachers rated their textbook least positively on the degree to which it is well suited for the needs of a diverse population of students (Figure 6).

Research Question #6: How do the teachers rate online technology tools?

The questionnaire included questions asking how often the teachers used computer-based instructional tools (item 1.5f), the extent to which insufficient access to computers is a problem in their school (item 2.1a), and how much they agreed or disagreed with the proposition that "Computer-based instructional tools (software) are helping Algebra I students in my Target Class" (item 1.6).

The data indicated that the average response to how frequently these tools are used was about 1 (= less than once a week) on a scale that ranged from (0 = never) to (4 = everyday) (Table 9 and Appendix D). The generally low levels of computer use does not appear to be a reflection of insufficient access. About half (49%) of the teachers reported that insufficient access to computers was not a problem in their schools and another 28% reported insufficient access to be a minor problem (Table 9). The teachers' ratings of the helpfulness of computer-based instructional tools were mixed, with 29% agreeing somewhat or agreeing strongly with the proposition that computers were helpful and 38% disagreeing somewhat or disagreeing strongly (34% neither agreed nor disagreed) (Figure 8).

Research Question #7: What is the role of the calculator in the algebra course?

Questionnaire item 1.5d asked how often the teacher uses graphing calculators in her or his target class. Overall, 33% of the teachers reported never using graphing calculators and another 29% report using them less than once a week. About 31% used them everyday (18%) or almost everyday (13%) (Table 10). Teachers' reports of insufficient access to graphing calculators was correlated with reports of low usage (Table 11).

Research Question #8: To what extent do the Algebra I teachers use physical objects (manipulatives) as instructional tools?

The relevant questionnaire item for this question asked how often the teacher uses physical objects, commonly referred to as manipulatives, in her or his target class (item I.5e). Overall, use of manipulatives on an occasional basis was widespread, but very few (9%) teachers report using them more than once a week or everyday. About 12% of the teachers reported never using manipulatives, and about 60% reported using them less than once a week (Table 12).

Research Question #9: How do teachers rate their professional training?

Questionnaire items pertaining to professional training and development included questionnaire items 3.4a,b and possibly 4.19; items 2.1f and j are also relevant. These items were examined by the teachers' years of teaching experience, and school classification variables.

- Teachers were asked to report the extent to which they believe inadequate preparation to teach Algebra 1 and opportunities for professional development are problems in their respective schools. The responses showed that both of these were generally viewed as very minor problems in their schools (see Table 13). When asked to rate their own Algebra I-related preservice training and in-service professional development opportunities, the teachers on average rated these experiences as just "adequate" (see Table 13 and Figures 10 and 11), suggesting room for improvement in the preservice training programs and professional development opportunities experienced by some teachers.

Research Question #10: Is there sufficient and effective remedial help for students who are struggling in algebra? What sort of assistance-based interventions would struggling students benefit from the most?

Questionnaire items 2.8a–b asked the teachers to rate the availability and quality of tutoring or other remedial services for students struggling with Algebra I in their schools.

- On average, teachers were generally satisfied with the services available (Table 14).

- Controlling for other demographic variables, remedial services were rated somewhat higher by teachers in schools with high minority enrollments. Also controlling for other demographic variables, female and black teachers are less satisfied with their schools' remedial services. (See Appendix Table C-8.)

Research Question #11: Would students learn more if they were grouped by ability for instruction, or is this approach counterproductive?

Questionnaire item 2.2 asked whether the school offers different levels of Algebra I based on ability; and 46% of the teachers indicated their schools did differentiate. Questionnaire item 2.1h asked teachers to rate the extent to which they see different levels of students in the same class as a problem in their school. A substantial number of teachers considered mixed-ability groupings to be a "moderate" (28%) or "serious" (23%) problem (see Figure 12). Teachers in schools that did not offer different levels of Algebra I based on ability were more likely than their counterparts in schools that do use ability grouping to consider mixed-ability classrooms to be a moderate or serious problem (Table 15).

Research Question #12: Do teachers find more parents helpful in encouraging students in their mathematics studies, or do too many parents make excuses for their children's lack of accomplishment?

Questionnaire item 2.1i asked teachers to rate the extent to which they see "too little parent/family support" as a problem in their school. The responses indicate that about 28% of the algebra teachers felt family participation is a serious problem and another 32% believed lack of family participation is a moderate problem (Figure 13).

Research Question #13: What do they see as the single most challenging aspect of teaching Algebra I successfully?

This question (4.20) included 10 response options: explaining material to students, handling accelerated students, teaching procedures, explaining concepts, using diagrams or models effectively, interpreting student errors and difficulties, working with unmotivated students, working with advanced students, helping students whose home language is not English, making mathematics accessible and comprehensible, and an “other” option.

The overwhelmingly most frequent response to this question was “working with unmotivated students.” This was chosen by 58% of the middle school teachers and 65% of the high school teachers (Table 16). The next most frequent response was “making mathematics accessible and comprehensible to all my students,” selected by 14% of the middle school teachers and 9% of the high school teachers.

Conclusions

The Algebra I teachers generally reported that students were not adequately prepared for their courses. The teachers rated as especially problematic students’ preparation in rational numbers, solving word problems, and basic study skills. A lack of student motivation was by far the most commonly cited biggest challenge reported by the teachers. The problems the teachers identified with the pre-Algebra I mathematics curriculum and instruction and with the lack of parental support for mathematics were likely to be contributing factors to the lack of adequate student preparation and motivation.

In contrast, the teachers generally held favorable views with respect to their own professional preparation and the Algebra I curriculum and instructional services. Taken together with the generally negative ratings of students’ preparation and motivation suggests that careful attention to pre-algebra curriculum and instruction in the elementary grades is needed, both to remedy the specific skill deficiencies reported by the Algebra I teachers and to identify ways in which negative attitudes toward mathematics develop and might be changed.

I. Introduction

The National Survey of Algebra Teachers (NSAT) conducted by the National Opinion Research Center (NORC) surveyed a national sample of public school Algebra I teachers during the 2007 spring school semester. The survey was designed to collect detailed information about the teachers' views on student preparation, motivation, work habits, and skills—as well as teachers' insights on how math is now taught, how earlier math education could be improved to prepare more children to succeed at algebra, and what would help all math teachers do a better job. The survey was designed to shed light on the experiences of algebra teachers in different kinds of school systems—for example, low-income, mainly minority schools versus higher income, mainly white schools. Learning algebra is often a turning point in a student's math education—when the student either thrives and moves forward or struggles and perhaps gives up on math—and the algebra teachers have a unique perspective on math education that is well worth understanding in some detail.

The NSAT was designed to provide a nationally representative sample of Algebra I teachers in public schools. A sample of 310 schools was selected from a comprehensive list of public schools which included the eighth grade or higher. The list was stratified by the type of grade configuration in the school (middle or junior high school, high school only, combined middle and high school), the number of students from low-income households, the number of racial and ethnic minority students enrolled in the school, and school location (urban, suburban, rural). Within the strata defined by these variables, schools were selected with probabilities of selection proportional to the estimated numbers of Algebra I teachers. Of the 310 schools selected, 258 agreed to provide rosters of their Algebra I teachers. A total of 1,026 teachers were identified on this basis, and 743 (72.4%) returned completed questionnaires by the July 1 close of data collection.

This report presents the survey results and provides initial analyses to identify important sources of variability in the teacher reports. It begins with a demographic and professional profile of the public school Algebra I teachers, and then presents findings related to the research questions identified by the National Mathematics Advisory Panel to guide the study. The survey methodology and data collection results are described in Appendix A. A full set of tabulations of the main survey variables is included in Appendix B. Tables and figures are used throughout the report to improve readability, and the numbers upon which they are based are displayed in the Appendix B tables. Multiple regression models are estimated to provide compact summaries of the influences of several variables on the outcomes focused on in the report, and the regression tables are included in Appendix C along with a descriptions of the independent variables used in the models. Appendix D is the means and confidence intervals for items in the National Survey of Algebra Teachers. Appendix E is a copy of the questionnaire used to collect the data. The report concludes with a summary of the main findings and a discussion of their implications.

II. Analysis of Survey Variables

Teacher Background and Work Situation

A profile of the demographic and professional backgrounds of the academic year 2006–07 Algebra I teachers in U.S. public schools is shown in Table 1. These teachers were predominately female (66%), white (91%), and had a median age of 41 years old. The Algebra I teachers' median years of teaching experience was nine years and had taught algebra for a median of six years. In terms of education, all had at least a baccalaureate degree and about half had an M.A. or M.S., or other advanced degree. About 44% majored in mathematics and another 24% minored in mathematics during college; about 15% of those who earned an advanced degree specialized in mathematics (Table 1).

Table 1: Demographic and Professional Characteristics of Algebra I Teachers: 2007

Characteristic	Values	Valid N	Weighted %
Teacher is female	0–1	733	65.5
Teacher racial/ethnic background			
Hispanic	0–1	727	5.7
American Indian or Alaska Native	0–1	715	2.1
Native Hawaiian or other Pacific Islander	0–1	715	0.2
Asian	0–1	715	2.5
Black or African-American	0–1	715	3.6
White	0–1	715	91.0
Teacher age (quartiles)	1st: 22–30 yrs		27.4
	2nd: 31–40 yrs		21.6
	3rd: 41–50 yrs		25.1
	4th: 51–65 yrs		26.0
	All	729	100.0
Teacher's total years teaching experience (quartiles)	1st: 0–3 yrs		31.1
	2nd: 4–9 yrs		30.6
	3rd: 10–18 yrs		21.6
	4th: 19–41 yrs		16.7
	All	733	100.0
Teacher's years teaching algebra (quartiles)	1st: 0–2 yrs		24.4
	2nd: 3–6 yrs		24.4
	3rd: 7–14 yrs		26.4
	4th: 15–40 yrs		24.8
	All	733	100.0
Teacher's highest degree	Bachelor's		51.4
	Master's		40.9
	Other advanced degree		7.7
	All	737	100.0
Baccalaureate math background	Math major	738	43.6
	Math minor	729	24.2
Graduate degree math background	Math specialty	400	15.2
Teacher has regular or standard state certification	0–1	733	82.4

Source: Based on responses to the National Opinion Research Center's National Survey of Algebra Teachers, 2007.

The distribution of Algebra I teachers by grade level (8–12) and by the main school-level classification variables used throughout the report is shown in Table 2. The first three of these school-level variables largely reflect student enrollment patterns across the country:

- *Type of locale*: the standard three-level indicator of urban (27%), suburban (39%), and rural (34%) school location.
- *Percentage of students receiving free or reduced-price lunch*: the percentage variable was recoded into quartiles of the distribution of Algebra I teachers (median was 10% of the students are eligible).
- *Percentage of students who are black or Hispanic*: the percentage variable was recoded into quartiles of the distribution of Algebra I teachers (median is 27% of the students are black or Hispanic).

The grade level variable at the bottom of Table 2 indicates that 32% of the algebra teachers were teaching at the middle or junior high school level, while 50% were teaching in high schools and 18% were in combined middle-high schools.

Table 2: Characteristics of Schools With Algebra I Teachers: 2007

School Characteristics	Values	UnWtd. N	Wtd. N	Wtd. %
School urbanicity	Urban	252	23,088	26.9
	Suburban	381	33,796	39.4
	Rural	110	28,891	33.7
	Total	743	85,775	100
Percentage minority—quartiles	Low thru 10%	119	22,923	26.7
	11 thru 27%	184	20,100	23.4
	28 thru 48%	265	24,549	28.6
	49 thru 81%	175	18,202	21.2
	Total	743	85,775	100
Percentage free/reduced-price lunch status—quartiles	Low thru 3%	219	21,998	25.6
	4 thru 10%	227	24,537	28.6
	11 thru 40%	182	22,318	26
	41 thru 82%	103	16,358	19.1
	Total	731	85,210	99.3
School grade level	Middle, junior high, or K–8 school	128	27,508	32.1
	High school (9–12 or 10–12)	532	43,234	50.4
	Other schools	83	15,033	17.5
	All schools	743	85,775	100

Source: Based on responses to the National Opinion Research Center's National Survey of Algebra Teachers, 2007.

The Algebra I teachers were asked to report several characteristics about a "target" Algebra I class they were currently teaching. The following table shows the portion of algebra teachers and their classes that fit various criteria. Most teachers reported that their class meets everyday (83%) and that they have enough time to teach algebra adequately (77%). About half of the teachers' schools offer different levels of algebra based on student needs, and about one-third of teachers reported that their class is part of block scheduling in their school.

The teachers were asked which student grade levels they were currently teaching in their Algebra I classes. The ninth grade was reported most often, by 58% of all the algebra teachers. Tenth grade was next (43%), followed by eighth grade (38%) and 11th grade (28%). A significant portion taught seniors (17%), and only 7% reported teaching seventh-graders. A significant number of the teachers (15%) reported teaching special education students in their Algebra I class(es). (See Table 3.)

Table 3: Percentages of Algebra I Teachers Reporting Various Characteristics of Their Classes and Schools: 2007

Classes and School	**Lower 95% CI**	**Mean**	**Higher 95% CI**
Target class meets everyday	76.1%	**82.8%**	89.4%
Feel they have enough time to adequately teach	70.7%	**76.3%**	81.9%
School offers different levels of Algebra I based on ability	39.3%	**46.6%**	54.0%
Target class is part of block scheduling	26.4%	**33.9%**	41.4%
Teachers Who Teach Algebra I to			
7th-graders	3.7%	**6.7%**	9.7%
8th-graders	31.2%	**38.4%**	45.7%
9th-graders	50.6%	**57.5%**	64.5%
10th-graders	36.9%	**43.2%**	49.5%
11th-graders	22.3%	**27.6%**	32.8%
12th-graders	12.3%	**16.8%**	21.3%
Special education students	10.8%	**15.1%**	19.4%
Teachers' Estimates of How Many Students Will Fail Their Algebra I Course			
None of the students in target class	15.6%	**21.7%**	27.9%
1–10% of the students in target class	33.9%	**40.7%**	47.4%
11–20% of the students in target class	12.4%	**18.0%**	23.6%
21–30% of the students in target class	5.3%	**8.3%**	11.4%
31–40% of the students in target class	3.5%	**5.6%**	7.6%
41–50% of the students in target class	2.2%	**3.3%**	4.4%
50% or more of the students in target class	1.4%	**2.5%**	3.7%

Note: CI = confidence interval, calculated as +/- two standard errors from the mean. Standard errors adjusted for design effects.

Source: Based on responses to the National Opinion Research Center's National Survey of Algebra Teachers, 2007.

With regards to rates of failing Algebra I, 22% of the teachers believed that none of the students in their target class would fail, and another 41% expected 1–10% of their students would fail. A substantial proportion of the teachers (20%) expected to fail more than 20% of their students.

Time allocations. Teachers were asked to report the number of minutes spent on various activities. On average, a class period of algebra lasted about one hour. Teachers also averaged about 1 hour per day preparing for their classes during the school day. Teachers also spent time outside of school in preparation, which averaged 54 minutes per day. In comparison, teachers expected their students to spend about 25 minutes per day on their Algebra I homework.

Table 4: Average Time (in Minutes) Algebra I Teachers Spent on Various Activities: 2007

Activity	**Lower 95% CI**	**Mean**	**Higher 95% CI**
In class per period	59.28	**62.14**	65.00
In preparation during a school day	57.25	**61.16**	65.07
In preparation for algebra outside of school	50.14	**54.38**	58.62
Expected time needed for target class students to complete homework per day	23.28	**24.81**	26.33

Note: CI = confidence interval, calculated as +/- two standard errors from the mean. Standard errors adjusted for design effects.

Source: Based on responses to the National Opinion Research Center's National Survey of Algebra Teachers, 2007.

As for the students in their target class, teachers were generally satisfied with their in-class behavior. On average, teachers felt that most of their students came to class on time and attended class regularly. Teachers also felt that more than half of their students generally came to class prepared, paid attention, participate, take notes, and care about the grades they receive. Disruptions do not appear be a major problem, as teachers report that few of their students create behavior problems. Finally, teachers felt that few of their students have serious difficulties reading English.

Further analyses found that teachers in urban schools were more likely to report that their students presented behavior problems, while teachers in rural schools reported the best-behaved students.

Table 5: Teacher-Reported Algebra I Target Class Student Behavior Characteristics: 2007

Student Behavior Characteristics	**Lower 95% CI**	**Mean**	**Higher 95% CI**
Come to class on time	3.49	**3.57**	3.65
Attend class regularly	3.39	**3.46**	3.54
Come to class prepared with appropriate supplies and books	2.79	**2.92**	3.05
Create serious behavior problems	0.53	**0.61**	0.69
Regularly pay attention in class	2.70	**2.82**	2.93
Actively participate in class activities	2.57	**2.69**	2.80
Take notes	2.59	**2.72**	2.86
Have serious difficulties reading English	0.41	**0.47**	0.54
Care about what grade they receive	2.78	**2.90**	3.02

Note: Scale: 0 = None, 1 = Some, 2 = About Half, 3 = Most, 4 = Nearly All

CI = confidence interval, calculated as +/- two standard errors from the mean. Standard errors adjusted for design effects.

Source: Based on responses to the National Opinion Research Center's National Survey of Algebra Teachers, 2007.

Size of target class. Most teachers have classes between 15 and 30 students, with 21–25 students reported most often. However, NORC's analysis found a strong correlation (r = 0.54) between the size of a teacher's target class and whether or not he or she felt that class size is a problem (see Table 6). Of those that felt it was not a problem, 90% of those teachers had class sizes of 25 students or below. Of those that felt it was a serious problem, almost 75% of those teachers had a class size above 25 students. There is a clear connection between class size and teachers feeling that it is a problem; this correlation is across the board.

Table 6: Size of Target Class, by Extent to Which the Teacher Considers Large Class Sizes to Be a Problem in the School: 2007

	How much of a problem is class size?				
Size of Target Class	**Not a Problem**	**Minor Problem**	**Moderate Problem**	**Serious Problem**	**All Teachers**
Less than 15 students	19.19%	4.05%	2.00%	0.41%	**9.90%**
15–20 students	40.44%	21.93%	11.24%	4.24%	**26.11%**
21–25 students	29.56%	41.89%	24.07%	19.84%	**30.82%**
26–30 students	7.58%	28.13%	51.19%	38.46%	**24.37%**
31–35 students	1.99%	2.78%	10.05%	30.37%	**6.90%**
More than 36 students	1.24%	1.21%	1.45%	6.67%	**1.90%**
Total	**100%**	**100%**	**100%**	**100%**	**100%**

Note: Chi-square = 296.6 ($p < 0.000$), Correlation = 0.54 ($p < 0.00$)

Source: Based on responses to the National Opinion Research Center's National Survey of Algebra Teachers, 2007.

Student Preparation

Research Question #1: How do the teachers rate the preparation of students coming into their Algebra I classes? Are there widespread problems, or are problems confined to individual students?

As noted in the previous section, the teachers were asked to report several characteristics about a target Algebra I class they were currently teaching. The questionnaire items asking about students' preparation are in Section 1, question #4 (items 4a–4o). The topics are listed in Table 7 and ranked from the biggest problem (on the bottom) to the smallest (the top). These items range from 1 = excellent [preparation] to 4 = poor [preparation].

Table 7: Teachers' Survey Responses on Student Preparation for Algebra I: 2007

Based on your experience with incoming Algebra I students in your Target Class, how would you rate students' background in each of the following areas of mathematics?	**Mean**	**95% CI**	
		Low	**High**
Whole numbers and operations with whole numbers	1.86	1.80	1.92
Working cooperatively with other students	2.32	2.26	2.37
Plotting points, and graphing lines on the four-quadrant coordinate plane	2.44	2.37	2.51
The concept of variables	2.48	2.42	2.54
Computation skills	2.53	2.47	2.60
Positive & negative integers and operations with positive & negative integers	2.58	2.51	2.64
Working independently	2.58	2.52	2.64
Solving simple linear equations and inequalities	2.80	2.74	2.86
Measurement formulas of basic geometric shapes	2.81	2.75	2.87
Manipulation of variables	2.82	2.76	2.88
Ratios, percents, rates, and proportions	2.83	2.77	2.90
Ability to use math in context that are identified as real-world situations	2.94	2.89	3.00
Basic study skills and work habits necessary for success in math	3.00	2.94	3.06
Rational numbers and operations involving fractions and decimals	3.10	3.04	3.16
Solving word problems	3.26	3.20	3.32

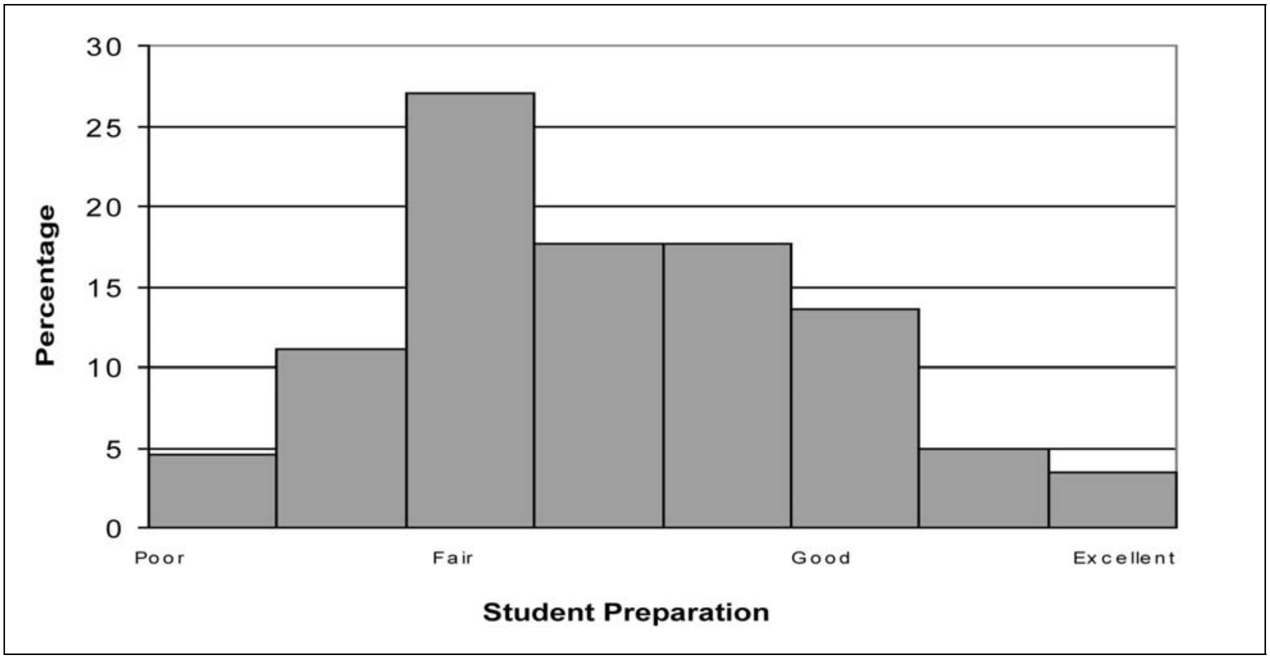
Note: CI = confidence interval, calculated as +/- two standard errors from the mean. Standard errors adjusted for design effects.

Source: Based on responses to the National Opinion Research Center's National Survey of Algebra Teachers, 2007.

As Table 7 shows, the three skill areas in which teachers report their students have the poorest preparation are solving word problems, rational numbers and operations involving fractions and decimals, and basic study skills and work habits. Student preparation is relatively strong in whole numbers and operations with whole numbers, working cooperatively with other students, and plotting points and graphing lines on the four-quadrant coordinate plane.

The teachers' responses to the various items in this battery are highly correlated with one another and can be combined into a single "student preparation" summary scale. As is evident in Figure 1, teachers generally feel their students are fair-to-poorly prepared for their algebra class (alpha = 0.94).

Figure 1: Percentage Distribution of Composite Student Preparation Scale Score: 2007



Source: Based on responses to the National Opinion Research Center's National Survey of Algebra Teachers, 2007.

Differences in the teachers' scale scores associated with types of classes and schools were assessed using regression analysis. The estimated regression coefficients of the class-type and school-level covariates are reported in Appendix Table C-1.

- The most consistent finding from the analyses is that, holding other factors constant, teachers of classes of primarily seventh- or eighth-graders rated their students' backgrounds more highly, by 0.88 standard deviations ($p < .001$). The grade level of the class is likely to be a proxy for the ability level of the class, with eighth grade being the advanced group, ninth grade the average group, and tenth and higher the lower groups.

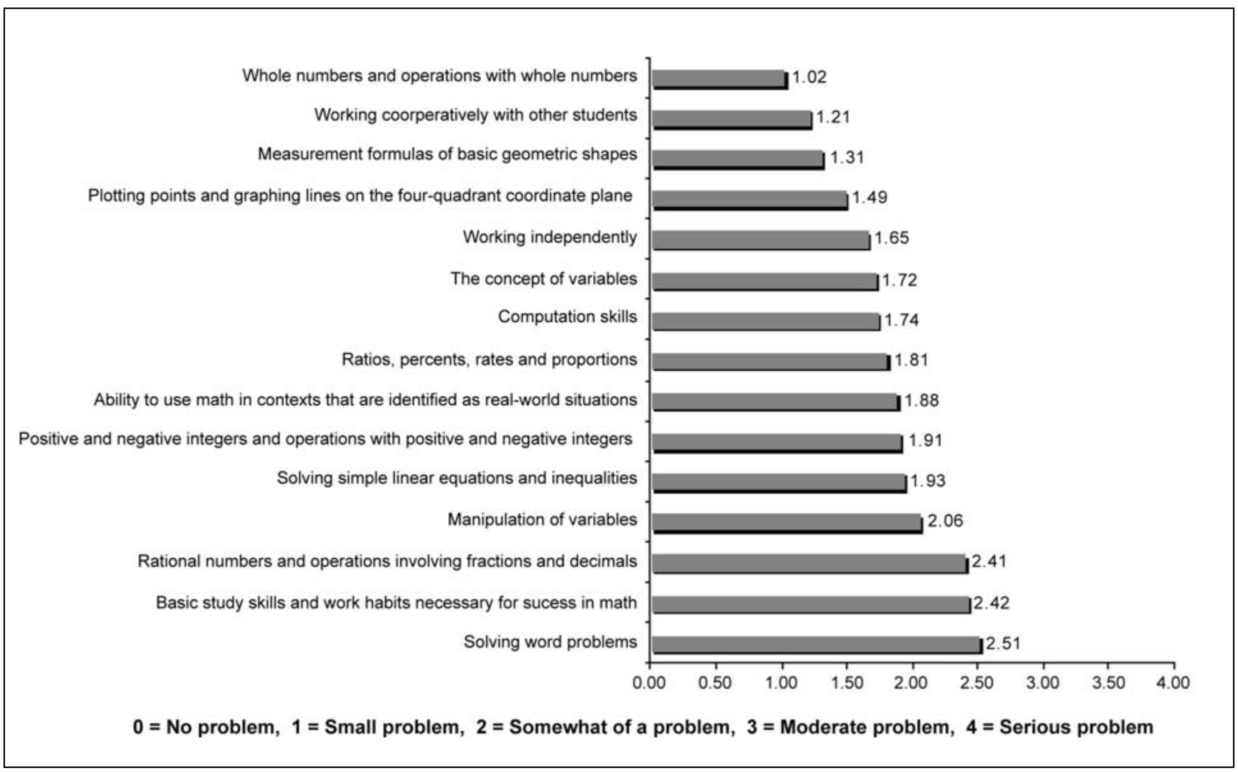
The regression analysis also finds that some school-level covariates were associated with whether teachers feel their students are prepared. Teachers in schools with a high concentration of minority students (greater than 81%) felt that their incoming students were less prepared, but this difference was reduced and not statistically significant in the full regression equation. Interestingly, there was only a weak association of teacher ratings with the schools' free and reduced-price lunch concentrations. Teachers' opinions of their students' preparations varied across urban-suburban-rural lines, with urban teachers having the lowest opinion and rural teachers having the highest, but these differences were not significant in the full regression.

Research Question #2: To the degree that the teachers believe students need to be better prepared, what are the major shortcomings?

The teachers were asked to rate the importance of a "solid foundation" in the each the 15 skill and knowledge areas asked about with respect to their target class students' background preparation (see questionnaire items 3.1a–o). We addressed this research question by combining the teachers' responses to the 15 student preparation items (1.4a–o) with teacher responses to the questionnaire items asking how important each of the preparation items is for success in Algebra I (3.1a–o). Information from the two batteries was combined to weight the preparation rating by its importance. A "preparation problem" score for each item was calculated by multiplying the teacher's rating of his or her students' preparation by that teacher's rating of the importance of a solid foundation in that particular area to students' success in Algebra I.

- Referring to Figure 2, weighting each topic by the teachers' level of importance, yields a similar pattern to that shown in Table 7 for the teachers' ratings of student backgrounds, with only minor differences in the ordering of the items.

Figure 2: Teachers' Ratings of Student Preparation Problems in Various Areas of Mathematics: 2007



Source: Based on responses to the National Opinion Research Center's National Survey of Algebra Teachers, 2007.

The set of preparation-problem items are highly intercorrelated and, like the background-preparation items, can be combined into a summary scale to facilitate analysis of factors related to differences among teachers in their ratings. For NORC's analysis, a summary "preparation problem" scale was constructed using the full set of weighted items and it was regressed based on the standard classroom and school classification variables.

- The regressions of this scale on the classroom, school, and teacher variables also confirm the patterns from the ratings of background preparation—students in the seventh- and eighth-grade Algebra I classes are better prepared than those taking Algebra I in Grade 9 and higher (see Appendix Table C-2).

The consistency of Table 7 and Figure 2 reflects the fact that virtually all of the "how important" items (3.1a–n) were rated as "very important" or "extremely important" by almost all respondents. Because these are largely invariant across the whole sample, the weighting method just outlined did not yield different results than the analysis of the preparation items discussed under research question #1.

Research Question #3: Given their experience with incoming students, would teachers change the level of emphasis placed on mathematics topics at the elementary level? If so, how would they change it?

- Would they put more or less emphasis on basic understandings of arithmetic and whole number, fraction, and decimals operations?

- Would they put more or less emphasis on helping students master basic concepts?

These questions are covered to some extent in the open-ended item 3.2, "Please provide a brief description of any changes you would like to see in the curriculum leading up to Algebra I in your district." Of the 743 teachers who returned completed questionnaires, 578 provided verbatim responses to this item.

A substantial number of the 578 would like to see a greater focus in primary education placed on mastery of basic mathematical concepts. For example:

> *"Students need to be better prepared in basic math skills and not be quite so calculator dependent. Also, more training in thinking skills."*

> *"Make sure the 1st–8th grade teachers teach the foundations of math and that the students know their basic skills."*

> *"More focus on basics - students should already know order of operations, positive vs. neg. numbers, fractions, and decimals."*

> *"Stronger basic math facts, less rigor and rushing to higher math and more arithmetic."*

> *"Please do not allow students to use calculators, especially fraction calculators."*

As these examples suggest, responses to this item will also be the best source in the questionnaire for answers to the National Math Panel's research question "What are the teachers' views on students using calculators in the early grades?" Of those that wrote an answer for item 3.2, (N = 578), 13% (N = 75) specifically mentioned that they would like to see less use of calculators before students take their Algebra I class.

Additionally, 8% of the teachers (N = 46) also mentioned changing pre-algebra standards. These responses not only include teachers stating that students need to prove their pre-algebra competence before entering Algebra I, but also indicate that pre-algebra is not even offered to all students before entering Algebra I. For example:

> *"Make pre-alg or Alg I a requirement for middle schools."*
>
> *"I would like to see a pre-algebra class as a requirement prior to taking Algebra."*
>
> *"Most students in my class have a different curriculum in middle school, so they do not officially have pre-algebra. A better diagnostic and year end assessment is essential. Many students are dependent on calculators."*
>
> *"The curriculum issue is being address next year. We are adding general math and pre-algebra and we will hopefully insist on mastery before allowing students to take Algebra I."*
>
> *"Students should have at least 80% proficiency in pre-algebra skills. Class for high schools students not proficient in these skills. Alternative classes or students with behavior and/or attendance issues."*
>
> *"Student mastery of pre-alg concepts before enrolling in Alg."*
>
> *"Mandatory success in a pre-algebra course."*

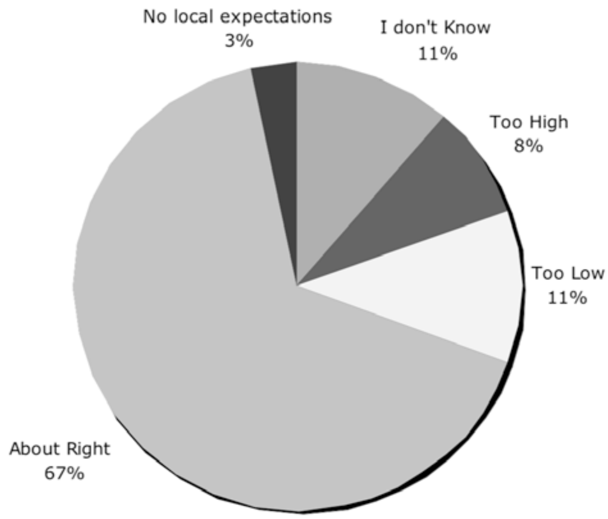
Curriculum and Instruction

Research Question #4: How do the Algebra I teachers rate their state and local district curricular expectations in algebra for PreK–12? How do they rate the state or local school district mathematics standards and math tests that they currently use? Are they setting the right expectations? Too low or unrealistically high? Clear and helpful, or confused and counterproductive? [This combines two separate research questions as requested by the National Mathematics Advisory Panel (NMP) subcommittee].

The questionnaire included one item asking the teachers to rate their local district's expectations for student proficiency in Algebra I (3.3) and two items asking about state standards and assessment tools (3.7a,b). A fourth related question asked whether students are required to pass Algebra I in order to graduate high school (3.6). These responses were examined these responses by the school classification variables.

- The modal response (67%) from teachers is that they feel that local expectations for student proficiency in Algebra I are "about right," while about equal numbers rated them as "too high" (8%) or "too low" (11%) (see Figure 3).

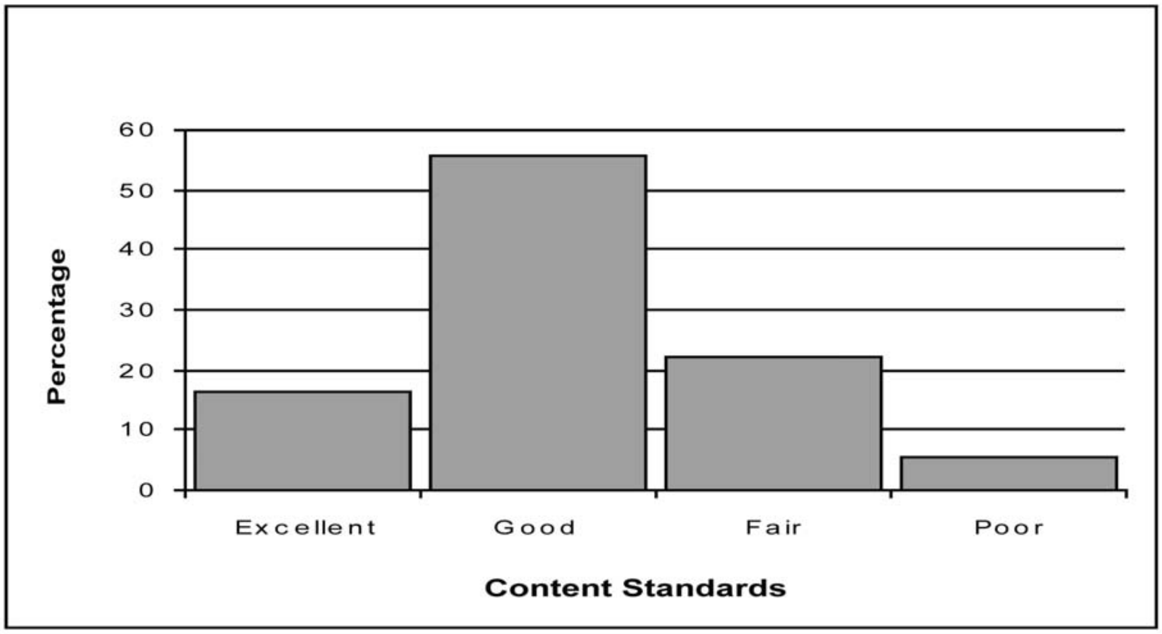
Figure 3: Teachers' Ratings of Local District Expectations for Student Proficiency in Algebra I: 2007



Source: Based on responses to the National Opinion Research Center's National Survey of Algebra Teachers, 2007.

The teachers were also generally favorable about content standards for Algebra I in their state or local district. A majority (54%) of teachers felt that the content standards are good and 19% rate them as excellent. Only about 3% rated their content standards as poor (see Figure 4). However, the regression analysis shows that teachers who teach in schools in the second quartile of minority student population also feel that the standards are better (.37 SD), compared with the feelings of teachers with low levels of minority students (see Appendix Table C-3).

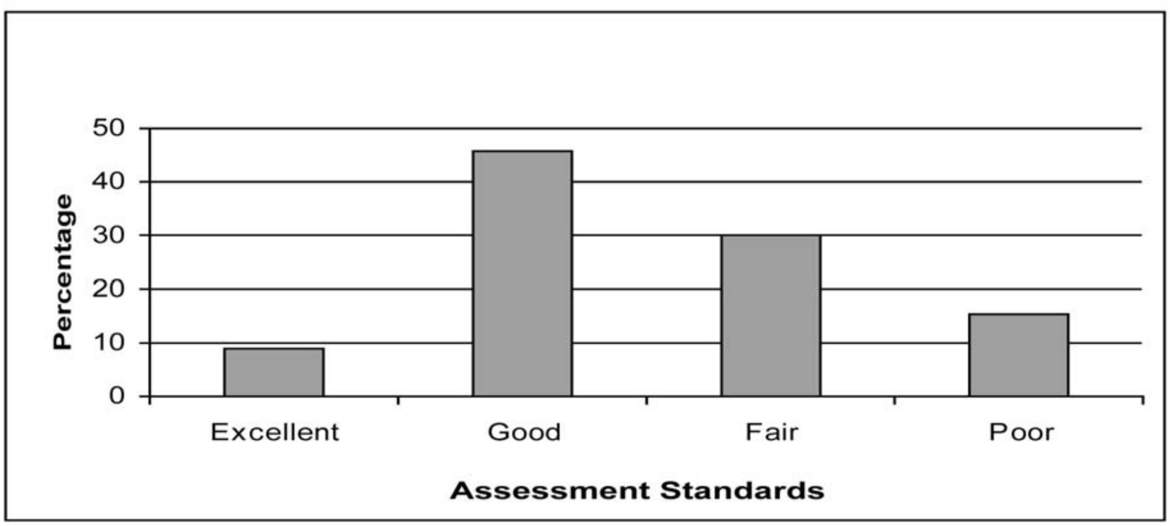
Figure 4: Teachers' Ratings of State or Local School District Mathematics Content Standards for Algebra I: 2007



Source: Based on responses to the National Opinion Research Center's National Survey of Algebra Teachers, 2007.

Teachers were less positive about state and local assessment standards, but the modal response was still that they were "good" (see Figure 5). The regression analysis did not find any differences based on teacher or school characteristics (see Appendix Table C-4).

Figure 5: Teachers' Ratings of State or Local School District Mathematics Assessment Standards of Algebra I Outcomes: 2007



Source: Based on responses to the National Opinion Research Center's National Survey of Algebra Teachers, 2007.

School Problems. The NSAT questionnaire also included a battery of questions regarding possible problems with the teacher's school, and the next table reports the means and 95% confidence intervals for these items. From poor computer access to inadequate administrative support, examination of the confidence intervals show that teachers have a problem with each aspect of their school to a similar degree. Teachers feel that each aspect is, on average, a minor problem.

Table 8: School Problems Reported by Algebra Teachers: 2007

School Problem	Lower 95% CI	Mean	Higher 95% CI
Insufficient access to computers	1.68	**1.86**	2.04
Inadequate access to graphing calculators	1.58	**1.70**	1.81
Poor quality or out-of-date textbooks	1.43	**1.59**	1.75
Too large class sizes	1.84	**1.97**	2.10
Too little coordination between classes in the mathematics	1.62	**1.75**	1.87
Lack of teacher planning time	1.63	**1.74**	1.85
Inadequate administrative support	1.52	**1.64**	1.75

Note: Scale: 1 = Not a problem, 2 = Minor, 3 = Moderate, 4 = Serious problem

CI = confidence interval, calculated as +/- two standard errors from the mean. Standard errors adjusted for design effects.

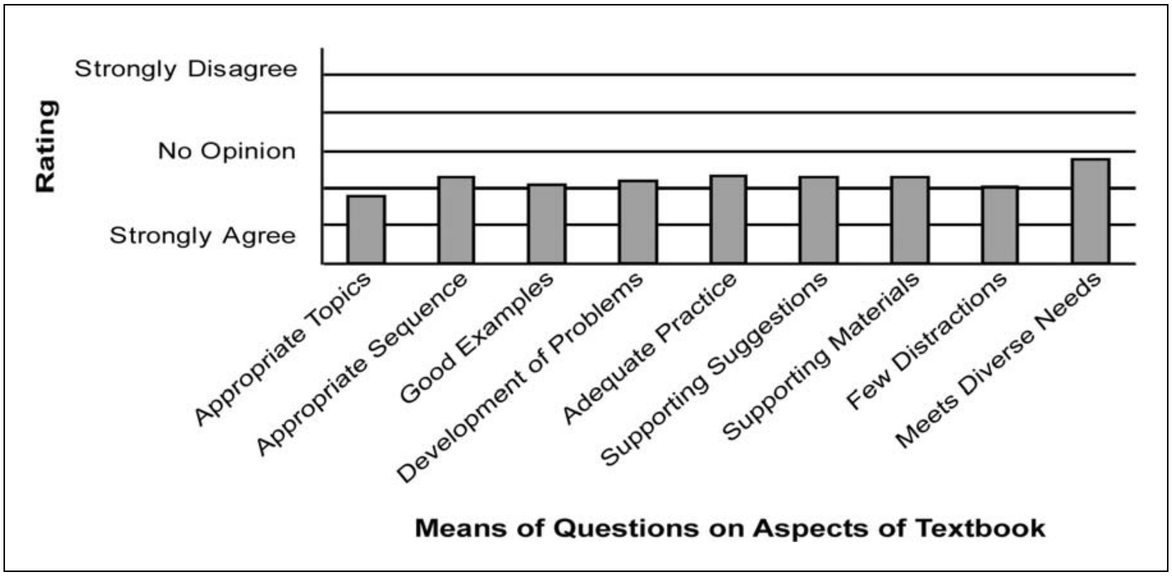
Source: Based on responses to the National Opinion Research Center's National Survey of Algebra Teachers, 2007.

Research Question #5: How do they rate their textbook (or textbooks in general) regarding algebra instruction?

The questionnaire included several items asking for the teacher's evaluation of the textbook they use in the target class (items 1.8a–i). In NORC's review, these were first examined item-by-item and then assessed whether they form a scale. The items and scale are then broken down by school classification variables and grade level of the Algebra I class.

- Figure 6 shows, item by item, how strongly the teachers agreed that their textbooks were well suited for a specific task. This figure shows there is little variation across items. For the most part, teachers were satisfied with their texts' list of topics. The only point of (possible) contention is that some teachers felt that their textbooks were not well suited for the needs of a diverse population of students.

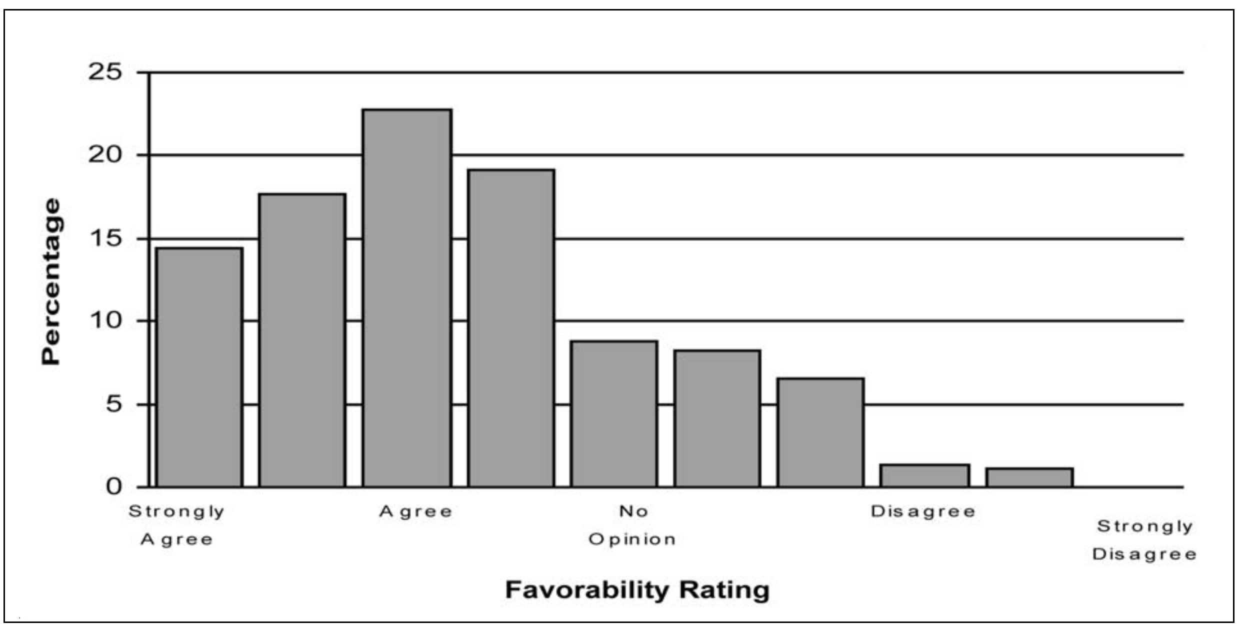
Figure 6: Teachers' Ratings of Various Aspects of the Algebra I Textbook Used in Target Class: 2007



Source: Based on responses to the National Opinion Research Center's National Survey of Algebra Teachers, 2007.

The data indicate that the nine items form a strong scale, with reliability of alpha = .90. Figure 7 shows the average composite scale score of the textbook rating questions across respondents. As is clear, the majority of the teachers have a positive view of their text.

Figure 7: Percentage Distribution of Teacher Composite Textbook Favorability Ratings Scale Score: 2007



Source: Based on responses to the National Opinion Research Center's National Survey of Algebra Teachers, 2007.

- The regression results for this composite scale show that teachers of smaller classes had more favorable ratings of their textbooks (Appendix Table C-5). Teachers with small classes (15 or fewer) like their text more by 0.56 standard deviations. Likewise, teachers in rural schools also like their books more, in this case by 0.35 standard deviations. However, teachers in schools with a high concentration of minority students have a less favorable view of their texts. On average, they like their texts less by .52 standard deviations.

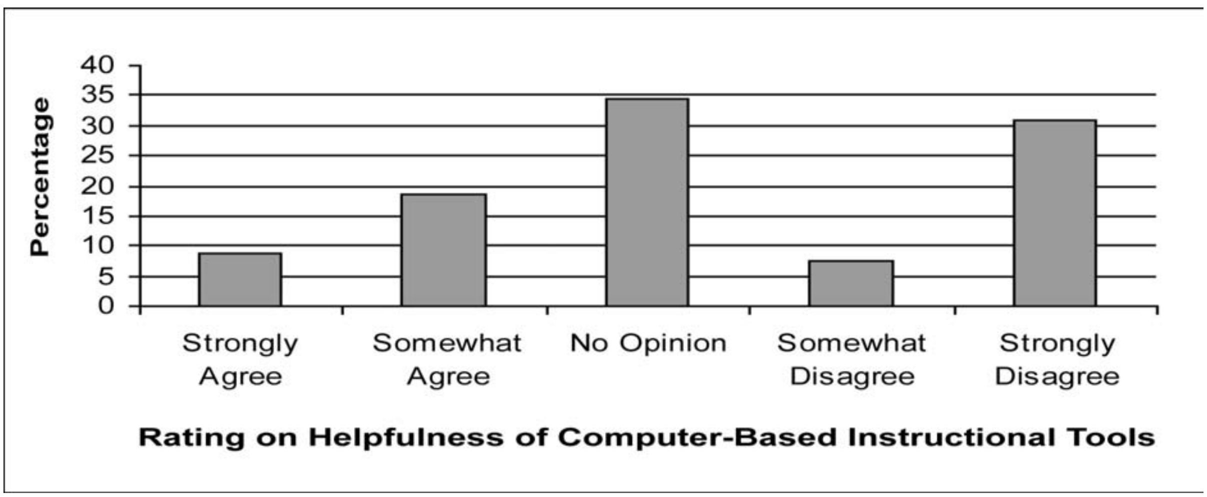
This generally positive evaluation was corroborated by the teachers' responses to an item asking them to rate the extent to which "poor quality or out-of-date textbooks" are a problem in their school. On a scale that ranged from 1 = not a problem to 4 = serious problem, the average rating was 1.59, indicating that poor textbooks are considered about midway between 1 = not a problem and 2 = a minor problem (Table 8).

Research Question #6: How do the teachers rate online technology tools?

The questionnaire included questions asking how often the teachers used computer-based instructional tools (item 1.5f), the extent to which insufficient access to computers is a problem in their school (item 2.1a), and how much they agreed or disagreed with the proposition that "Computer-based instructional tools (software) are helping Algebra I students in my Target Class" (item 1.6). These responses were examined by the grade level of the class and the standard school classification variables in the regression analysis (see Appendix Table C-6).

The data indicated that the average response to how frequently these tools are used was about 1 (= less than once a week) on a scale that ranged from 0 = never to 4 = everyday. The teachers' ratings of the helpfulness of computer-based instructional tools were mixed, with 29% agreeing somewhat or agreeing strongly with the proposition that computers were helpful and 38% disagreeing somewhat or disagreeing strongly (34% neither agreed nor disagreed).

Figure 8: Teachers' Ratings on Helpfulness of Computer-Based Instructional Tools in Algebra I Target Class: 2007



Source: Based on responses to the National Opinion Research Center's National Survey of Algebra Teachers, 2007.

Use of computers and access. The generally low levels of computer use does not appear to be a reflection of insufficient access. About half (49%) of the teachers reported that insufficient access to computers was not a problem in their schools and another 28% reported insufficient access to be a minor problem. Similar portions of those who do not feel access is a problem use computers less than once a week or never (74%) as do those who feel access is a serious problem (73%). This suggests that if those without access did get computers they would not use them much.

Table 9: Frequency of Using Computers in the Target Class, by Extent to Which Insufficient Access to Computers Is a Problem in the School: 2007

	How much of a problem is insufficient access to computers?				
Use of Computers and Software	**Not a Problem**	**Minor Problem**	**Moderate Problem**	**Serious Problem**	**Use Total**
Never	40.75%	46.80%	38.69%	51.72%	**43.40%**
Less than once a week	33.42%	33.17%	46.79%	20.58%	**33.66%**
About once a week	10.76%	9.49%	9.37%	9.02%	**10.03%**
Several times a week	6.62%	3.30%	1.14%	2.53%	**4.52%**
Everyday	8.47%	7.24%	4.00%	16.15%	**8.39%**
Total	**100%**	**100%**	**100%**	**100%**	**100%**

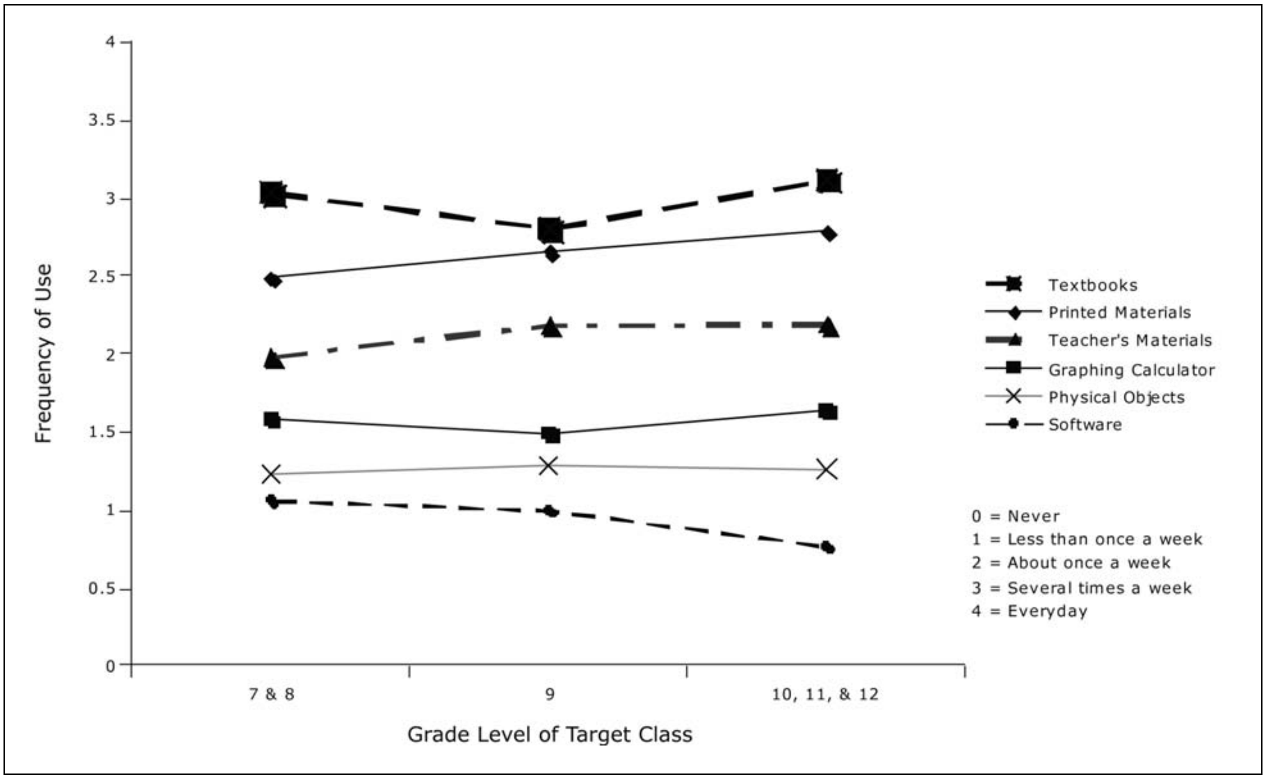
Note: Chi-square = 27.1 (p = 0.46), Correlation = 0.03 (p = 0.73)

CI = confidence interval, calculated as +/- two standard errors from the mean. Standard errors adjusted for design effects.

Source: Based on responses to the National Opinion Research Center's National Survey of Algebra Teachers, 2007.

Figure 9 shows the frequency of use of various materials across grades. As the chart shows, the level of use for texts and technology generally remains constant across grades. In other words, no matter what the age is of the students, the level of use for each material is about the same. Software is used least of all.

Figure 9: Frequency of Using Various Instructional Materials and Tools in Algebra I, by Grade Level of Target Class: 2007



Source: Based on responses to the National Opinion Research Center's National Survey of Algebra Teachers, 2007.

Research Question #7: What is the role of the calculator in the algebra course?

Questionnaire item 1.5d asked how often the teacher uses graphing calculators in her or his target class. Overall, 33% of the teachers report never using graphing calculators and another 29% report using them less than once a week. About 31% use them everyday (18%) or almost everyday (13%). (See Table 10).

Table 10 shows rates of graphing calculator use by grade and urbanicity. Teachers in urban schools were less likely to use graphing calculators than their suburban and rural counterparts, and teachers of eighth-grade Algebra I were more likely than others to use them in all three types of locale.

Table 10: Frequency of Graphing Calculator Use, by Grade Level of Target Class and Urbanicity: 2007

Frequency of Use	Grade 7 & 8	Grade 9	Grade 10–12	Total
Never	22.8%	39.4%	38.7%	**33.0%**
Less than once a week	41.9%	22.6%	15.6%	**29.4%**
About once a week	7.1%	5.7%	8.5%	**6.4%**
Several times a week	10.1%	14.2%	17.5%	**13.2%**
Everyday	17.4%	18.1%	19.7%	**18.0%**
Total	**100%**	**100%**	**100%**	**100%**
Sample Size (Total)	**128**	**518**	**73**	**719**
	Urban			
Never	**18.6%**	**39.4%**	**44.3%**	31.8%
Less than once a week	**44.4%**	**22.8%**	**17.8%**	30.7%
About once a week	**8.6%**	**6.4%**	**13.6%**	7.4%
Several times a week	**20.9%**	**19.9%**	**9.0%**	20.0%
Everyday	**7.5%**	**11.6%**	**15.3%**	10.1%
Total	100%	100%	100%	100%
Sample Size (Urban)	*37*	*202*	*10*	*249*
	Suburban			
Never	30.3%	44.8%	36.5%	**38.6%**
Less than once a week	43.3%	18.8%	10.1%	**26.7%**
About once a week	9.5%	7.2%	11.6%	**8.6%**
Several times a week	7.6%	11.3%	22.1%	**11.2%**
Everyday	9.3%	17.9%	19.7%	**15.0%**
Total	**100%**	**100%**	**100%**	**100%**
Sample Size (Suburban)	**66**	**247**	**55**	**368**
	Rural			
Never	18.0%	32.9%	42.1%	**27.2%**
Less than once a week	38.4%	27.0%	27.5%	**31.8%**
About once a week	3.3%	3.2%	0.0%	**3.0%**
Several times a week	6.9%	12.2%	9.5%	**9.9%**
Everyday	33.4%	24.8%	20.9%	**28.1%**
Total	**100%**	**100%**	**100%**	**100%**
Sample Size (Rural)	**25**	**69**	**8**	**102**

Note: Cells are weighted percentages within each urbanicity.

Source: Based on responses to the National Opinion Research Center's National Survey of Algebra Teachers, 2007.

Use of graphing calculators and access. While only about 30% of teachers use graphing calculators more than about once a week, many of those who use them with less frequency do report that access to this technology is a problem (Table 11). Of those that feel that access is not a problem, only 26% never use them. This contrasts with the over 50% that never use them among those who report insufficient access is a moderate or serious problem.

The correlation coefficient summarizing the linear relationship between the two items is moderately high (r = 0.32). This suggests that that if they had access, more—though by no means all—of the Algebra I teachers would use graphing calculators.

Table 11: Frequency of Using Graphing Calculators, by Extent to Which Insufficient Access to Graphing Calculators Is a Problem in the School: 2007

	How much of a problem is insufficient access to graphing calculators?				
Use of Graphing Calculators	**Not a Problem**	**Minor Problem**	**Moderate Problem**	**Serious Problem**	**Use Total**
Never	25.9%	32.1%	50.0%	58.1%	**32.7%**
Less than once a week	22.7%	42.7%	35.4%	23.2%	**29.6%**
About once a week	7.8%	2.7%	8.6%	4.7%	**6.5%**
Several times a week	14.6%	18.4%	2.3%	4.6%	**13.3%**
Everyday	29.0%	4.1%	3.7%	9.4%	**18.0%**
Total	**100%**	**100%**	**100%**	**100%**	**100%**

Chi-square = 121.6 ($p < .000$), Correlation = 0.32 ($p < 0.000$)

Source: Based on responses to the National Opinion Research Center's National Survey of Algebra Teachers, 2007.

Research Question #8: What about the use of manipulatives as instructional tools?

The relevant questionnaire item for this question asked how often the teacher uses physical objects (manipulatives) in her or his target class (item 1.5e). Overall, use of manipulatives on an occasional basis is widespread, but very few (9%) teachers report using them more than once a week. About 12% of the teachers reported never using manipulatives, and about 60% reported using them less than once a week (Table 12). As evident in Table 12, there does not seem to be a relationship between the class grade level and the frequency of use.

Table 12: Frequency of Physical Object Use, by Grade Level of Target Class: 2007

Frequency of Use	**Grade 7 & 8**	**Grade 9**	**Grade 10–12**	**Total**
Never	11.4%	12.9%	12.8%	**12.3%**
Less than once a week	62.1%	57.8%	53.7%	**59.1%**
About once a week	19.2%	18.5%	28.9%	**19.5%**
Several times a week	7.4%	10.1%	3.9%	**8.6%**
Everyday	0.0%	0.7%	0.7%	**0.4%**
Total	**100%**	**100%**	**100%**	**100%**
Sample Size	**128**	**518**	**73**	**719**

Note: Cells are weighted percentages.

Source: Based on responses to the National Opinion Research Center's National Survey of Algebra Teachers, 2007.

Views on Changing Secondary School Math Education

Research Question #9: How do teachers rate their professional training?

Questionnaire items pertaining to professional training and development include items 3.4a,b and possibly 4.19; items 2.1f and j are also relevant. These items were examined by the teachers' years of teaching experience, and school classification variables. With one exception, satisfaction with training did not vary by teacher characteristics; Hispanic teachers reported more satisfaction with preservice training by .64 standard deviations.

Looking at Table 13, teachers reported that inadequate preparation to teach Algebra I and inadequate professional development opportunities for Algebra I teachers are not problems for the teachers in their respective schools. When asked to report on their own preservice and professional development experiences as preparations for teaching Algebra I, Figures 10 and 11 show that most teachers evaluated the experiences as preparing themselves adequately or very well. However, substantial minorities of the teachers indicated that improvements can be made in preservice training programs and professional development opportunities.

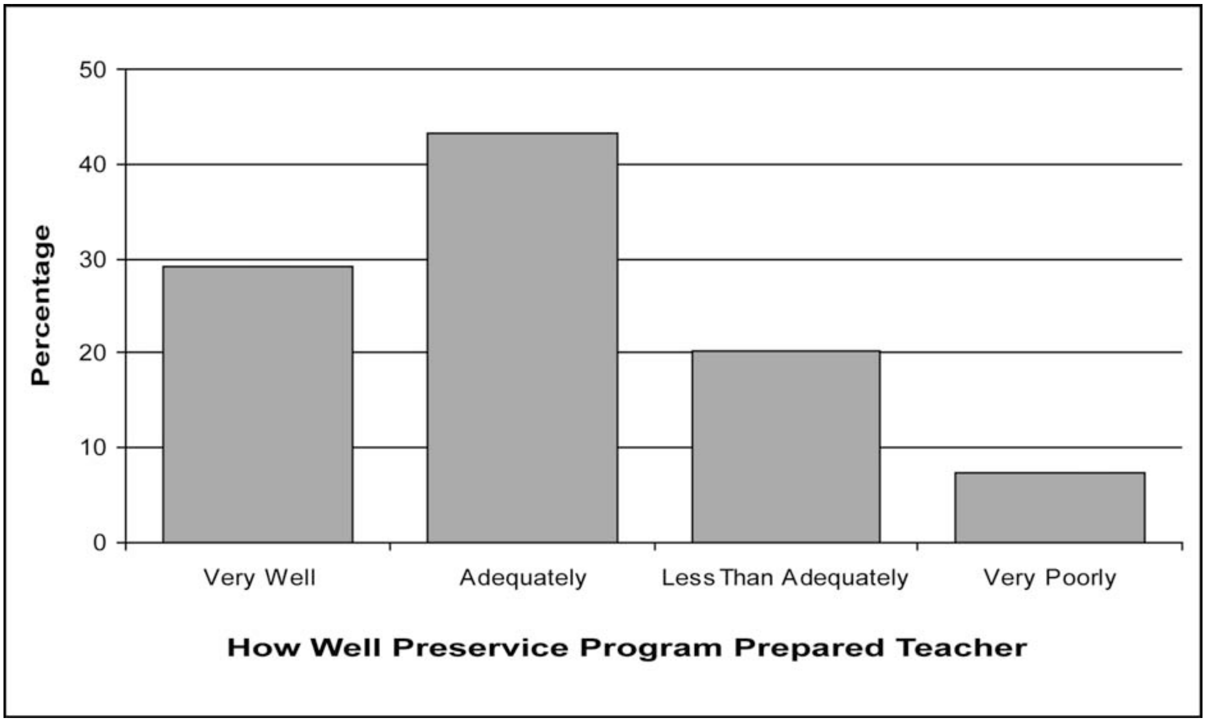
Table 13: Teachers' Evaluation of Selected Professional Development Factors: 2007

Professional Development Factor	**Scale**	**Mean**	**95% CI**	
			Low	**High**
Inadequately prepared teachers	1 = Not a Problem … 4 = Serious Problem	1.49	1.43	1.55
Inadequate opportunities for professional development	1 = Not a Problem … 4 = Serious Problem	1.65	1.59	1.71
Rating of own pre-service teacher education	1 = Prepared Teacher Very Well … 4 = Very Poorly	1.96	1.89	2.02
Rating of own professional development opportunities	1 = Help Teach Very Well … 4 = Very Poorly	1.98	1.91	2.04

Note: CI = confidence interval, calculated as +/- two standard errors from the mean. Standard errors adjusted for design effects.

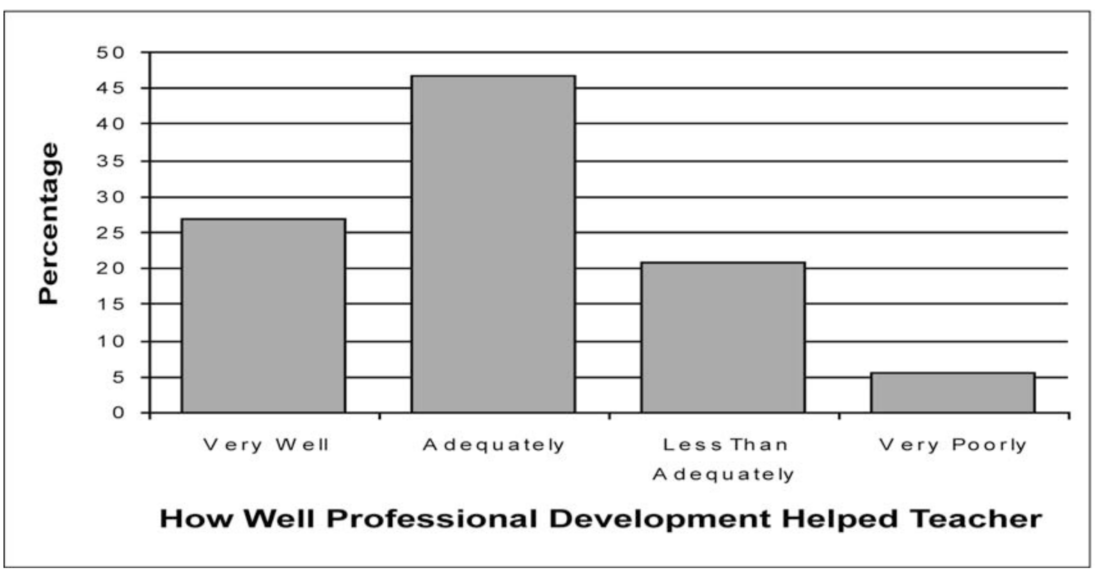
Source: Based on responses to the National Opinion Research Center's National Survey of Algebra Teachers, 2007.

Figure 10: Distribution of Teachers' Ratings of How Well Their Preservice Education Program Prepared Them to Teach Algebra I: 2007



Source: Based on responses to the National Opinion Research Center's National Survey of Algebra Teachers, 2007.

Figure 11: Distribution of Teachers' Ratings of How Well Their Professional Development Opportunities Have Helped Them Teach Algebra I: 2007



Source: Based on responses to the National Opinion Research Center's National Survey of Algebra Teachers, 2007.

Research Question #10: Is there sufficient and effective remedial help for students who are struggling in algebra? What sort of assistance-based interventions would struggling students benefit from the most?

Questionnaire items 2.8a–b asked the teachers to rate the availability and quality of tutoring or other remedial services for students struggling with Algebra I in their school. The average ratings by the school classification variables were examined.

- On average, looking at Table 14, teachers were generally satisfied with the services available, even if not extremely so.
- These services were rated more favorably by teachers in high minority schools.
- Female and black teachers were less satisfied with their schools' remedial services.

Table 14: Teachers' Ratings on Availability and Quality of Remedial Help for Algebra I Students: 2007

Evaluation of Remedial Help	**Lower 95% CI**	**Mean**	**Higher 95% CI**
Availability of remedial help	2.35	**2.52**	2.69
Quality of remedial help	2.26	**2.42**	2.58

Note: Scale: 1 = Excellent, 2 = Good, 3 = Fair, 4 = Poor

CI = confidence interval, calculated as +/- two standard errors from the mean. Standard errors adjusted for design effects.

Source: Based on responses to the National Opinion Research Center's National Survey of Algebra Teachers, 2007.

Research Question #11: Do teachers believe that students would learn more if they were grouped by ability for instruction, or is this approach counterproductive?

Questionnaire item 2.2 asked whether the school offers different levels of Algebra I based on ability; 46% of the teachers indicated their schools did differentiate. Questionnaire item 2.1h asked teachers to rate the extent to which they see different levels of students in the same class as a problem in their school.

A substantial number of teachers considered mixed-ability groupings to be a "moderate" (28%) or "serious" (23%) problem (see Figure 12). Teachers in schools that did not offer different levels of Algebra I based on ability were more likely than their counterparts in schools that do use ability grouping to consider mixed-ability classrooms to be a moderate or serious problem (Table 15).

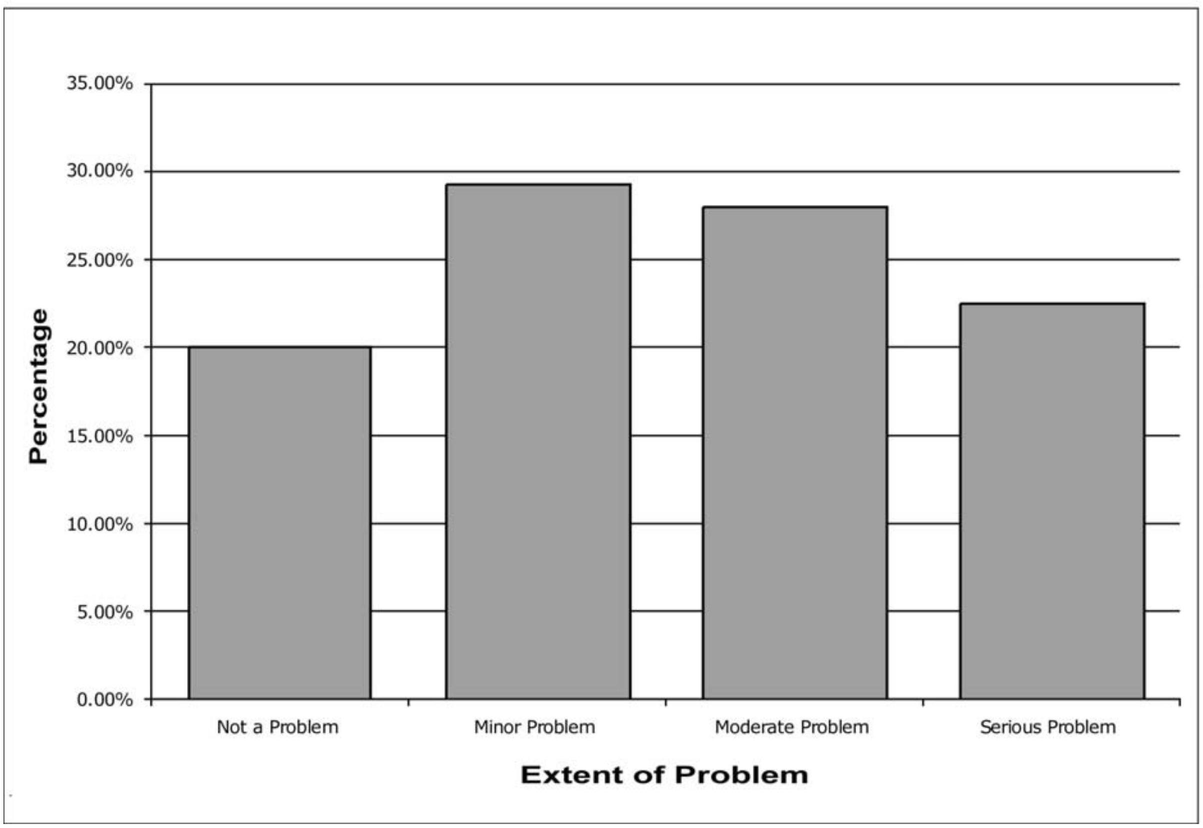
Table 15: Percentage of Algebra I Teachers Reporting Students With Different Abilities And Skills Taking the Same Class is a Problem, by Whether School Offers Different Levels Based on Ability: 2007

Level of Problem	Available at Teachers' School	Not Available at Teachers' School	All Teachers
Not a problem	21.3%	19.3%	**20.2%**
Minor problem	33.4%	25.9%	**29.4%**
Moderate problem	26.2%	29.5%	**27.9%**
Serious Problem	19.2%	25.4%	**22.5%**
Total	**100%**	**100%**	**100%**

Note: Twelve respondents did not know whether or not their school mixed ability levels.

Source: Based on responses to the National Opinion Research Center's National Survey of Algebra Teachers, 2007.

Figure 12: Extent to Which Students With Different Abilities and Interests Taking the Same Algebra I Class Is a Problem: 2007



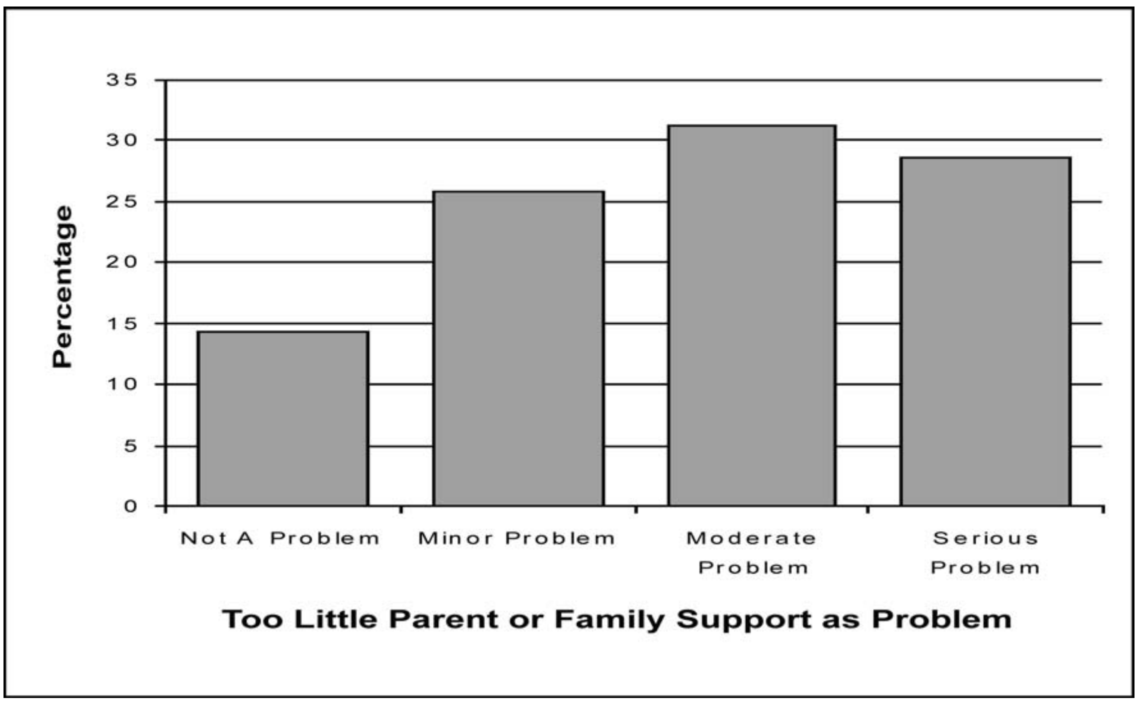
Source: Based on responses to the National Opinion Research Center's National Survey of Algebra Teachers, 2007.

Looking at Appendix Table C-9, the data indicate that for larger classes, high school teachers do feel that mixed-ability classes are a problem. Also, data obtained indicate that black teachers were more favorable of the practice. In this case teachers were describing their feelings about the practice in general. Teachers with larger classes and later grades are less likely to feel that it is a good practice.

Research Question #12: Do teachers find more parents helpful in encouraging students in their mathematics studies, or do too many parents make excuses for their children's lack of accomplishment?

Questionnaire item 2.1i asked teachers to rate the extent to which they see "too little parent/family support" as a problem in their school. The data in Figure 13 shows that more teachers feel that family participation is a moderate (32%) or serious (28%) problem than feel it is a minor problem (26%) or not a problem at all (14%).

Figure 13: Extent to Which Too Little Parent or Family Support Is a Problem in School: 2007



Source: Based on responses to the National Opinion Research Center's National Survey of Algebra Teachers, 2007.

To estimate relationships between the teachers' family participation rating and the teacher and school background variables, regression analysis was used (see Appendix Table C-10). High school teachers were much more likely than middle school and other teachers to report lack of family participation as a problem (the effect size is 0.65 SD units). Also, teachers in schools with higher percentages of free and reduced-priced lunch students also felt that lack of family participation was more of a problem, the second quartile by .31 standard deviations, the third by .46 SD units, and the fourth quartile by .54 SD units. Female teachers, on the other hand, feel that lack of family participation is less of a problem by .22 standard deviations.

Research Question #13: What do teachers see as the single most challenging aspect of teaching Algebra I successfully?

This question (4.20) included 10 response options: explaining material to students, handling accelerated students, teaching procedures, explaining concepts, using diagrams or models effectively, interpreting student errors and difficulties, working with unmotivated students, working with advanced students, helping students whose home language is not English, making mathematics accessible and comprehensible, and an "other" option.

Table 16 shows the percentages of each response for the categories of high schools and middle/other schools. The overwhelmingly most frequent response to this question was "working with unmotivated students." This was chosen by 65% of the high school teachers and 58% of the middle school teachers.

Table 16: Frequencies of Reported Challenges to Teaching Algebra I by Class Grade Level and Type of School: 2007

Reported Challenge	**High School Teachers**	**Middle/Other School Teachers**	**High and Middle/ Other School Teachers**
Working with unmotivated students	65.4%	58.2%	61.8%
Making mathematics accessible and comprehensible	9.1%	13.6%	11.3%
Explaining concepts	5.5%	3.1%	4.4%
Explaining material to struggling students	2.1%	4.1%	3.1%
Interpreting students errors and difficulties	0.3%	2.7%	1.5%
Handling accelerated students	1.4%	1.4%	1.4%
Helping students whose home language is different than English	1.6%	0.6%	1.1%
Using diagrams or models effectively	0.5%	1.4%	0.9%
Working with advanced students	0.0%	1.2%	0.6%
Teaching procedures	0.0%	0.6%	0.3%
Other, verbatim responses	14.1%	13.2%	13.7%
Sample Size	**100%**	**100%**	**100%**
Column N	**530**	**207**	**737**

Note: Cells are weighted percentages.

Source: Based on responses to the National Opinion Research Center's National Survey of Algebra Teachers, 2007.

The next most frequent response was "making mathematics accessible and comprehensible to all my students," selected by 14% of the middle school teachers and 9% of the high school teachers.

Many teachers wrote in additional challenges in response to this question. The written-in verbatim responses most often mentioned included handling different skill levels in a single classroom, motivation issues, and student study skills. Some notable responses were:

> *Walking into a class of 30 students in which 1/3 of them don't have the prerequisite skills necessary to be in the class. Many of whom don't know their basic arithmetic facts and know they aren't going to be successful from day one no matter how hard they try.*
>
> *Students come to me without a basic understanding of math. I am constantly reteaching concepts that should have been mastered in the earlier grades.*
>
> *Parents not letting me do my job as I see fit. (Autonomy in the classroom.)*
>
> *Getting students and parents to believe that education is important. Students don't do their homework ... you call the parents ... they say that the student will start doing the work (and coming to tutorials). The students still don't do the h.w.— and still don't come to tutorials.*
>
> *Engaging students who have come to believe that they are stupid because they are struggling with my state's cognitively inappropriate standards.*

NORC staff examined whether there is a relationship between the types of challenges identified and the experience of an algebra teacher. Table 17 displays the percentages selecting the three most frequently-selected responses separately by the teacher's years of teaching experience. The differences among age groups in the percentages selecting "working with unmotivated students" were slight and not statistically significant; this is evidently not a challenge related to teaching experience. In contrast, the least experienced teachers were more likely than others to identify "making mathematics accessible and comprehensible" as their greatest challenge (18%). The most experienced teachers were much less likely to view that as their greatest challenge (6%).

Table 17: Reported Challenges to Teaching Algebra I by Years of Experience: 2007

Reported Challenges	Years of Experience				All Teachers
	Up to 3	**4 to 9**	**10 to 18**	**19 or more**	
Working with unmotivated students	61.3%	60.0%	61.4%	65.6%	**61.6%**
Making mathematics accessible and comprehensible	17.5%	7.8%	11.9%	6.0%	**11.3%**
Other and Rest of Items	21.2%	32.3%	26.7%	28.3%	**27.0%**
Total	**100%**	**100%**	**100%**	**100%**	**100%**
Column N	**209**	**229**	**167**	**122**	**727**

Note: Cells are weighted percentages.

Source: Based on responses to the National Opinion Research Center's National Survey of Algebra Teachers, 2007.

III. Summary and Conclusions

The main findings of the survey can be summarized in terms of the guiding research questions for the project.

Student Preparation. The first question concerned the adequacy of student preparation coming into the Algebra I classes. In an important sense, any rating of the knowledge areas and skills asked about in the questionnaire of less than "good" represents an important problem that should be addressed in the math classes leading up to Algebra I. The topics that were rated as especially problematic were rational numbers, solving word problems, and basic study skills. But the only item that had an average rating better than "good" was "whole number operations." Coupled with the teachers' verbatim responses to the question asking for changes they would like to see in the curriculum leading up to Algebra I (item 3.2), the teachers indicate that students are often ill prepared to think about how to solve novel or more complex problems than familiar arithmetic operations. In sum, the teachers generally rate their students' background as less than satisfactory, and this no doubt poses additional challenges to teaching Algebra I.

The teachers' ratings of student preparation varied mainly according the grade level of the students, with preparation rated highest for the Grade 7 and 8 Algebra I classes and rated lowest for the Grade 10 and higher classes. This likely reflects the ability-grouping regime, whereby the higher achievers take the class earlier. The staggering of entry grades is intended to enable each group of students to reach a good level of preparation for success, and not simply open the way for the highest achievers to advance through the high school mathematics curriculum. In any case, these finding emphasize the importance of improving student performance among those entering Algebra I after the eighth grade.

Curriculum and Instruction. In contrast to their views on student preparation, the teachers are relatively favorable about the algebra curriculum and instructional materials at their disposal. Local expectations for student proficiency in algebra are viewed as reasonable, and local and state content and assessment standards for algebra are generally regarded favorably. The teachers gave their textbooks high average marks on all aspects identified in the questionnaire. The composite-scale ratings were somewhat less favorable among teachers in schools with higher minority student enrollments, and this likely reflects a more negative evaluation among those teachers on the specific point of how adequately "the textbook and accompanying materials provide useful suggestions for meeting the needs of diverse learners" (item 1.8.i. and see Appendix Table C-8).

The teachers generally reported favorable views of their own preservice training for teaching and of the helpfulness of the in-service professional development opportunities they have had. At the same time, it should be noted that about a quarter of the teachers evaluated their preservice as "less than adequate" or "very poor," and about the same number rated their in-service professional development as such. Further analysis to try to identify systematic factors related to those negative evaluations is needed in order to suggest remedies.

Views on Changing Secondary School Math Education. When asked to identify the single most challenging aspect of teaching Algebra I successfully, the teachers overwhelmingly indicated "working with unmotivated students." This was selected by 62% of the teachers; the next most frequent item was "making mathematics accessible and comprehensible to all my students" selected by a distant-second 11% of the teachers.

In light of the generally favorable views the teachers report with respect to curriculum and instruction, the issue of unmotivated students implicitly is something the teachers view as more of a "algebra-student problem" than an "algebra-teacher problem." The generally negative views expressed by the teachers of parental support for mathematics reinforce that attribution. Taken together with the generally negative ratings of background preparation, the lack of student motivation suggests that careful attention to pre-algebra curriculum and instruction in the elementary grades is needed, both to remedy the specific skill deficiencies as well as to identify ways in which negative attitudes toward mathematics are developed.

APPENDIX A: Survey Methodology

In February 2007, NORC began work under direction of the National Mathematics Advisory Panel established within the U.S. Department of Education to conduct the National Survey of Algebra Teachers (NSAT). The main tasks on the project were to (a) develop the survey instrument, (b) design the sampling plan and draw the sample, (c) collect rosters of the Algebra I teachers in each school, (d) contact the teachers and collect the survey data, and (e) produce data files and perform statistical analysis. This section summarizes these activities.

Instrument Development

The questionnaire development was done in close consultation with the National Math Panel to ensure that key areas of analytic interest were covered. A first draft of the NSAT questionnaire was assembled by NORC and submitted to the Panel in early February. This draft included questions directly mapped to the key items identified by the Panel, as well as additional items which helped develop the key research questions or provide analytical leverage in addressing them. These items were drawn from a variety of sources including the Education Longitudinal Study of 2002 (Teacher questionnaire), the National Education Longitudinal Study of 1988, the National Education Association's Status of the American Public School Teacher 2000–01 survey, the Consortium for Chicago School Research 2005 teacher survey, and the Longitudinal Study of American Youth (LSAY, beginning in 1987) math teacher questionnaires.

NORC project staff then met with local Chicago-area teachers, other education researchers with experience on mathematics teacher surveys, and NORC questionnaire design experts to test the instrument and obtain feedback. In general, the teachers responded positively to the survey and had a few minor changes to the wording and ordering of the questions. Almost all of the teachers interviewed wanted additional items or questions added that focused on the pre-algebra skills. They provided a list of additional questions targeted towards students' pre-algebra skills. NORC's questionnaire design team had few issues with the content of the questions being asked, and they provided essential feedback on questionnaire wording and answer categories. Additionally, they suggested that a few items be dropped (see the comments in the questionnaire), either due to their repetitive nature or because they did not add much analytic value.

Comments from the Panel on the first draft of the survey were received by NORC mid-February. NORC incorporated comments provided by the Panel, the teachers, and NORC's questionnaire design team into the second draft of the questionnaire. The final version of the questionnaire was submitted for OMB approval on February 20, 2007.

Sampling

NORC utilized the U.S. Department of Education's Common Core of Data (CCD) file for the 2004–05 school year (this was the most recent year available as of February 2007) to compile the sample frame of public schools. All schools listed in the CCD as located within the 50 states and District of Columbia with an eighth grade or higher, and which were not classified by CCD as special education, vocational education centers, or alternative schools were considered eligible for the sample.

To ensure the sample would represent public school Algebra I teachers in different types of schools and settings across the country, the frame was stratified by four variables, all defined from data included in the CCD file:

1) *Type of locale*. A standard three-level indicator of urban, suburban, or rural school location was used for this variable.

2) *Percentage of students eligible to receive free or reduced-price lunch.* This was simplified to a dichotomous indicator of "40 percent or lower" versus "more than 40 percent."

3) *Percentage of students who are black, Hispanic, and American Indian*. This was also simplified to a dichotomous indicator of "40 percent or lower" versus "more than 40 percent."

4) *Graded configuration of the school*. Since Algebra I instruction starts in earnest in the eighth grade and continues throughout high school, eligible school configurations include K–8 elementary schools, grade 6–8 middle schools, grade 7–9 junior high schools, grade 9–12 and 10–12 high schools, and K–12 combined elementary and secondary schools. The various configurations were trichotomized into "grade 9–12 and 10–12 high schools," "K–8 elementary schools, grade 6–8 middle schools, and grade 7–9 junior high schools," and "all other schools where Algebra I is taught."

The cross-classification of the stratification variables created 36 sampling strata. Approximately 2,300 of the 36,353 eligible schools were missing information on the percentage of students eligible for free or reduced-price lunch, and a total of 440 of the New York City Public School District schools were listed as having zero students eligible. Since this is certainly incorrect for many if not most of these NYCPSD schools, NORC staff recoded the *Percentage of Students Eligible to Receive Free or Reduced-price Lunch* from 0 to missing for all of them. To mitigate the impact of the missing data on the sample design, the missing data was replaced with the same data from the 2003–04 school year CCD file if available. If the data were also missing in the 2003–04 CCD, the missing data was replaced with data from the 2002–03 CCD if available. After consultation with the NMP it was decided to define a special supplemental stratum consisting of schools with missing stratification data in the final sample file, and to sample schools from that stratum.

Target numbers of 300 schools and 1,000 Algebra I teachers were defined for the survey, based on project objectives and statistical power calculations. These targets were supplemented with a target of 10 schools and 40 teachers from the missing data stratum noted above. To select the sample, the target number of 310 schools was systematically sampled from the frame with the selection probability proportional to the estimated number of Algebra I teachers per school. The number of Algebra I teachers per school was estimated on the basis of grade-specific enrollment data from the CCD, coupled with data on the number of Algebra I teachers collected in February from a small sample of schools and average rates of Algebra I course-taking and class-size data obtained from recent national surveys. Because the schools were selected with probability proportional to the number of Algebra I teachers, schools with more Algebra I teachers are more likely to be selected into the sample. Therefore, a fixed number of sample schools will represent a greater number of teachers than under simple random sampling.

Roster and Data Collection

On March 21, 2007, NORC mailed letters to all district superintendents and principals of the selected school. This letter informed them that a school in their district (for superintendents) or their school (for principals) had been selected to participate in the study and alerted them that a NORC staff member would be calling the school in the next few weeks to obtain roster information on their Algebra I staff. The letter also included NORC's contact information should the district or school like to request more information on the study. NORC began roster collection on March 26. This process included collecting Algebra I teacher information (names, e-mails, number of Algebra I classes taught, other classes taught, last day of school) from either the school principal, the office secretary, or the head of the math department. At this point, it was determined if a school was ineligible or if a school refused to participate. Refused or ineligible schools were replaced with other schools with the same strata qualifications. Of the 300 schools in the original sample, 52 schools had to be replaced. Ineligible supplemental samples were not replaced. Rosters were collected from a total of 258 schools. All data collected were entered into a receipt control system which also helped to keep track of sent and returned mail to districts, principals, and teachers. This system was also utilized to track and prompt nonrespondents to the survey during data collection.

The following table breaks down the number of rosters collected by the possible 36 different strata, as well as three additional schools drawn from those lacking information on the number of students eligible for the federal free and reduced-price lunch program.

Table A-1: Numbers of Sampled Schools, Schools That Provided Rosters of Algebra I Teachers, and Algebra I Teachers, by Sample Stratum: 2007

Strata	Total # of Schools in Sample	Total # of Schools That Provided Roster Information	Total # of Teachers
Missing FRPL Information	3	2	12
1. Rrl HS < 40 % Mnr & < 40 % FRPL	25	22	70
2. Rrl HS < 40 % Mnr & > 40 % FRPL	6	6	17
3. Rrl HS > 40 % Mnr & < 40 % FRPL	2	2	5
4. Rrl HS > 40 % Mnr & > 40 % FRPL	5	4	10
5. Rrl M/JH < 40 % Mnr & < 40 % FRPL	7	7	17
6. Rrl M/JH < 40 % Mnr & > 40 % FRPL	4	3	4
7. Rrl M/JH > 40 % Mnr & < 40 % FRPL	1	0	0
8. Rrl M/JH > 40 % Mnr & > 40 % FRPL	2	2	4
9. Rrl OtherS < 40 % Mnr & < 40 % FRPL	8	8	18
10. Rrl OtherS < 40 % Mnr & > 40 % FRPL	4	4	7
11. Rrl OtherS > 40 % Mnr & < 40 % FRPL	1	1	2
12. Rrl OtherS > 40 % Mnr & > 40 % FRPL	2	2	3
13. Srb HS < 40 % Mnr & < 40 % FRPL	61	51	233
14. Srb HS < 40 % Mnr & > 40 % FRPL	5	5	18
15. Srb HS > 40 % Mnr & < 40 % FRPL	12	7	56
16. Srb HS > 40 % Mnr & > 40 % FRPL	16	11	63
17. Srb M/JH < 40 % Mnr & < 40 % FRPL	23	22	57
18. Srb M/JH < 40 % Mnr & > 40 % FRPL	7	6	15
19. Srb M/JH > 40 % Mnr & < 40 % FRPL	2	1	5
20. Srb M/JH > 40 % Mnr & > 40 % FRPL	10	9	17
21. Srb OtherS < 40 % Mnr & < 40 % FRPL	7	5	12
22. Srb OtherS < 40 % Mnr & > 40 % FRPL	1	0	0
23. Srb OtherS > 40 % Mnr & < 40 % FRPL	1	1	9
24. Srb OtherS > 40 % Mnr & > 40 % FRPL	3	3	20
25. Urb HS < 40 % Mnr & < 40 % FRPL	18	16	82
26. Urb HS < 40 % Mnr & > 40 % FRPL	3	2	14
27. Urb HS > 40 % Mnr & < 40 % FRPL	9	8	48
28. Urb HS > 40 % Mnr & > 40 % FRPL	28	18	136
29. Urb M/JH < 40 % Mnr & < 40 % FRPL	5	5	12
30. Urb M/JH < 40 % Mnr & > 40 % FRPL	4	3	10
31. Urb M/JH > 40 % Mnr & < 40 % FRPL	1	1	1
32. Urb M/JH > 40 % Mnr & > 40 % FRPL	14	12	25
33. Urb OtherS < 40 % Mnr & < 40 % FRPL	2	2	12
34. Urb OtherS < 40 % Mnr & > 40 % FRPL	1	1	6
35. Urb OtherS > 40 % Mnr & < 40 % FRPL	1	1	4
36. Urb OtherS > 40 % Mnr & > 40 % FRPL	6	5	16
All Strata	310	258	1,040

Note: FRPL: Free and Reduced-Price Lunch
Mnr: Minority
Rrl HS: Rural High School
Rrl M/JH: Rural Middle/Junior High
Srb HS: Suburban High School
Srb M/JH: Suburban Middle/Junior High
Urb HS: Urban High School
Urb M/JH: Urban Middle/Junior High

Source: Based on responses to the National Opinion Research Center's National Survey of Algebra Teachers, 2007.

Because roster collection was an ongoing process, NORC conducted the necessary mail outs in batches to collect the teacher information. Prior to mailing the questionnaires to the teachers, NORC sent out prenotice letters informing the teachers of the survey, and notifying them that their principals had consented for them to participate. A week later each teacher was sent (via FedEx) a questionnaire, along with a $20 check, a business reply envelope, and a letter informing them of the survey and requesting their participation. A week after each initial questionnaire mailing NORC sent out a postcard to all teachers reminding them of the survey and requesting their participation. This was followed approximately two weeks later by a second questionnaire mailing to all nonrespondents. NORC staff began phone and e-mail prompting of all remaining nonrespondents at this time. Marian Banfield provided assistance in the prompting process by sending out e-mails from the Department of Education to teachers requesting their participation. A final, third questionnaire was sent one to two weeks after the second questionnaire depending on when the school was going to be closed for the summer. Appendix Table A-2 summarizes the exact mail-out dates for each mail-out cohort or batch.

Table A-2: Questionnaire and Follow-Up Mailing Dates and Numbers of Algebra I Teachers, by Mail-Out Cohort: 2007

Disposition	**Cohort 1**	**Cohort 2**	**Cohort 3**	**Cohort 4**	**Cohort 5**	**Cohort 6**	**Cohort 7**	**Total**
# of Teachers	**147**	**147**	**189**	**274**	**134**	**68**	**81**	**1040**
Prenotice	4/9/2007	4/16/2007	4/23/2007	4/30/2007	5/7/2007	5/14/2007	5/21/2007	1040
Quex 1 Mail Out	4/17/2007	4/20/2007	4/25/2007	5/2/2007	5/10/2007	5/16/2007	5/23/2007	1040
Post card								
Mail-Out Date	4/27/2007	4/27/2007	5/4/2007	5/11/2007	5/17/2007	5/25/2007	6/1/2007	
# Mailed	136	147	183	262	134	68	68	998
Quex 2 Mail Out								
Mail-Out Date	5/9/2007	5/9/2007	5/16/2007	5/18/2007	5/23/2007	6/1/2007	6/8/2007	
# Mailed	64	76	120	178	94	56	64	652
Quex 3 Mail Out								
Mail-Out Date	5/23/2007	5/23/2007	5/30/2007	6/1/2007	6/8/2007	6/15/2007	6/22/2007	
# Mailed	39	49	77	98	55	35	38	391

Source: Based on responses to the National Opinion Research Center's National Survey of Algebra Teachers, 2007.

Response Rates

Of the 1,040 teachers NORC prompted to complete the survey, 743 completed questionnaires were received. An additional 14 teachers also notified us that they, in fact, were not Algebra I teachers and therefore were ineligible to participate in the survey, while two teachers explicitly refused to participate. Appendix Table A-3 provides a breakdown of how many teachers completed the survey by each of the four sample stratification variables, and Appendix Table A-4 shows the results for each of the 36 strata.

Table A-3: Number of Algebra I Teachers Sampled, Ineligible, Refusing, and Completing the Questionnaire, and Survey Response Rate, by Sample Stratification Variables: 2007

Sample Stratification Variable	Total # of Teachers	Total # of Teachers Who Are Ineligible	Total # of Teachers Who Refused to Complete Questionnaire	Total # of Teachers Who Completed Questionnaire	Response Rate (%)
Urbanicity					
Urban	505	6	2	370	74.1%
Suburban	366	7	0	251	69.9%
Rural	157	1	0	110	70.5%
School Type					
High School	752	12	1	521	70.4%
Middle School or Junior High	167	1	1	128	77.1%
Other Type of School	109	1	0	82	75.9%
Percentage of Students Who Are Minority					
Less than 40%	604	10	2	432	72.7%
More than 40%	424	4	0	299	71.2%
Percentage of Students Who Are Eligible or Receive Free or Reduced-Price Lunch					
Less than 40%	643	7	2	462	72.6%
More than 40%	385	7	0	269	71.2%

Note: Response rates were calculated on the basis of eligible teachers.

Source: Based on responses to the National Opinion Research Center's National Survey of Algebra Teachers, 2007.

Table A-4: Number of Algebra I Teachers Sampled, Ineligible, Refusing, and Completing the Questionnaire, and Survey Response Rate, by Sample Stratum: 2007

Sample Stratum	Total # of Teachers	Total # of Teachers Who Are Ineligible	Total # of Teachers Who Refused to Complete Questionnaire	Total # of Teachers Who Completed Questionnaire	Response Rate (%)
Supplemental Stratum (Missing Data on FRPL)	12	0	0	12	100
1. Rrl HS < 40 % Mnr & < 40 % FRPL	70	1	0	45	65.2
2. Rrl HS < 40 % Mnr & > 40 % FRPL	17	0	0	9	52.9
3. Rrl HS > 40 % Mnr & < 40 % FRPL	5	0	0	3	60.0
4. Rrl HS > 40 % Mnr & > 40 % FRPL	10	0	0	8	80.0
5. Rrl M/JH < 40 % Mnr & < 40 % FRPL	17	0	0	14	82.4
6. Rrl M/JH < 40 % Mnr & > 40 % FRPL	4	0	0	3	75.0
7. Rrl M/JH > 40 % Mnr & < 40 % FRPL	0	0	0	0	N/A
8. Rrl M/JH > 40 % Mnr & > 40 % FRPL	4	0	0	3	75.0
9. Rrl OtherS < 40 % Mnr & < 40 % FRPL	18	0	0	14	77.8
10. Rrl OtherS < 40 % Mnr & > 40 % FRPL	7	0	0	7	100
11. Rrl OtherS > 40 % Mnr & < 40 % FRPL	2	0	0	2	100
12. Rrl OtherS > 40 % Mnr & > 40 % FRPL	3	0	0	2	66.7
13. Srb HS < 40 % Mnr & < 40 % FRPL	233	4	1	167	72.9
14. Srb HS < 40 % Mnr & > 40 % FRPL	18	1	0	12	70.6
15. Srb HS > 40 % Mnr & < 40 % FRPL	56	0	0	40	71.4
16. Srb HS > 40 % Mnr & > 40 % FRPL	63	1	0	50	80.6
17. Srb M/JH < 40 % Mnr & < 40 % FRPL	57	0	1	43	75.4
18. Srb M/JH < 40 % Mnr & > 40 % FRPL	15	0	0	10	66.7
19. Srb M/JH > 40 % Mnr & < 40 % FRPL	5	0	0	3	60.0
20. Srb M/JH > 40 % Mnr & > 40 % FRPL	17	0	0	13	76.5
21. Srb OtherS < 40 % Mnr & < 40 % FRPL	12	0	0	8	66.7
22. Srb OtherS < 40 % Mnr & > 40 % FRPL	0	0	0	0	N/A
23. Srb OtherS > 40 % Mnr & < 40 % FRPL	9	0	0	7	77.8
24. Srb OtherS > 40 % Mnr & > 40 % FRPL	20	0	0	17	85.0
25. Urb HS < 40 % Mnr & < 40 % FRPL	82	1	0	59	72.8
26. Urb HS < 40 % Mnr & > 40 % FRPL	14	2	0	8	66.7
27. Urb HS > 40 % Mnr & < 40 % FRPL	48	1	0	31	66.0
28. Urb HS > 40 % Mnr & > 40 % FRPL	136	1	0	89	65.9
29. Urb M/JH < 40 % Mnr & < 40 % FRPL	12	0	0	9	75.0
30. Urb M/JH < 40 % Mnr & > 40 % FRPL	10	0	0	10	100
31. Urb M/JH > 40 % Mnr & < 40 % FRPL	1	0	0	1	100
32. Urb M/JH > 40 % Mnr & > 40 % FRPL	25	1	0	19	79.2
33. Urb OtherS < 40 % Mnr & < 40 % FRPL	12	0	0	12	100
34. Urb OtherS < 40 % Mnr & > 40 % FRPL	6	1	0	2	40.0
35. Urb OtherS > 40 % Mnr & < 40 % FRPL	4	0	0	4	100
36. Urb OtherS > 40 % Mnr & > 40 % FRPL	16	0	0	7	43.8
Total	1,040	14	2	743	72.4

Note: Response rates were calculated on the basis of eligible teachers.

FRPL: Free and Reduced-Price Lunch
Mnr: Minority
Rrl HS: Rural High School
Rrl M/JH: Rural Middle/Junior High
Srb HS: Suburban High School
Srb M/JH: Suburban Middle/Junior High
Urb HS: Urban High School
Urb M/JH: Urban Middle/Junior High

Source: Based on responses to the National Opinion Research Center's National Survey of Algebra Teachers, 2007.

APPENDIX B: Table Means for Survey Variables, By School Classification Variables

Table B-1: Means for Survey Variables by School Locale

Variable Name	Variable Label	Overall Sample		Locale					
				Urban		Suburban		Rural	
		Wtd. Mean	Wtd. SD	Wtd. Mean	Wtd. SE	Wtd. Mean	Wtd. SE	Wtd. Mean	Wtd. SE
TC_Student	Target Class - total number of students	2.98	1.16	3.26	0.07	3.29	0.07	2.39	0.07
TC_Studnt7	Target Class 7th-grade students	0.21	0.80	0.29	0.08	0.24	0.05	0.13	0.05
TC_Studnt8	Target Class 8th-grade students	1.65	1.93	1.67	0.16	1.56	0.12	1.75	0.13
TC_Studnt9	Target Class 9th-grade students	2.00	1.73	2.11	0.13	1.98	0.11	1.92	0.11
TC_Studt10	Target Class 10th-grade students	0.68	0.91	0.60	0.05	0.81	0.07	0.59	0.06
TC_Studt11	Target Class 11th-grade students	0.33	0.52	0.30	0.04	0.36	0.04	0.31	0.03
TC_Studt12	Target Class 12th-grade students	0.17	0.39	0.23	0.04	0.17	0.03	0.15	0.03
TC_StudtSE	Target Class special ed students	0.61	0.69	0.63	0.06	0.61	0.05	0.59	0.04
TC_StudtBi	Target Class bilingual students	0.34	0.73	0.43	0.06	0.36	0.05	0.25	0.05
Come_Time	Come to class on time	3.57	0.61	3.26	0.05	3.59	0.04	3.80	0.03
Attend_Reg	Attend class regularly	3.47	0.63	3.28	0.05	3.42	0.04	3.66	0.04
Come_Prep	Come to class prepared	2.92	0.90	2.53	0.07	2.92	0.05	3.23	0.04
Creat_Prob	Create serious behavior problems	0.61	0.68	0.76	0.04	0.62	0.04	0.48	0.04
Pay_Attn	Regularly pay attention	2.82	0.83	2.63	0.06	2.80	0.05	2.98	0.05
Activ_Part	Actively participate	2.69	0.89	2.64	0.06	2.61	0.05	2.81	0.06
Take_Note	Take notes	2.72	1.01	2.68	0.07	2.60	0.06	2.90	0.06
Diff_ReadE	Serious difficulties reading English	0.47	0.64	0.69	0.05	0.46	0.04	0.31	0.04
Care_grade	Care about what grade received	2.90	0.88	2.69	0.08	2.92	0.05	3.04	0.04
Whole_Numb	Whole number-background	1.86	0.80	2.01	0.06	1.86	0.05	1.73	0.05
Pos_Neg	Positive and negative integers-background	2.58	0.91	2.82	0.06	2.59	0.06	2.37	0.06
Rat_Numb	Rational numbers-background	3.10	0.86	3.20	0.06	3.23	0.05	2.86	0.06
RatoPrRteP	Ratio_percent_rate_propor-background	2.83	0.84	3.04	0.06	2.92	0.05	2.56	0.05
Wd_Prob	Solving word problems-background	3.26	0.81	3.35	0.06	3.27	0.05	3.18	0.05
variables	Concept of variables-background	2.48	0.80	2.66	0.06	2.49	0.05	2.32	0.05
Mani_Var	Manipulation of variables-background	2.82	0.78	3.06	0.05	2.84	0.05	2.60	0.05
Simp_eq	Solve simple linear equations & inequalities-background	2.80	0.83	2.91	0.06	2.84	0.05	2.64	0.05
PlotGraph	Plotting and graphing-background	2.44	0.93	2.65	0.07	2.48	0.06	2.22	0.06
Geo_Shapes	Formulas for geometric shapes-background	2.81	0.82	2.93	0.06	2.78	0.05	2.76	0.05
StudyHabit	Study skills & work habits-background	3.00	0.87	3.18	0.06	2.99	0.05	2.87	0.05
ComputeSk	Computation skills-background	2.53	0.89	2.69	0.06	2.56	0.05	2.37	0.06
Use_real	Use math in real-world-background	2.94	0.77	2.97	0.06	3.01	0.05	2.84	0.05
Work_Indep	Work independently-background	2.58	0.85	2.78	0.06	2.60	0.06	2.38	0.05
Work_Coop	Working cooperatively-background	2.32	0.77	2.56	0.05	2.36	0.05	2.07	0.04
Textbooks	Textbooks	2.92	1.18	2.48	0.10	2.94	0.07	3.25	0.07
PrintMat	Printed instructional materials	2.60	0.93	2.73	0.07	2.62	0.06	2.46	0.06
TeacherMat	Teacher written materials	2.11	1.17	2.23	0.08	2.21	0.07	1.89	0.08
GrCalculat	Graphing calculators	1.53	1.50	1.45	0.10	1.37	0.09	1.80	0.11
PhyObj	Physical objects-manipulatives	1.26	0.80	1.45	0.07	1.15	0.04	1.22	0.05
Software	Computer-based instructional tools-software	1.00	1.21	1.20	0.10	0.85	0.06	1.02	0.09
Computer_help	Computer-based tools help	3.33	1.32	3.14	0.10	3.41	0.07	3.40	0.09
TextTopic	Appropriate textbook topics	1.77	0.83	1.85	0.06	1.86	0.05	1.59	0.05
TextSeqCon	Appropriate math concept sequences	2.23	1.03	2.46	0.08	2.41	0.06	1.84	0.05
TextExampl	Examples & lessons on concepts	2.09	0.98	2.28	0.08	2.22	0.06	1.77	0.06
TextProbSo	Development of problem-solving skills	2.16	0.98	2.47	0.08	2.21	0.06	1.87	0.05
TextPrac	Practice on topics	2.29	1.14	2.60	0.09	2.29	0.06	2.04	0.07
TextSugges	Textbook suggestions for homework	2.24	1.05	2.55	0.08	2.28	0.06	1.93	0.07
TextSupp	Adequate textbook support materials	2.27	1.10	2.53	0.08	2.35	0.07	1.97	0.07
TextTitle_A	Textbook title	2.01	0.89	2.27	0.07	2.07	0.05	1.74	0.05
TextDivers	Textbook suggestions for diverse learner	2.73	1.10	2.84	0.08	2.94	0.06	2.38	0.07
StudentFail	Number of Target Class student fail	2.55	1.44	3.11	0.12	2.57	0.09	2.08	0.06

Continued on p. 9-40

Table B-1, continued

Variable Name	Variable Label	Overall Sample		Locale					
				Urban		Suburban		Rural	
		Wtd. Mean	Wtd. SD	Wtd. Mean	Wtd. SE	Wtd. Mean	Wtd. SE	Wtd. Mean	Wtd. SE
TimeAssign	Time on assignments	3.10	0.86	3.17	0.07	3.10	0.05	3.05	0.04
ComAssign	Frequency of completes	2.06	0.97	2.42	0.07	2.02	0.06	1.81	0.06
Min_Meet	Average minutes of class time	271.80	84.52	258.70	5.76	280.55	5.58	272.15	4.58
Class_Period	Minutes of class period	62.77	25.23	63.95	1.91	63.09	1.27	61.45	1.75
InsuffComA	Insufficient access to computers	1.86	1.01	1.73	0.07	2.00	0.06	1.79	0.07
InsuffGrCa	Insufficient access to graphing calculators	1.70	0.92	1.72	0.06	1.94	0.06	1.40	0.04
PoorTextBk	Poor quality or out-of-date textbooks	1.59	1.01	1.70	0.07	1.60	0.06	1.50	0.07
LargeClas	Class sizes are too large	1.97	1.04	2.08	0.08	2.22	0.07	1.60	0.05
Insuffcoor	Insufficient access to computers	1.75	0.92	1.96	0.07	1.81	0.06	1.50	0.05
InadTeach	Inadequately prepared teachers	1.41	0.75	1.64	0.06	1.39	0.04	1.25	0.04
LackPlan	Lack of teacher planning time	1.74	0.93	1.97	0.07	1.69	0.05	1.62	0.06
DiffStudnt	Diverse students take same class	2.53	1.05	2.81	0.08	2.47	0.06	2.38	0.06
LittleFamS	Too little parent/family support	2.74	1.03	2.98	0.07	2.73	0.07	2.57	0.06
InadProLng	Inadequate opportunities for professional learning	1.66	0.84	1.87	0.06	1.66	0.05	1.50	0.05
InadAdminS	Inadequate administrative support	1.64	0.91	1.88	0.07	1.63	0.05	1.45	0.05
Class_Wk	Class periods per week	18.71	9.83	18.88	0.67	19.27	0.57	17.93	0.65
Min_Prep	Average minutes for class preparation	63.06	40.88	62.65	2.94	65.07	2.68	61.05	2.27
UnschdPrep	Average Min for unscheduled class prep	61.66	81.22	69.47	4.98	68.50	6.28	47.54	3.26
AvailTutor	Availability of tutoring or other	2.52	1.10	2.31	0.08	2.47	0.07	2.74	0.07
QualTutor	Quality of tutoring or other	2.42	1.05	2.36	0.07	2.37	0.06	2.52	0.07
WholNumIm	Whole number operations-importance	4.65	0.59	4.60	0.05	4.61	0.04	4.74	0.03
PosNegIm	Positive & negative integers-importance	4.77	0.46	4.75	0.03	4.80	0.03	4.76	0.03
RatNumbIm	Rational numbers-importance	4.59	0.59	4.52	0.04	4.56	0.04	4.70	0.03
RatoPrRtePIm	Ratio_percent_rate_propor-importance	4.19	0.78	4.18	0.05	4.08	0.05	4.32	0.05
Wd_ProbIm	Solving word problems-importance	4.51	0.62	4.47	0.05	4.51	0.04	4.54	0.04
variablesIm	Concept of variables-importance	4.61	0.67	4.49	0.06	4.68	0.04	4.63	0.04
Mani_VarIm	Manipulation of variables-importance	4.55	0.75	4.39	0.06	4.60	0.05	4.61	0.04
Simp_eqIm	Solve simple linear equations & inequalities-importance	4.44	0.84	4.26	0.07	4.46	0.05	4.55	0.05
PlotGraphIm	Plotting and graphing-importance	4.35	0.80	4.23	0.06	4.35	0.05	4.44	0.05
Geo_ShapesIm	Formulas for geometric shapes-importance	3.45	0.97	3.47	0.07	3.37	0.06	3.52	0.06
StudyHabitIm	Study skills & work habits-importance	4.72	0.50	4.69	0.04	4.74	0.03	4.71	0.03
ComputeSk_A	Computation skills-importance	4.54	0.65	4.56	0.04	4.50	0.04	4.56	0.04
Use_realIm	Use math in real-world-importance	4.10	0.83	4.12	0.06	4.09	0.05	4.10	0.05
Work_IndepIm	Work independently-importance	4.34	0.71	4.25	0.05	4.35	0.04	4.39	0.04
Work_CoopIm	Working cooperatively-importance	4.02	0.86	4.03	0.06	4.04	0.05	3.98	0.06
AlgbraProf	Expected student algebra proficiency	2.30	0.93	2.39	0.07	2.23	0.05	2.32	0.06
Preservice	Preservice teacher education	2.06	0.89	2.02	0.06	2.09	0.06	2.05	0.06
ProfDev	Professional development	2.05	0.84	2.08	0.07	2.00	0.05	2.09	0.05
ContentStd	Algebra I content	2.29	0.94	2.26	0.06	2.20	0.06	2.43	0.06
AssessOut	Assessments of Algebra I outcomes	2.66	1.01	2.57	0.06	2.57	0.06	2.82	0.07
T_Age	Teacher's age	41.11	11.69	42.33	0.86	41.29	0.69	39.94	0.71
ElemYrs	Elementary years taught	2.07	4.86	1.57	0.41	3.35	0.51	1.06	0.24
SecYrs	Secondary years taught	12.15	9.99	11.83	0.74	11.78	0.59	12.81	0.64
TotalYrs	Total years taught	12.77	10.35	12.16	0.91	13.08	0.71	12.84	0.76
T_YrsSchool	Teacher's years in current school	8.00	8.09	6.86	0.48	8.29	0.48	8.59	0.56
T_YrsExp	Teacher's years of algebra experience	9.49	8.56	8.88	0.57	9.15	0.48	10.38	0.59
T_ColegeYr	Teacher's college graduation year	1993.70	10.97	1993.65	0.77	1993.39	0.68	1994.08	0.67
T_Skill	Teacher's skill	1.33	0.58	1.34	0.04	1.35	0.04	1.31	0.03

Note: SE's are not adjusted for design effect.

Source: Based on responses to the National Opinion Research Center's National Survey of Algebra Teachers, 2007.

Table B-2: Means for Survey Variables by School Minority Concentration

Variable Name	Variable Label	Percent Minority							
		1st Quartile (Low)		2nd Quartile		3rd Quartile		4th Quartile (high)	
		Wtd. Mean	Wtd. SE	Wtd. Mean	Wtd. SE	Wtd. Mean	Wtd. SE	Wtd. Mean	Wtd. SE
TC_Student	Target Class - Total Number of Students	2.52	0.08	3.25	0.09	3.05	0.08	3.16	0.08
TC_Studnt7	Target Class 7th-grade students	0.19	0.06	0.26	0.06	0.20	0.07	0.22	0.08
TC_Studnt8	Target Class 8th-grade students	1.78	0.15	1.83	0.16	1.50	0.15	1.48	0.17
TC_Studnt9	Target Class 9th-grade students	1.87	0.13	1.88	0.14	2.13	0.12	2.10	0.14
TC_Studt10	Target Class 10th-grade students	0.49	0.06	0.77	0.09	0.78	0.06	0.70	0.07
TC_Studt11	Target Class 11th-grade students	0.29	0.04	0.32	0.04	0.33	0.04	0.40	0.06
TC_Studt12	Target Class 12th-grade students	0.13	0.03	0.08	0.03	0.21	0.03	0.30	0.04
TC_StudtSE	Target Class Special ED Students	0.57	0.04	0.56	0.06	0.72	0.06	0.55	0.06
TC_StudtBi	Target Class bilingual Students	0.20	0.06	0.29	0.05	0.38	0.05	0.52	0.08
Come_Time	Come to class on time	3.83	0.03	3.72	0.04	3.51	0.04	3.17	0.06
Attend_Reg	Attend class regularly	3.62	0.04	3.63	0.04	3.42	0.04	3.15	0.05
Come_Prep	Come to class prepared	3.12	0.05	3.25	0.06	2.88	0.06	2.36	0.08
Creat_Prob	Create serious behavior problems	0.47	0.04	0.56	0.06	0.65	0.05	0.80	0.05
Pay_Attn	Regularly pay attention	2.97	0.05	2.95	0.07	2.77	0.06	2.53	0.07
Activ_Part	Actively participate	2.85	0.06	2.80	0.07	2.57	0.06	2.52	0.07
Take_Note	Take notes	2.85	0.07	2.75	0.08	2.68	0.06	2.58	0.09
Diff_ReadE	Serious difficulties reading English	0.38	0.04	0.32	0.05	0.49	0.04	0.75	0.06
Care_grade	Care about what grade received	3.02	0.05	3.10	0.06	2.92	0.06	2.51	0.08
Whole_Numb	Whole number-background	1.80	0.05	1.69	0.06	1.95	0.06	1.98	0.06
Pos_Neg	Positive and negative integers-background	2.39	0.05	2.45	0.07	2.59	0.07	2.91	0.07
Rat_Numb	Rational numbers-background	2.93	0.06	3.01	0.07	3.25	0.06	3.20	0.07
RatoPrRteP	Ratio_percent_rate_propor-background	2.66	0.06	2.63	0.07	2.91	0.06	3.17	0.05
Wd_Prob	Solving word problems-background	3.17	0.06	3.10	0.07	3.30	0.06	3.49	0.06
variables	Concept of variables-background	2.44	0.05	2.31	0.07	2.46	0.06	2.75	0.06
Mani_Var	Manipulation of variables-background	2.68	0.05	2.66	0.07	2.88	0.05	3.07	0.06
Simp_eq	Solve simple linear equations & inequalities-background	2.75	0.05	2.64	0.07	2.85	0.06	2.94	0.07
PlotGraph	Plotting and graphing-background	2.29	0.06	2.32	0.07	2.48	0.07	2.69	0.07
Geo_Shapes	Formulas for geometric shapes-background	2.79	0.05	2.51	0.07	2.89	0.06	3.06	0.06
StudyHabit	Study skills & work habits-background	2.80	0.05	2.81	0.08	3.13	0.06	3.27	0.06
ComputeSk	Computation skills-background	2.51	0.06	2.37	0.07	2.53	0.07	2.73	0.06
Use_real	Use math in real world-background	2.77	0.05	2.88	0.06	3.06	0.06	3.06	0.06
Work_Indep	Work independently-background	2.38	0.05	2.46	0.08	2.64	0.06	2.86	0.06
Work_Coop	Working cooperatively-background	2.14	0.04	2.25	0.07	2.34	0.06	2.57	0.05
Textbooks	Textbooks	3.16	0.07	3.16	0.08	2.85	0.09	2.43	0.10
PrintMat	Printed instructional materials	2.48	0.06	2.46	0.08	2.67	0.07	2.80	0.07
TeacherMat	Teacher written materials	1.91	0.08	2.10	0.09	2.15	0.09	2.30	0.08
GrCalculat	Graphing calculators	1.61	0.11	1.53	0.11	1.41	0.11	1.60	0.12
PhyObj	Physical objects-manipulatives	1.23	0.06	1.15	0.06	1.13	0.05	1.57	0.07
Software	Computer-based instructional tools-software	1.03	0.09	0.99	0.08	0.70	0.07	1.39	0.12
Computer_help	Computer-based tools help	3.53	0.10	3.33	0.10	3.44	0.09	2.94	0.11
TextTopic	Appropriate textbook topics	1.63	0.05	1.72	0.07	1.86	0.06	1.87	0.06
TextSeqCon	Appropriate math concept sequences	1.90	0.06	2.16	0.08	2.30	0.07	2.66	0.09
TextExampl	Examples & lessons on concepts	1.83	0.06	2.17	0.09	2.10	0.07	2.33	0.08
TextProbSo	Development of problem-solving skills	1.86	0.06	1.99	0.07	2.21	0.07	2.71	0.09
TextPrac	Practice on topics	2.14	0.08	2.34	0.09	2.22	0.08	2.52	0.09
TextSugges	Textbook suggestions for homework	2.05	0.08	2.26	0.08	2.12	0.07	2.62	0.09
TextSupp	Adequate textbook support materials	2.02	0.08	2.36	0.09	2.18	0.07	2.65	0.09
TextTitle_A	Textbook title	1.73	0.05	1.93	0.07	2.10	0.06	2.36	0.07
TextDivers	Textbook suggestions for diverse learner	2.57	0.09	2.88	0.09	2.59	0.07	2.98	0.09

Continued on p. 9-42

Table B-2, continued

Variable Name	Variable Label	Percent Minority							
		1st Quartile (Low)		2nd Quartile		3rd Quartile		4th Quartile (high)	
		Wtd. Mean	Wtd. SE	Wtd. Mean	Wtd. SE	Wtd. Mean	Wtd. SE	Wtd. Mean	Wtd. SE
StudentFail	Number of Target Class student fail	2.10	0.07	2.10	0.09	2.81	0.11	3.26	0.13
TimeAssign	Time on assignments	3.08	0.05	3.11	0.06	3.16	0.06	3.04	0.08
ComAssign	Frequency of completes	1.79	0.06	1.74	0.07	2.17	0.06	2.59	0.09
Min_Meet	Average Minutes of class time	264.29	4.48	259.75	5.84	278.83	6.49	284.99	7.91
Class_Period	Minutes of class period	60.42	2.09	58.20	1.33	68.25	1.91	63.37	1.76
InsuffComA	Insufficient access to computers	1.93	0.08	1.72	0.07	1.85	0.07	1.92	0.08
InsuffGrCa	Insufficient access to graphing calculators	1.56	0.06	1.60	0.07	1.64	0.06	2.04	0.08
PoorTextBk	Poor quality or out-of-date textbooks	1.47	0.07	1.34	0.06	1.60	0.07	2.02	0.09
LargeClas	Class sizes are too large	1.68	0.06	1.91	0.08	2.12	0.08	2.21	0.09
Insuffcoor	Insufficient access to computers	1.52	0.06	1.53	0.05	1.82	0.07	2.17	0.08
InadTeach	Inadequately prepared teachers	1.22	0.04	1.26	0.05	1.49	0.05	1.69	0.08
LackPlan	Lack of teacher planning time	1.49	0.05	1.61	0.06	1.86	0.08	2.05	0.07
DiffStudnt	Diverse students take same class	2.42	0.07	2.23	0.08	2.61	0.07	2.90	0.08
LittleFamS	Too little parent/family support	2.60	0.07	2.31	0.08	2.82	0.07	3.30	0.07
InadProLng	Inadequate opportunities for professional learning	1.53	0.05	1.45	0.05	1.74	0.06	1.95	0.08
InadAdminS	Inadequate administrative support	1.40	0.05	1.42	0.06	1.74	0.07	2.04	0.09
Class_Wk	Class periods per week	18.00	0.73	19.75	0.73	19.04	0.65	18.01	0.81
Min_Prep	Average minutes for class preparation	63.31	2.56	60.90	2.90	63.64	2.82	64.35	4.10
UnschdPrep	Average Min for unscheduled class prep	49.69	3.66	59.32	6.81	57.45	6.40	85.31	6.91
AvailTutor	Availability of tutoring or other	2.85	0.08	2.40	0.08	2.55	0.08	2.19	0.08
QualTutor	Quality of tutoring or other	2.62	0.08	2.24	0.08	2.50	0.08	2.27	0.08
WholNumIm	Whole number operations-importance	4.64	0.04	4.66	0.05	4.71	0.04	4.57	0.05
PosNegIm	Positive & negative integers-importance	4.72	0.03	4.81	0.04	4.85	0.03	4.69	0.04
RatNumbIm	Rational numbers-importance	4.65	0.04	4.60	0.05	4.62	0.04	4.48	0.05
RatoPrRtePIm	Ratio_percent_rate_propor-importance	4.26	0.06	3.97	0.07	4.29	0.05	4.21	0.06
Wd_ProbIm	Solving word problems-importance	4.50	0.04	4.50	0.05	4.57	0.04	4.46	0.05
variablesIm	Concept of variables-importance	4.55	0.05	4.69	0.05	4.67	0.05	4.50	0.06
Mani_VarIm	Manipulation of variables-importance	4.48	0.05	4.68	0.06	4.60	0.06	4.43	0.06
Simp_eqIm	Solve simple linear equations & inequalities-importance	4.41	0.06	4.55	0.06	4.47	0.06	4.29	0.07
PlotGraphIm	Plotting and graphing-importance	4.31	0.06	4.37	0.06	4.41	0.06	4.27	0.07
Geo_ShapesIm	Formulas for geometric shapes-importance	3.44	0.07	3.17	0.07	3.55	0.07	3.64	0.07
StudyHabitIm	Study skills & work habits-importance	4.71	0.04	4.69	0.04	4.78	0.03	4.67	0.04
ComputeSk_A	Computation skills-importance	4.52	0.05	4.50	0.05	4.60	0.05	4.50	0.05
Use_realIm	Use math in real world-importance	4.02	0.05	4.06	0.07	4.08	0.06	4.30	0.06
Work_IndepIm	Work independently-importance	4.33	0.05	4.44	0.05	4.29	0.05	4.28	0.06
Work_CoopIm	Working cooperatively-importance	3.95	0.06	4.03	0.07	3.96	0.07	4.17	0.06
AlgbraProf	Expected student algebra proficiency	2.31	0.06	2.20	0.06	2.31	0.06	2.39	0.08
Preservice	Preservice teacher education	2.19	0.06	2.08	0.08	1.94	0.06	2.02	0.06
ProfDev	Professional development	2.15	0.06	1.99	0.06	2.03	0.06	2.03	0.07
ContentStd	Algebra I content	2.41	0.06	2.24	0.08	2.21	0.06	2.30	0.08
AssessOut	Assessments of algebra I outcomes	2.82	0.07	2.61	0.08	2.65	0.07	2.51	0.07
T_Age	Teacher’s age	40.62	0.83	41.22	0.90	40.50	0.75	42.43	1.03
ElemYrs	Elementary years taught	1.27	0.36	3.70	0.69	1.26	0.31	2.22	0.51
SecYrs	Secondary years taught	13.59	0.74	11.66	0.73	12.86	0.73	9.89	0.76
TotalYrs	Total years taught	13.51	0.94	13.71	0.89	12.57	0.78	10.90	1.03
T_YrsSchool	Teacher’s years in current school	9.40	0.65	8.75	0.63	7.06	0.50	6.68	0.58
T_YrsExp	Teacher’s years of algebra experience	10.96	0.68	8.82	0.56	9.86	0.58	7.88	0.66
T_ColegeYr	Teacher’s college graduation year	1993.99	0.82	1994.21	0.84	1993.54	0.69	1992.99	0.94
T_Skill	Teacher’s skill	1.38	0.04	1.26	0.04	1.29	0.04	1.43	0.05

Note: SE’s are not adjusted for design effect.

Source: Based on responses to the National Opinion Research Center’s National Survey of Algebra Teachers, 2007.

Table B-3: Means for Survey Variables by Percentage of Students in the School Eligible for Free and Reduced-Price Lunch

Variable Name	Variable Label	Percent Free/Reduced-Price Lunch							
		1st Quartile (Low)		2nd Quartile		3rd Quartile		4th Quartile (high)	
		Wtd. Mean	Wtd. SE	Wtd. Mean	Wtd. SE	Wtd. Mean	Wtd. SE	Wtd. Mean	Wtd. SE
TC_Student	Target Class - total number of students	3.13	0.09	2.83	0.08	2.89	0.08	3.15	0.09
TC_Studnt7	Target Class 7th-grade students	0.23	0.06	0.30	0.08	0.17	0.06	0.12	0.06
TC_Studnt8	Target Class 8th-grade students	1.74	0.15	1.12	0.14	1.70	0.15	2.36	0.19
TC_Studnt9	Target Class 9th-grade students	1.96	0.14	2.31	0.12	1.86	0.13	1.73	0.15
TC_Studt10	Target Class 10th-grade students	0.68	0.07	0.74	0.07	0.71	0.07	0.50	0.06
TC_Studt11	Target Class 11th-grade students	0.31	0.04	0.37	0.04	0.36	0.04	0.22	0.05
TC_Studt12	Target Class 12th-grade students	0.13	0.03	0.18	0.03	0.21	0.03	0.17	0.04
TC_StudtSE	Target Class Special ED students	0.55	0.05	0.66	0.05	0.68	0.06	0.46	0.06
TC_StudtBi	Target Class bilingual Students	0.30	0.06	0.36	0.07	0.29	0.04	0.41	0.08
Come_Time	Come to class on time	3.71	0.04	3.66	0.04	3.53	0.05	3.31	0.06
Attend_Reg	Attend class regularly	3.56	0.04	3.53	0.04	3.40	0.05	3.34	0.06
Come_Prep	Come to class prepared	3.08	0.06	2.96	0.06	2.79	0.07	2.85	0.08
Creat_Prob	Create serious behavior problems	0.59	0.05	0.60	0.04	0.62	0.05	0.63	0.06
Pay_Attn	Regularly pay attention	2.82	0.06	2.88	0.05	2.85	0.06	2.67	0.08
Activ_Part	Actively participate	2.68	0.06	2.77	0.07	2.59	0.07	2.71	0.07
Take_Note	Take notes	2.71	0.08	2.67	0.07	2.81	0.07	2.73	0.09
Diff_ReadE	Serious difficulties reading English	0.44	0.05	0.35	0.04	0.51	0.04	0.66	0.06
Care_grade	Care about what grade received	3.00	0.06	3.02	0.06	2.85	0.07	2.68	0.08
Whole_Numb	Whole number-background	1.76	0.06	1.86	0.06	1.96	0.07	1.84	0.06
Pos_Neg	Positive and negative integers-background	2.43	0.06	2.66	0.06	2.68	0.07	2.50	0.08
Rat_Numb	Rational numbers-background	3.06	0.06	3.11	0.06	3.26	0.06	2.91	0.08
RatoPrRteP	Ratio_percent_rate_propor-background	2.75	0.06	2.85	0.06	2.92	0.06	2.80	0.08
Wd_Prob	Solving word problems-background	3.24	0.06	3.27	0.06	3.35	0.05	3.12	0.08
variables	Concept of variables-background	2.37	0.06	2.42	0.06	2.62	0.06	2.55	0.06
Mani_Var	Manipulation of variables-background	2.66	0.06	2.85	0.05	2.95	0.06	2.82	0.07
Simp_eq	Solve simple linear equations & inequalities-background	2.69	0.06	2.76	0.06	2.90	0.06	2.84	0.07
PlotGraph	Plotting and graphing-background	2.41	0.07	2.50	0.06	2.47	0.07	2.33	0.08
Geo_Shapes	Formulas for geometric shapes-background	2.64	0.06	2.85	0.06	2.93	0.06	2.83	0.07
StudyHabit	Study skills & work habits-background	2.88	0.06	2.89	0.07	3.15	0.06	3.10	0.07
ComputeSk	Computation skills-background	2.58	0.06	2.41	0.07	2.66	0.07	2.48	0.06
Use_real	Use math in real world-background	2.92	0.06	2.93	0.05	3.07	0.05	2.80	0.07
Work_Indep	Work independently-background	2.56	0.06	2.45	0.07	2.70	0.06	2.63	0.07
Work_Coop	Working cooperatively-background	2.28	0.05	2.25	0.06	2.30	0.06	2.48	0.06
Textbooks	Textbooks	3.24	0.07	2.89	0.09	2.61	0.10	2.95	0.08
PrintMat	Printed instructional materials	2.45	0.07	2.45	0.07	2.83	0.06	2.72	0.07
TeacherMat	Teacher written materials	2.13	0.09	2.10	0.09	2.24	0.09	1.87	0.09
GrCalculat	Graphing calculators	1.42	0.10	1.27	0.10	1.52	0.11	2.02	0.14
PhyObj	Physical objects-manipulatives	1.07	0.05	1.21	0.05	1.21	0.06	1.65	0.08
Software	Computer-based instructional tools-software	0.96	0.08	0.87	0.07	0.77	0.07	1.56	0.14
Computer_help	Computer-based tools help	3.35	0.10	3.42	0.09	3.52	0.09	2.96	0.12
TextTopic	Appropriate textbook topics	1.84	0.07	1.69	0.06	1.90	0.06	1.62	0.06
TextSeqCon	Appropriate math concept sequences	2.23	0.08	2.08	0.07	2.38	0.07	2.29	0.09
TextExampl	Examples & lessons on concepts	2.14	0.08	1.99	0.07	2.24	0.08	1.99	0.07
TextProbSo	Development of problem-solving skills	2.04	0.07	1.97	0.06	2.35	0.08	2.37	0.09
TextPrac	Practice on topics	2.51	0.09	2.07	0.08	2.31	0.08	2.30	0.09
TextSugges	Textbook suggestions for homework	2.34	0.08	2.09	0.07	2.28	0.08	2.25	0.09
TextSupp	Adequate textbook support materials	2.36	0.09	2.12	0.08	2.28	0.07	2.36	0.10
TextTitle_A	Textbook title	1.99	0.07	1.94	0.07	2.15	0.06	1.99	0.07
TextDivers	Textbook suggestions for diverse learner	2.88	0.08	2.58	0.09	2.81	0.07	2.66	0.09

Continued on p. 9-44

Table B-3, continued

Variable Name	Variable Label	Percent Free/Reduced-Price Lunch							
		1st Quartile (Low)		2nd Quartile		3rd Quartile		4th Quartile (high)	
		Wtd. Mean	Wtd. SE	Wtd. Mean	Wtd. SE	Wtd. Mean	Wtd. SE	Wtd. Mean	Wtd. SE
StudentFail	Number of Target Class student fail	2.22	0.09	2.36	0.10	2.75	0.11	2.98	0.13
TimeAssign	Time on assignments	3.18	0.06	3.20	0.05	2.95	0.07	3.08	0.08
ComAssign	Frequency of completes	1.83	0.06	1.97	0.07	2.24	0.08	2.23	0.08
Min_Meet	Average Minutes of class time	265.71	5.44	262.28	5.30	275.06	6.99	291.62	7.26
Class_Period	Minutes of class period	63.77	2.48	61.44	1.56	63.90	1.52	62.39	1.73
InsuffComA	Insufficient access to computers	1.75	0.07	1.73	0.06	1.65	0.07	2.48	0.10
InsuffGrCa	Insufficient access to graphing calculators	1.75	0.07	1.63	0.06	1.68	0.07	1.78	0.08
PoorTextBk	Poor quality or out-of-date textbooks	1.36	0.06	1.41	0.06	1.75	0.09	1.98	0.09
LargeClas	Class sizes are too large	2.04	0.07	1.86	0.07	1.94	0.08	2.12	0.08
Insuffcoor	Insufficient access to computers	1.59	0.06	1.72	0.07	1.71	0.07	2.04	0.08
InadTeach	Inadequately prepared teachers	1.35	0.05	1.35	0.05	1.43	0.05	1.54	0.08
LackPlan	Lack of teacher planning time	1.64	0.07	1.77	0.06	1.80	0.07	1.76	0.07
DiffStudnt	Diverse students take same class	2.36	0.08	2.49	0.07	2.59	0.08	2.76	0.08
LittleFamS	Too little parent/family support	2.35	0.07	2.69	0.07	2.88	0.07	3.18	0.08
InadProLng	Inadequate opportunities for professional learning	1.58	0.06	1.67	0.06	1.51	0.05	1.97	0.08
InadAdminS	Inadequate administrative support	1.54	0.06	1.59	0.06	1.67	0.07	1.81	0.08
Class_Wk	Class periods per week	19.97	0.68	18.63	0.68	18.94	0.68	16.71	0.91
Min_Prep	Average minutes for class preparation	65.01	2.96	62.16	2.99	68.41	3.05	54.67	3.13
UnschdPrep	Average Min for unscheduled class prep	61.46	6.35	63.76	4.22	63.81	8.33	56.94	3.93
AvailTutor	Availability of tutoring or other	2.54	0.08	2.48	0.07	2.33	0.08	2.84	0.09
QualTutor	Quality of tutoring or other	2.36	0.08	2.36	0.07	2.30	0.08	2.78	0.08
WholNumIm	Whole number operations-importance	4.58	0.05	4.73	0.03	4.73	0.04	4.54	0.06
PosNegIm	Positive & negative integers-importance	4.81	0.03	4.81	0.03	4.78	0.03	4.66	0.04
RatNumbIm	Rational numbers-importance	4.60	0.05	4.61	0.04	4.62	0.04	4.55	0.05
RatoPrRtePIm	Ratio_percent_rate_propor-importance	4.16	0.07	4.11	0.06	4.32	0.05	4.18	0.05
Wd_ProbIm	Solving word problems-importance	4.49	0.05	4.53	0.04	4.55	0.05	4.47	0.05
variablesIm	Concept of variables-importance	4.66	0.05	4.60	0.05	4.59	0.05	4.58	0.05
Mani_VarIm	Manipulation of variables-importance	4.62	0.06	4.55	0.05	4.49	0.06	4.55	0.05
Simp_eqIm	Solve simple linear equations & inequalities-importance	4.44	0.06	4.48	0.06	4.36	0.06	4.46	0.07
PlotGraphIm	Plotting and graphing-importance	4.41	0.06	4.34	0.06	4.31	0.06	4.31	0.07
Geo_ShapesIm	Formulas for geometric shapes-importance	3.35	0.07	3.44	0.07	3.36	0.07	3.73	0.08
StudyHabitIm	Study skills & work habits-importance	4.68	0.04	4.77	0.03	4.73	0.04	4.69	0.04
ComputeSk_A	Computation skills-importance	4.50	0.05	4.58	0.05	4.49	0.05	4.59	0.05
Use_realIm	Use math in real world-importance	4.11	0.06	4.03	0.06	4.00	0.07	4.34	0.06
Work_IndepIm	Work independently-importance	4.37	0.05	4.41	0.05	4.25	0.06	4.29	0.06
Work_CoopIm	Working cooperatively-importance	3.97	0.07	4.01	0.06	3.89	0.07	4.27	0.05
AlgbraProf	Expected student algebra proficiency	2.20	0.06	2.25	0.06	2.46	0.08	2.32	0.07
Preservice	Preservice teacher education	2.10	0.07	2.05	0.07	2.00	0.06	2.08	0.07
ProfDev	Professional development	2.08	0.06	2.00	0.06	2.03	0.06	2.11	0.08
ContentStd	Algebra I content	2.23	0.07	2.41	0.07	2.37	0.07	2.12	0.07
AssessOut	Assessments of Algebra I outcomes	2.67	0.08	2.74	0.08	2.76	0.07	2.39	0.07
T_Age	Teacher's age	41.55	0.82	41.08	0.78	40.34	0.85	41.84	1.08
ElemYrs	Elementary years taught	3.92	0.73	1.04	0.26	1.72	0.35	2.05	0.63
SecYrs	Secondary years taught	12.25	0.75	12.95	0.70	12.17	0.74	10.91	0.85
TotalYrs	Total years taught	13.31	0.85	14.37	0.82	11.83	0.84	10.00	1.23
T_YrsSchool	Teacher's years in current school	8.81	0.56	8.37	0.58	7.08	0.58	7.72	0.69
T_YrsExp	Teacher's years of algebra experience	9.69	0.63	10.53	0.63	9.08	0.58	8.27	0.68
T_ColegeYr	Teacher's college graduation year	1994.34	0.84	1993.72	0.75	1993.72	0.75	1992.61	0.96
T_Skill	Teacher's skill	1.35	0.04	1.27	0.03	1.32	0.04	1.43	0.06

Note: SE's are not adjusted for design effect.

Source: Based on responses to the National Opinion Research Center's National Survey of Algebra Teachers, 2007.

Table B-4: Means for Survey Variables by Grade Level of the Target Class and School Grade Level

Variable Name	Variable Label	Target Class Grade						School Grade (High School vs. Others)			
		7th & 8th Grade		9thGrade		10th, 11th, & 12th Grade		High School		Others	
		Wtd. Mean	Wtd. SE	Wtd. Mean	Wtd. SE	Wtd. Mean	Wtd. SE	Wtd. Mean	Wtd. SE	Wtd. Mean	Wtd. SE
TC_Student	Target Class - total number of students	3.35	0.07	2.78	0.06	2.55	0.12	--	--	--	--
TC_Studnt7	Target Class 7th-grade students	0.50	0.07	0.00	0.00	0.00	0.00	--	--	--	--
TC_Studnt8	Target Class 8th-grade students	3.61	0.06	0.01	0.01	0.11	0.05	--	--	--	--
TC_Studnt9	Target Class 9th-grade students	0.05	0.02	3.31	0.04	0.42	0.09	--	--	--	--
TC_Studt10	Target Class 10th-grade students	0.00	0.00	0.82	0.03	2.83	0.13	--	--	--	--
TC_Studt11	Target Class 11th-grade students	0.00	0.00	0.46	0.03	1.07	0.11	--	--	--	--
TC_Studt12	Target Class 12th-grade students	0.00	0.00	0.27	0.03	0.48	0.09	--	--	--	--
TC_StudtSE	Target Class Special ED students	0.38	0.04	0.75	0.04	0.72	0.10	--	--	--	--
TC_StudtBi	Target Class bilingual students	0.29	0.05	0.36	0.04	0.47	0.15	--	--	--	--
Come_Time	Come to class on time	3.75	0.03	3.49	0.03	3.30	0.10	--	--	--	--
Attend_Reg	Attend class regularly	3.74	0.03	3.32	0.03	3.15	0.11	--	--	--	--
Come_Prep	Come to class prepared	3.32	0.04	2.71	0.05	2.49	0.12	--	--	--	--
Creat_Prob	Create serious behavior problems	0.46	0.04	0.70	0.03	0.74	0.09	--	--	--	--
Pay_Attn	Regularly pay attention	3.11	0.05	2.66	0.04	2.50	0.10	--	--	--	--
Activ_Part	Actively participate	2.92	0.05	2.60	0.04	2.17	0.11	--	--	--	--
Take_Note	Take notes	3.05	0.05	2.58	0.05	2.08	0.13	--	--	--	--
Diff_ReadE	Serious difficulties reading English	0.25	0.03	0.59	0.03	0.77	0.11	--	--	--	--
Care_grade	Care about what grade received	3.28	0.04	2.71	0.05	2.33	0.10	--	--	--	--
Whole_Numb	Whole number-background	1.49	0.04	2.07	0.04	2.19	0.10	--	--	--	--
Pos_Neg	Positive and negative integers-background	2.11	0.05	2.88	0.04	2.82	0.12	--	--	--	--
Rat_Numb	Rational numbers-background	2.64	0.05	3.37	0.04	3.46	0.10	--	--	--	--
RatoPrRteP	Ratio_percent_rate_propor-background	2.49	0.05	3.03	0.04	3.13	0.11	--	--	--	--
Wd_Prob	Solving word problems-background	2.75	0.05	3.57	0.03	3.55	0.09	--	--	--	--
variables	Concept of variables-background	2.17	0.05	2.68	0.04	2.65	0.11	--	--	--	--
Mani_Var	Manipulation of variables-background	2.52	0.04	3.03	0.04	2.85	0.10	--	--	--	--
Simp_eq	Solve simple linear equations & inequalities-background	2.60	0.05	2.92	0.04	2.91	0.12	--	--	--	--
PlotGraph	Plotting and graphing-background	2.03	0.05	2.67	0.04	2.86	0.12	--	--	--	--
Geo_Shapes	Formulas for geometric shapes-background	2.52	0.05	2.98	0.04	3.18	0.08	--	--	--	--
StudyHabit	Study skills & work habits-background	2.56	0.06	3.24	0.04	3.46	0.08	--	--	--	--
ComputeSk	Computation skills-background	2.11	0.05	2.76	0.04	3.05	0.10	--	--	--	--
Use_real	Use math in real world-background	2.46	0.04	3.21	0.03	3.45	0.08	--	--	--	--
Work_Indep	Work independently-background	2.22	0.05	2.75	0.04	3.12	0.10	--	--	--	--
Work_Coop	Working cooperatively-background	2.05	0.05	2.43	0.04	2.82	0.10	--	--	--	--
Textbooks	Textbooks	3.03	0.06	2.81	0.07	3.12	0.13	--	--	--	--
PrintMat	Printed instructional materials	2.49	0.05	2.66	0.05	2.79	0.11	--	--	--	--
TeacherMat	Teacher written materials	1.98	0.07	2.19	0.06	2.19	0.16	--	--	--	--
GrCalculat	Graphing calculators	1.58	0.08	1.49	0.08	1.64	0.22	--	--	--	--
PhyObj	Physical objects-manipulatives	1.23	0.05	1.28	0.04	1.26	0.11	--	--	--	--
Software	Computer-based instructional tools-software	1.06	0.07	1.00	0.06	0.77	0.15	--	--	--	--
Computer_help	Computer-based tools help	3.36	0.08	3.30	0.07	3.45	0.18	--	--	--	--
TextTopic	Appropriate textbook topics	1.67	0.05	1.83	0.05	1.83	0.12	--	--	--	--
TextSeqCon	Appropriate math concept sequences	2.07	0.06	2.36	0.06	2.22	0.14	--	--	--	--
TextExampl	Examples & lessons on concepts	2.02	0.06	2.14	0.05	2.09	0.15	--	--	--	--
TextProbSo	Development of problem-solving skills	1.99	0.05	2.30	0.06	2.11	0.12	--	--	--	--
TextPrac	Practice on topics	2.23	0.06	2.32	0.06	2.35	0.18	--	--	--	--
TextSugges	Textbook suggestions for homework	2.10	0.06	2.32	0.06	2.30	0.16	--	--	--	--
TextSupp	Adequate textbook support materials	2.15	0.06	2.34	0.06	2.43	0.17	--	--	--	--
TextTitle_A	Textbook title	1.89	0.05	2.09	0.05	2.12	0.13	--	--	--	--
TextDivers	Textbook suggestions for diverse learner	2.60	0.06	2.78	0.06	3.02	0.14	--	--	--	--

Continued on p. 9-46

Table B-4, continued

Variable Name	Variable Label	Target Class Grade						School Grade (High School vs. Others)			
		7th & 8th Grade		9thGrade		10th, 11th, & 12th Grade		High School		Others	
		Wtd. Mean	Wtd. SE	Wtd. Mean	Wtd. SE	Wtd. Mean	Wtd. SE	Wtd. Mean	Wtd. SE	Wtd. Mean	Wtd. SE
StudentFail	Number of Target Class student fail	1.81	0.06	2.92	0.08	3.69	0.22	--	--	--	--
TimeAssign	Time on assignments	3.34	0.04	2.97	0.05	2.87	0.13	--	--	--	--
ComAssign	Frequency of completes	1.63	0.04	2.28	0.05	2.65	0.13	--	--	--	--
Min_Meet	Average Minutes of class time	263.72	4.64	277.61	4.47	271.19	10.55	--	--	--	--
Class_Period	Minutes of class period	58.53	0.96	64.93	1.47	68.96	3.51	--	--	--	--
InsuffComA	Insufficient access to computers	--	--	--	--	--	--	1.83	0.05	1.89	0.05
InsuffGrCa	Insufficient access to graphing calculators	--	--	--	--	--	--	1.74	0.05	1.65	0.05
PoorTextBk	Poor quality or out-of-date textbooks	--	--	--	--	--	--	1.64	0.05	1.54	0.05
LargeClas	Class sizes are too large	--	--	--	--	--	--	2.07	0.05	1.87	0.06
Insuffcoor	Insufficient access to computers	--	--	--	--	--	--	1.82	0.05	1.67	0.05
InadTeach	Inadequately prepared teachers	--	--	--	--	--	--	1.44	0.04	1.38	0.04
LackPlan	Lack of teacher planning time	--	--	--	--	--	--	1.76	0.05	1.72	0.05
DiffStudnt	Diverse students take same class	--	--	--	--	--	--	2.68	0.05	2.38	0.06
LittleFamS	Too little parent/family support	--	--	--	--	--	--	3.08	0.05	2.40	0.05
InadProLng	Inadequate opportunities for professional learning	--	--	--	--	--	--	1.76	0.05	1.56	0.04
InadAdminS	Inadequate administrative support	--	--	--	--	--	--	1.74	0.05	1.53	0.05
Class_Wk	Class periods per week	--	--	--	--	--	--	19.46	0.49	17.94	0.53
Min_Prep	Average minutes for class preparation	--	--	--	--	--	--	70.60	2.42	55.42	1.76
UnschdPrep	Average Min for Unscheduled class prep	--	--	--	--	--	--	66.57	5.16	56.71	3.15
AvailTutor	Availability of tutoring or other	--	--	--	--	--	--	2.30	0.06	2.75	0.06
QualTutor	Quality of tutoring or other	--	--	--	--	--	--	2.27	0.05	2.58	0.06
WholNumIm	Whole number operations-importance	--	--	--	--	--	--	4.61	0.03	4.69	0.03
PosNegIm	Positive & negative integers-importance	--	--	--	--	--	--	4.71	0.03	4.83	0.02
RatNumbIm	Rational numbers-importance	--	--	--	--	--	--	4.48	0.03	4.71	0.03
RatoPrRtePIm	Ratio_percent_rate_propor-importance	--	--	--	--	--	--	4.14	0.04	4.24	0.04
Wd_ProbIm	Solving word problems-importance	--	--	--	--	--	--	4.43	0.04	4.59	0.03
variablesIm	Concept of variables-importance	--	--	--	--	--	--	4.55	0.04	4.67	0.03
Mani_VarIm	Manipulation of variables-importance	--	--	--	--	--	--	4.48	0.04	4.61	0.04
Simp_eqIm	Solve simple linear equations & inequalities-importance	--	--	--	--	--	--	4.44	0.05	4.44	0.04
PlotGraphIm	Plotting and graphing-importance	--	--	--	--	--	--	4.29	0.04	4.40	0.04
Geo_ShapesIm	Formulas for geometric shapes-importance	--	--	--	--	--	--	3.32	0.05	3.59	0.05
StudyHabitIm	Study skills & work habits-importance	--	--	--	--	--	--	4.69	0.03	4.75	0.03
ComputeSk_A	Computation skills-importance	--	--	--	--	--	--	4.45	0.04	4.63	0.03
Use_realIm	Use math in real world-importance	--	--	--	--	--	--	3.99	0.05	4.21	0.04
Work_IndepIm	Work independently-importance	--	--	--	--	--	--	4.25	0.04	4.42	0.04
Work_CoopIm	Working cooperatively-importance	--	--	--	--	--	--	3.96	0.05	4.08	0.05
AlgbraProf	Expected student algebra proficiency	--	--	--	--	--	--	2.21	0.05	2.39	0.05
Preservice	Preservice teacher education	--	--	--	--	--	--	1.97	0.04	2.15	0.05
ProfDev	Professional development	--	--	--	--	--	--	2.04	0.04	2.06	0.04
ContentStd	Algebra I content	--	--	--	--	--	--	2.13	0.04	2.45	0.06
AssessOut	Assessments of Algebra I outcomes	--	--	--	--	--	--	2.59	0.05	2.73	0.06
T_Age	Teacher’s age	--	--	--	--	--	--	40.80	0.62	41.42	0.60
ElemYrs	Elementary years taught	--	--	--	--	--	--	1.14	0.24	2.89	0.40
SecYrs	Secondary years taught	--	--	--	--	--	--	11.58	0.53	12.71	0.53
TotalYrs	Total years taught	--	--	--	--	--	--	12.41	0.62	13.18	0.65
T_YrsSchool	Teacher’s years in current school	--	--	--	--	--	--	7.26	0.39	8.76	0.45
T_YrsExp	Teacher’s years of algebra experience	--	--	--	--	--	--	9.83	0.45	9.15	0.45
T_ColegeYr	Teacher’s college graduation year	--	--	--	--	--	--	1994.44	0.56	1992.93	0.59
T_Skill	Teacher’s skill	--	--	--	--	--	--	1.37	0.03	1.29	0.03

Note: SE’s are not adjusted for design effect.
-- : No data available

Source: Based on responses to the National Opinion Research Center’s National Survey of Algebra Teachers, 2007.

APPENDIX C: Variables Used in the Regression Equations And Tables of Regression Estimates

This appendix contains results of ordinary least squares regression analysis for the main outcome variables described in the report. The dependent variables used in the regressions were all transformed to standardized z-scores, such that the estimated effects of the independent variables refer to standard deviation units of the dependent variable. Sample weights were used to weight the observations, and the standard errors of the estimates were adjusted for design effects.

The regressions referred to in the report use a common set of predictor or independent variables. These are defined as follows:

- *Type of locale*: the standard three-level indicator of urban, suburban, or rural school location. This was dichotomized into two variables, one indicating an urban school and another indicating a rural school; each references the difference between those schools and suburban schools.

- *Percentage of students receiving free or reduced-price lunch*: a dichotomous indicator of "40 percent or lower" versus "more than 40 percent" was used as a stratifying variable in the sample design. The analysis was structured to capture more linear effects by using quartile indicators. Dichotomous variables were created to indicate in which quartile (of the sample) of students receiving free or reduced-priced lunch a school was located. The sample was divided into the following quartiles based on the following cut points:

First Quartile (low)	**0.102**
Second Quartile	0.274
Third Quartile	0.478
Forth Quartile (high)	0.809

With dummy variables indicating membership in the second, third, or fourth quartile (referenced to the first quartile, low number of students receiving free or reduced-price lunch).

- *Percentage of students who are black or Hispanic*: a similar dichotomous indicator of "40 percent or lower" versus "more than 40 percent" was used as a sample stratification variable. For the regression analysis, the percentile range was recoded into quartiles and separate dummy variables for the second, third, and fourth quartiles were used (the first quartile was the reference group) based on the following cut points:

First Quartile (low)	**0.028**
Second Quartile	0.099
Third Quartile	0.401
Forth Quartile (high)	0.816

- *School size*. The percentile distribution of school enrollment size was recoded into quartiles, and dummy variables defined based on these cut points:

First Quartile (low)	213
Second Quartile	**436**
Third Quartile	725
Forth Quartile (high)	1681

Note, however, that these dummies reference the second quartile, not the first.

Classroom Variables

- *Graded configuration of the school*: a three-level indicator of “grade 9–12 and 10–12 high schools,” “grade 6–8 middle schools and grade 7–9 junior high schools,” and “all other schools where pre-algebra or algebra are taught.” These are used in the regressions of **non-target class dependent variables** only.

  Results showed that there were differences among high schools, middle schools, and other types of schools teaching algebra. However, on further inspection, it was found that the effects were generated not by the types of schools, but by the grades of those schools. In other words, it is not the middle school that is different than the high school, but that it is 7th-grade that is different from 9th-grade classes. For this reason, two dummy variables were included in the models of target class dependent variables, one that indicates that the class is primarily 7th- and 8th-grade students, and another dummy variable indicating that the class is primarily either a 10th-, 11th-, or 12th-grade class. The effects of each reference the difference between those classes and the traditional ninth-grade class.

- NORC controlled for the size of the classroom with dummies that indicate smaller classes (15 or fewer, 16 to 20, 26 to 30, 31 to 35, and more than 36 students). These variables reference the typical size of 20 to 25. While these refer to the target class, dummies were also included in the regressions of the nontarget dependent variables on the assumption that they proxy student-teacher ratios in the school more generally.

Teacher Background Variables

- All of the regression tables included controls for teacher sex, age, and race/ethnicity (dummy variables for Hispanic and for non-Hispanic black; reference group is all other identifications). Teacher age is centered on age = 40 to improve interpretability of the regression intercept (constant) term.

Table C-1: Regressions of Teachers' Summary Ratings of Student Background Preparation for Algebra I on Grade Level and Class Size of the Target Class, and School and Teacher Demographic Variables, 2007

Independent Variable	Model (1)	(2)	(3)	(4)	(5)	(6)
Class is 7th or 8th grade (ref: 9th)	-0.879***	-0.978***				
	(0.17)	(0.14)				
Class is 10th, 11th, or 8th grade (ref: 9th)	0.288	0.209				
	(0.17)	(0.15)				
Class size LE 15 students (ref: 21–25)	0.0323		0.0515			
	(0.28)		(0.23)			
Class size 16–20 students (ref: 21–25)	-0.0462		-0.0620			
	(0.12)		(0.16)			
Class size 26–30 students (ref: 21–25)	-0.108		-0.256			
	(0.12)		(0.15)			
Class size 31–35 students (ref: 21–25)	0.263		0.0598			
	(0.22)		(0.25)			
Class size GE 36 students (ref: 21–25)	-0.285		-0.775*			
	(0.29)		(0.31)			
School size: 1st quartile (ref: 2nd)	0.140		0.472*			
	(0.20)		(0.21)			
School size: 3rd quartile (ref: 2nd)	0.239		0.287			
	(0.21)		(0.23)			
School size: 4th quartile (ref: 2nd)	0.212		0.799***			
	(0.20)		(0.19)			
Sch N of minority students: 2nd quartile (ref: 1st)	-0.277			-0.0426		-0.224
	(0.18)			(0.20)		(0.20)
Sch N of minority students: 3rd quartile (ref: 1st)	-0.137			0.305		0.124
	(0.15)			(0.18)		(0.17)
Sch N of minority students: 4th quartile (ref: 1st)	0.0507			0.572**		0.341
	(0.19)			(0.18)		(0.19)
School N FRPL: 2nd quartile (ref: 1st)	0.0000157				0.0906	
	(0.15)				(0.16)	
School N FRPL: 3rd quartile (ref: 1st)	0.264				0.299	
	(0.19)				(0.16)	
School N FRPL: 4th quartile (ref: 1st)	0.0416				0.0836	
	(0.22)				(0.26)	
Urban school (ref: suburban)	0.150					0.0168
	(0.14)					(0.15)
Rural school (ref: suburban)	-0.264					-0.345*
	(0.15)					(0.14)
Teacher is female (ref: male)	-0.0614					
	(0.11)					
Teacher's age (centered on age 40)	-0.000181					
	(0.0046)					
Teacher is black (ref: white, Asian)	-0.121					
	(0.22)					
Teacher is Hispanic (ref: Non-Hispanic)	0.0814					
	(0.11)					
Constant	-0.0475	0.0353	-0.722***	-0.524***	-0.445***	-0.273
	(0.28)	(0.063)	(0.19)	(0.14)	(0.11)	(0.14)
Observations	660	720	723	725	713	725
R-squared	0.31	0.23	0.10	0.06	0.01	0.07

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. Standard errors in parentheses.

Note: The items used to construct the dependent summary scale range from 1 = excellent [preparation] to 4 = poor [preparation]. Negative coefficients in this table thus represent more favorable ratings and positive coefficients less favorable ratings.

LE: Less than or equal to; GE: Greater than or equal to; FRPL: Free or reduced-price lunch

Source: Based on responses to the National Opinion Research Center's National Survey of Algebra Teachers, 2007.

Table C-2: Regressions of Teachers' Summary Ratings of Importance-Weighted Preparation for Algebra I on Grade Level and Class Size of the Target Class, and School and Teacher Demographic Variables, 2007

	Model					
Independent Variable	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Class is 7th or 8th grade (ref: 9th)	-0.715***	-0.789***				
	(0.17)	(0.13)				
Class is 10th, 11th, or 8th grade (ref: 9th)	0.203	0.149				
	(0.13)	(0.13)				
Class size LE 15 students (ref: 21–25)	0.0551		0.132			
	(0.29)		(0.24)			
Class size 16–20 students (ref: 21–25)	-0.0974		-0.0932			
	(0.11)		(0.13)			
Class size 26–30 students (ref: 21–25)	-0.0952		-0.227			
	(0.12)		(0.12)			
Class size 31–35 students (ref: 21–25)	0.250		0.111			
	(0.22)		(0.23)			
Class size GE 36 students (ref: 21–25)	-0.367		-0.735*			
	(0.27)		(0.29)			
School size: 1st quartile (ref: 2nd)	0.0123		0.283			
	(0.19)		(0.17)			
School size: 3rd quartile (ref: 2nd)	0.134		0.199			
	(0.18)		(0.19)			
School size: 4th quartile (ref: 2nd)	0.204		0.677***			
	(0.19)		(0.15)			
Sch N of minority students: 2nd quartile (ref: 1st)	-0.166			0.000194		-0.121
	(0.16)			(0.17)		(0.17)
Sch N of minority students: 3rd quartile (ref: 1st)	0.0369			0.325*		0.232
	(0.16)			(0.15)		(0.15)
Sch N of minority students: 4th quartile (ref: 1st)	0.227			0.530**		0.416*
	(0.19)			(0.16)		(0.17)
School N FRPL: 2nd quartile (ref: 1st)	-0.0319				0.0711	
	(0.15)				(0.15)	
School N FRPL: 3rd quartile (ref: 1st)	0.140				0.225	
	(0.18)				(0.15)	
School N FRPL: 4th quartile (ref: 1st)	-0.0307				0.0752	
	(0.21)				(0.20)	
Urban school (ref: suburban)	0.0879					-0.0697
	(0.14)					(0.14)
Rural school (ref: suburban)	-0.113					-0.250
	(0.14)					(0.13)
Teacher is female (ref: male)	0.130					
	(0.098)					
Teacher's age (centered on age 40)	0.00270					
	(0.0044)					
Teacher is black (ref: white, Asian)	-0.128					
	(0.22)					
Teacher is Hispanic (ref: Non-Hispanic)	0.117					
	(0.14)					
Constant	-0.351	-0.00488	-0.594***	-0.500***	-0.387***	-0.320*
	(0.25)	(0.056)	(0.15)	(0.11)	(0.10)	(0.13)
Observations	640	697	700	702	690	702
R-squared	0.23	0.17	0.09	0.05	0.01	0.06

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. Standard errors in parentheses.

Note: The items used to construct the dependent summary scale range from 0 = not a problem [preparation] to 4 = serious problem [preparation]. Negative coefficients in this table thus represent more favorable ratings and positive coefficients less favorable ratings.

LE: Less than or equal to; GE: Greater than or equal to; FRPL: Free or reduced-price lunch

Source: Based on responses to the National Opinion Research Center's National Survey of Algebra Teachers, 2007.

Table C-3: Regressions of Teachers' Summary Ratings of Content Standards for Algebra I in Their State or Local District on School Grade Level and Class Size of the Target Class, and School and Teacher Demographic Variables, 2007

	Model					
Independent Variable	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
School is a middle or other school						
(ref: 9th to 12th grade high school)	0.187	0.212				
	(0.14)	(0.16)				
Class size LE 15 students (ref: 21–25)	0.432		0.495			
	(0.33)		(0.40)			
Class size 16–20 students (ref: 21–25)	-0.0409		-0.0104			
	(0.16)		(0.18)			
Class size 26–30 students (ref: 21–25)	-0.193		-0.139			
	(0.13)		(0.13)			
Class size 31–35 students (ref: 21–25)	-0.0323		-0.0191			
	(0.19)		(0.17)			
Class size GE 36 students (ref: 21–25)	0.0671		-0.0597			
	(0.19)		(0.19)			
School size: 1st quartile (ref: 2nd)	0.126		0.177			
	(0.28)		(0.26)			
School size: 3rd quartile (ref: 2nd)	-0.0607		0.0239			
	(0.18)		(0.18)			
School size: 4th quartile (ref: 2nd)	-0.162		-0.198			
	(0.18)		(0.12)			
Sch N of minority students: 2nd quartile (ref: 1st)	-0.365*			-0.495*		-0.402
	(0.16)			(0.22)		(0.21)
Sch N of minority students: 3rd quartile (ref: 1st)	-0.131			-0.197		-0.124
	(0.26)			(0.26)		(0.31)
Sch N of minority students: 4th quartile (ref: 1st)	-0.142			-0.220		-0.140
	(0.24)			(0.24)		(0.26)
School N FRPL: 2nd quartile (ref: 1st)	0.110				0.276	
	(0.18)				(0.22)	
School N FRPL: 3rd quartile (ref: 1st)	0.0563				0.279	
	(0.16)				(0.18)	
School N FRPL: 4th quartile (ref: 1st)	-0.374				-0.0607	
	(0.19)				(0.14)	
Urban school (ref: suburban)	0.179					0.0981
	(0.15)					(0.15)
Rural school (ref: suburban)	0.0501					0.221
	(0.17)					(0.22)
Teacher is female (ref: male)	-0.00636					
	(0.14)					
Teacher's age (centered on age 40)	0.00154					
	(0.0039)					
Teacher is black (ref: white, Asian)	0.158					
	(0.37)					
Teacher is Hispanic (ref: Non-Hispanic)	0.580					
	(0.38)					
Constant	0.0660	-0.0171	0.111	0.299	-0.0530	0.139
	(0.40)	(0.053)	(0.15)	(0.21)	(0.076)	(0.20)
Observations	663	721	719	721	710	721
R-squared	0.12	0.01	0.06	0.03	0.02	0.04

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. Standard errors in parentheses.

Note: The items used to construct the dependent summary scale range from 1 = excellent [standards] to 4 = poor [standards]. Negative coefficients in this table thus represent more favorable ratings and positive coefficients less favorable ratings.

LE: Less than or equal to; GE: Greater than or equal to; FRPL: Free or reduced-price lunch

Source: Based on responses to the National Opinion Research Center's National Survey of Algebra Teachers, 2007.

Table C-4: Regressions of Teachers' Summary Ratings of Assessment Standards for Algebra I in Their State or Local District on School Grade Level and Class Size of the Target Class, and School And Teacher Demographic Variables, 2007

	Model					
Independent Variable	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
School is a middle or other school (ref: 9th to 12th grade high school)	0.00919	0.0642				
	(0.14)	(0.15)				
Class size LE 15 students (ref: 21–25)	0.326		0.393			
	(0.29)		(0.32)			
Class size 16–20 students (ref: 21–25)	0.153		0.217			
	(0.18)		(0.19)			
Class size 26–30 students (ref: 21–25)	-0.0656		-0.0100			
	(0.13)		(0.14)			
Class size 31–35 students (ref: 21–25)	-0.0754		-0.131			
	(0.18)		(0.18)			
Class size GE 36 students (ref: 21–25)	-0.150		-0.243			
	(0.18)		(0.18)			
School size: 1st quartile (ref: 2nd)	0.0933		0.118			
	(0.28)		(0.27)			
School size: 3rd quartile (ref: 2nd)	0.215		0.169			
	(0.22)		(0.22)			
School size: 4th quartile (ref: 2nd)	0.0594		-0.0318			
	(0.21)		(0.20)			
Sch N of minority students: 2nd quartile (ref: 1st)	-0.308			-0.373		-0.284
	(0.19)			(0.24)		(0.22)
Sch N of minority students: 3rd quartile (ref: 1st)	-0.0853			-0.0576		0.000997
	(0.23)			(0.24)		(0.24)
Sch N of minority students: 4th quartile (ref: 1st)	-0.332			-0.281		-0.219
	(0.24)			(0.24)		(0.25)
School N FRPL: 2nd quartile (ref: 1st)	0.00595				0.151	
	(0.20)				(0.20)	
School N FRPL: 3rd quartile (ref: 1st)	0.148				0.254	
	(0.23)				(0.17)	
School N FRPL: 4th quartile (ref: 1st)	-0.210				-0.109	
	(0.24)				(0.17)	
Urban school (ref: suburban)	0.225					0.125
	(0.14)					(0.15)
Rural school (ref: suburban)	0.166					0.219
	(0.18)					(0.18)
Teacher is female (ref: male)	0.0360					
	(0.14)					
Teacher's age (centered on age 40)	0.00110					
	(0.0045)					
Teacher is black (ref: white, Asian)	-0.0489					
	(0.33)					
Teacher is Hispanic (ref: Non-Hispanic)	0.446					
	(0.27)					
Constant	-0.166	-0.00235	-0.0979	0.190	-0.0576	0.0311
	(0.38)	(0.068)	(0.19)	(0.21)	(0.097)	(0.19)
Observations	650	708	706	708	697	708
R-squared	0.07	0.00	0.03	0.02	0.02	0.03

Note: The items used to construct the dependent summary scale range from 1 = excellent [standards] to 4 = poor [standards]. Negative coefficients in this table thus represent more favorable ratings and positive coefficients less favorable ratings.

LE: Less than or equal to; GE: Greater than or equal to; FRPL: Free or reduced-price lunch

Source: Based on responses to the National Opinion Research Center's National Survey of Algebra Teachers, 2007.

Table C-5: Regressions of Teachers' Summary Ratings of Algebra I Textbooks on Grade Level and Class Size of the Target Class, and School and Teacher Demographic Variables, 2007

Independent Variable	Model					
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Class is 7th or 8th grade (ref: 9th)	-0.234	-0.251				
	(0.12)	(0.16)				
Class is 10th, 11th, or 8th grade (ref: 9th)	-0.0392	-0.00355				
	(0.19)	(0.21)				
Class size LE 15 students (ref: 21–25)	-0.561*		-0.726*			
	(0.23)		(0.32)			
Class size 16–20 students (ref: 21–25)	-0.105		-0.206			
	(0.14)		(0.18)			
Class size 26–30 students (ref: 21–25)	-0.0487		-0.155			
	(0.15)		(0.18)			
Class size 31–35 students (ref: 21–25)	-0.139		-0.145			
	(0.16)		(0.19)			
Class size GE 36 students (ref: 21–25)	-0.287		-0.367*			
	(0.16)		(0.17)			
School size: 1st quartile (ref: 2nd)	0.151		0.355			
	(0.23)		(0.29)			
School size: 3rd quartile (ref: 2nd)	0.117		0.401*			
	(0.18)		(0.18)			
School size: 4th quartile (ref: 2nd)	-0.00993		0.469**			
	(0.18)		(0.17)			
Sch N of minority students: 2nd quartile (ref: 1st)	0.0229			0.298		0.0631
	(0.18)			(0.21)		(0.19)
Sch N of minority students: 3rd quartile (ref: 1st)	0.0214			0.275		0.0283
	(0.16)			(0.17)		(0.16)
Sch N of minority students: 4th quartile (ref: 1st)	0.517*			0.705**		0.387*
	(0.23)			(0.23)		(0.19)
School N FRPL: 2nd quartile (ref: 1st)	-0.204				-0.255	
	(0.15)				(0.19)	
School N FRPL: 3rd quartile (ref: 1st)	-0.0127				0.0508	
	(0.18)				(0.16)	
School N FRPL: 4th quartile (ref: 1st)	-0.362				-0.0705	
	(0.21)				(0.27)	
Urban school (ref: suburban)	0.0482					0.0683
	(0.14)					(0.15)
Rural school (ref: suburban)	-0.347*					-0.441***
	(0.15)					(0.13)
Teacher is female (ref: male)	-0.172					
	(0.11)					
Teacher's age (centered on age 40)	0.00235					
	(0.0044)					
Teacher is black (ref: white, Asian)	-0.132					
	(0.27)					
Teacher is Hispanic (ref: Non-Hispanic)	-0.500*					
	(0.24)					
Constant	0.193	-0.0116	-0.284	-0.404**	-0.0353	-0.0844
	(0.26)	(0.11)	(0.17)	(0.15)	(0.12)	(0.14)
Observations	636	693	696	698	686	698
R-squared	0.17	0.02	0.07	0.06	0.02	0.10

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. Standard errors in parentheses.

Note: The items used to construct the dependent summary scale range from 1 = strongly agree [that the text has some quality] to 5 = strongly disagree [that the text has some quality]. Negative coefficients in this table thus represent more favorable ratings and positive coefficients represent negative ratings.

LE: Less than or equal to; GE: Greater than or equal to; FRPL: Free or reduced-price lunch

Source: Based on responses to the National Opinion Research Center's National Survey of Algebra Teachers, 2007.

Table C-6: Regressions of Teachers' Summary Ratings of Technology Use in Algebra I on Grade Level and Class Size of the Target Class, and School and Teacher Demographic Variables, 2007

Independent Variable	Model (1)	(2)	(3)	(4)	(5)	(6)
Class is 7th or 8th grade (ref: 9th)	-0.134 (0.16)	-0.00412 (0.17)				
Class is 10th, 11th, or 8th grade (ref: 9th)	0.142 (0.16)	0.181 (0.21)				
Class size LE 15 students (ref: 21–25)	0.292 (0.28)		0.385 (0.36)			
Class size 16–20 students (ref: 21–25)	0.545** (0.18)		0.604* (0.24)			
Class size 26–30 students (ref: 21–25)	0.254 (0.17)		0.255 (0.18)			
Class size 31–35 students (ref: 21–25)	0.374 (0.23)		0.413 (0.22)			
Class size GE 36 students (ref: 21–25)	0.276 (0.22)		0.350 (0.27)			
School size: 1st quartile (ref: 2nd)	-0.135 (0.30)		-0.115 (0.35)			
School size: 3rd quartile (ref: 2nd)	0.255 (0.22)		0.246 (0.19)			
School size: 4th quartile (ref: 2nd)	-0.0777 (0.25)		0.00865 (0.17)			
Sch N of minority students: 2nd quartile (ref: 1st)	-0.152 (0.16)			-0.0642 (0.19)		-0.116 (0.18)
Sch N of minority students: 3rd quartile (ref: 1st)	-0.0132 (0.16)			0.124 (0.18)		0.142 (0.19)
Sch N of minority students: 4th quartile (ref: 1st)	-0.417 (0.28)			-0.443 (0.32)		-0.398 (0.26)
School N FRPL: 2nd quartile (ref: 1st)	0.0692 (0.18)				0.0791 (0.16)	
School N FRPL: 3rd quartile (ref: 1st)	0.275 (0.20)				0.173 (0.15)	
School N FRPL: 4th quartile (ref: 1st)	-0.295 (0.29)				-0.475 (0.29)	
Urban school (ref: suburban)	-0.0101 (0.15)					-0.225 (0.19)
Rural school (ref: suburban)	-0.176 (0.17)					-0.146 (0.15)
Teacher is female (ref: male)	0.241* (0.12)					
Teacher's age (centered on age 40)	-0.000164 (0.0072)					
Teacher is black (ref: white, Asian)	-0.250 (0.22)					
Teacher is Hispanic (ref: Non-Hispanic)	0.273 (0.30)					
Constant	-0.229 (0.40)	-0.0194 (0.13)	-0.326 (0.21)	0.0638 (0.15)	0.0171 (0.11)	0.170 (0.15)
Observations	650	709	712	714	702	714
R-squared	0.14	0.00	0.05	0.04	0.05	0.04

** $p < 0.01$, * $p < 0.05$. Standard errors in parentheses.

Note: The items used to construct the dependent summary scale range from 1 = strongly agree [that the technology is helpful] to 5 = strongly disagree [that the technology is helpful]. Negative coefficients in this table thus represent more favorable ratings and positive coefficients represent negative ratings.

LE: Less than or equal to; GE: Greater than or equal to; FRPL: Free or reduced-price lunch

Source: Based on responses to the National Opinion Research Center's National Survey of Algebra Teachers, 2007.

Table C-7a: Regressions of Teachers' Summary Ratings on the Helpfulness of Pre-Service Teacher Training in Teaching Algebra I on Grade Level and Class Size of the Target Class, and School and Teacher Demographic Variables, 2007

	Model					
Independent Variable	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
School is a middle or other school (ref: 9th to 12th grade high school)	0.114	0.209				
	(0.14)	(0.14)				
Class size LE 15 students (ref: 21–25)	0.192		0.187			
	(0.32)		(0.35)			
Class size 16–20 students (ref: 21–25)	0.0981		0.0940			
	(0.17)		(0.16)			
Class size 26–30 students (ref: 21–25)	0.350*		0.389*			
	(0.16)		(0.16)			
Class size 31–35 students (ref: 21–25)	0.377		0.452*			
	(0.24)		(0.23)			
Class size GE 36 students (ref: 21–25)	0.737**		0.733**			
	(0.27)		(0.25)			
School size: 1st quartile (ref: 2nd)	0.110		0.0961			
	(0.25)		(0.24)			
School size: 3rd quartile (ref: 2nd)	-0.123		-0.0998			
	(0.23)		(0.18)			
School size: 4th quartile (ref: 2nd)	-0.290		-0.231			
	(0.21)		(0.14)			
Sch N of minority students: 2nd quartile (ref: 1st)	-0.123			-0.123		-0.200
	(0.21)			(0.22)		(0.22)
Sch N of minority students: 3rd quartile (ref: 1st)	-0.303			-0.287		-0.372
	(0.22)			(0.20)		(0.21)
Sch N of minority students: 4th quartile (ref: 1st)	-0.179			-0.187		-0.296
	(0.21)			(0.18)		(0.19)
School N FRPL: 2nd quartile (ref: 1st)	-0.000303				-0.0585	
	(0.18)				(0.21)	
School N FRPL: 3rd quartile (ref: 1st)	-0.0584				-0.117	
	(0.18)				(0.16)	
School N FRPL: 4th quartile (ref: 1st)	-0.0590				-0.0236	
	(0.19)				(0.18)	
Urban school (ref: suburban)	0.0632					0.00919
	(0.14)					(0.13)
Rural school (ref: suburban)	-0.274					-0.159
	(0.18)					(0.18)
Teacher is female (ref: male)	0.0959					
	(0.12)					
Teacher's age (centered on age 40)	0.000408					
	(0.0045)					
Teacher is black (ref: white, Asian)	0.0355					
	(0.34)					
Teacher is Hispanic (ref: Non-Hispanic)	-0.642***					
	(0.19)					
Constant	0.197	0.0143	0.0266	0.269	0.169	0.385*
	(0.35)	(0.054)	(0.16)	(0.17)	(0.12)	(0.18)
Observations	673	734	732	734	722	734
R-squared	0.08	0.01	0.04	0.01	0.00	0.02

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. Standard errors in parentheses.

Note: The items used to construct the dependent summary scale range from 1 = Very Well to 4 = Very Poorly. Negative coefficients in this table thus represent more favorable ratings and positive coefficients less favorable ratings.

LE: Less than or equal to; GE: Greater than or equal to; FRPL: Free or reduced-price lunch

Source: Based on responses to the National Opinion Research Center's National Survey of Algebra Teachers, 2007.

Table C-7b: Regressions of Teachers' Summary Ratings on the Helpfulness of Professional Development for Teaching Algebra I on Grade Level and Class Size of the Target Class, and School and Teacher Demographic Variables, 2007

	Model					
Independent Variable	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
School is a middle or other school (ref: 9th to 12th grade high school)	-0.123 (0.11)	0.0320 (0.11)				
Class size LE 15 students (ref: 21–25)	0.0342 (0.20)		0.0142 (0.19)			
Class size 16–20 students (ref: 21–25)	0.0811 (0.14)		0.0829 (0.13)			
Class size 26–30 students (ref: 21–25)	0.154 (0.15)		0.157 (0.15)			
Class size 31–35 students (ref: 21–25)	0.261 (0.23)		0.278 (0.22)			
Class size GE 36 students (ref: 21–25)	0.663** (0.21)		0.594** (0.23)			
School size: 1st quartile (ref: 2nd)	-0.0784 (0.23)		-0.0831 (0.21)			
School size: 3rd quartile (ref: 2nd)	0.120 (0.23)		0.0239 (0.21)			
School size: 4th quartile (ref: 2nd)	-0.263 (0.22)		-0.248 (0.17)			
Sch N of minority students: 2nd quartile (ref: 1st)	-0.245 (0.19)			-0.194 (0.16)		-0.194 (0.19)
Sch N of minority students: 3rd quartile (ref: 1st)	-0.195 (0.18)			-0.144 (0.15)		-0.182 (0.18)
Sch N of minority students: 4th quartile (ref: 1st)	-0.281 (0.25)			-0.145 (0.16)		-0.207 (0.19)
School N FRPL: 2nd quartile (ref: 1st)	-0.170 (0.16)				-0.0963 (0.14)	
School N FRPL: 3rd quartile (ref: 1st)	-0.100 (0.17)				-0.0636 (0.15)	
School N FRPL: 4th quartile (ref: 1st)	-0.0769 (0.24)				0.0324 (0.18)	
Urban school (ref: suburban)	0.184 (0.15)					0.128 (0.14)
Rural school (ref: suburban)	0.0300 (0.18)					0.0263 (0.15)
Teacher is female (ref: Male)	-0.00290 (0.11)					
Teacher's age (centered on age 40)	-0.00502 (0.0043)					
Teacher is black (ref: white, Asian)	0.464 (0.28)					
Teacher is Hispanic (ref: Non-Hispanic)	0.313 (0.36)					
Constant	0.539 (0.33)	0.0741 (0.062)	0.111 (0.19)	0.207 (0.11)	0.128 (0.100)	0.188 (0.16)
Observations	675	736	734	736	724	736
R-squared	0.05	0.00	0.02	0.01	0.00	0.01

** $p < 0.01$. Standard errors in parentheses.

Note: The items used to construct the dependent summary scale range from 1 = Very Well to 4 = Very Poorly. Negative coefficients in this table thus represent more favorable ratings and positive coefficients less favorable ratings.

LE: Less than or equal to; GE: Greater than or equal to; FRPL: Free or reduced-price lunch

Source: Based on responses to the National Opinion Research Center's National Survey of Algebra Teachers, 2007.

Table C-8: Regressions of Teachers' Summary Ratings of Remedial Help for Algebra I Students on Grade Level and Class Size of the Target Class, and School and Teacher Demographic Variables, 2007

Independent Variable	Model (1)	(2)	(3)	(4)	(5)	(6)
School is a middle or other school (ref: 9th to 12th grade high school)	0.243	0.440*				
	(0.16)	(0.17)				
Class size LE 15 students (ref: 21–25)	-0.0591		0.0301			
	(0.21)		(0.25)			
Class size 16–20 students (ref: 21–25)	-0.176		-0.00806			
	(0.17)		(0.20)			
Class size 26–30 students (ref: 21–25)	-0.237		0.0607			
	(0.17)		(0.20)			
Class size 31–35 students (ref: 21–25)	0.0786		0.408			
	(0.23)		(0.24)			
Class size GE 36 students (ref: 21–25)	-0.302		-0.173			
	(0.34)		(0.23)			
School size: 1st quartile (ref: 2nd)	-0.0702		-0.310			
	(0.32)		(0.37)			
School size: 3rd quartile (ref: 2nd)	-0.287		-0.486			
	(0.28)		(0.32)			
School size: 4th quartile (ref: 2nd)	-0.449		-0.767**			
	(0.28)		(0.29)			
Sch N of minority students: 2nd quartile (ref: 1st)	-0.258			-0.463		-0.412
	(0.22)			(0.24)		(0.26)
Sch N of minority students: 3rd quartile (ref: 1st)	-0.0232			-0.188		-0.113
	(0.27)			(0.27)		(0.32)
Sch N of minority students: 4th quartile (ref: 1st)	-0.814**			-0.587*		-0.485
	(0.26)			(0.23)		(0.27)
School N FRPL: 2nd quartile (ref: 1st)	-0.137				-0.0589	
	(0.20)				(0.18)	
School N FRPL: 3rd quartile (ref: 1st)	-0.0906				-0.107	
	(0.23)				(0.24)	
School N FRPL: 4th quartile (ref: 1st)	0.576*				0.351	
	(0.28)				(0.27)	
Urban school (ref: suburban)	-0.119					-0.0722
	(0.16)					(0.18)
Rural school (ref: suburban)	-0.188					0.0913
	(0.22)					(0.24)
Teacher is female (ref: male)	0.273*					
	(0.12)					
Teacher's age (centered on age 40)	-0.0128**					
	(0.0049)					
Teacher is black (ref: white, Asian)	0.520*					
	(0.21)					
Teacher is Hispanic (ref: Non-Hispanic)	-0.473					
	(0.35)					
Constant	1.100**	0.0259	0.682*	0.531**	0.225	0.464
	(0.41)	(0.082)	(0.29)	(0.20)	(0.14)	(0.24)
Observations	660	717	715	717	705	717
R-squared	0.20	0.04	0.07	0.04	0.02	0.04

** $p < 0.01$, * $p < 0.05$. Standard errors in parentheses.

Note: The items used to construct the dependent summary scale range from 1 = excellent [tutoring] to 5 = poor [tutoring]. Negative coefficients in this table thus represent more favorable ratings and positive coefficients represent negative ratings.

LE: Less than or equal to; GE: Greater than or equal to; FRPL: Free or reduced-price lunch

Source: Based on responses to the National Opinion Research Center's National Survey of Algebra Teachers, 2007.

Table C-9: Regressions of Teachers' Summary Ratings of Extent to Which They See Different Levels of Students in the Same Algebra I Class as a Problem in Their School On Grade Level and Class Size of the Target Class, and School and Teacher Demographic Variables

Independent Variable	Model (1)	(2)	(3)	(4)	(5)	(6)
School is a middle or other school (ref: 9th to 12th grade high school)	-0.314**	-0.292*				
	(0.12)	(0.13)				
Class size LE 15 students (ref: 21–25)	-0.152		-0.280			
	(0.22)		(0.24)			
Class size 16–20 students (ref: 21–25)	0.0738		-0.0492			
	(0.16)		(0.19)			
Class size 26–30 students (ref: 21–25)	0.252		0.155			
	(0.17)		(0.18)			
Class size 31–35 students (ref: 21–25)	0.620*		0.514*			
	(0.26)		(0.25)			
Class size GE 36 students (ref: 21–25)	-0.0995		-0.318			
	(0.26)		(0.31)			
School size: 1st quartile (ref: 2nd)	0.428		0.548*			
	(0.25)		(0.26)			
School size: 3rd quartile (ref: 2nd)	0.0937		0.126			
	(0.23)		(0.23)			
School size: 4th quartile (ref: 2nd)	0.200		0.357			
	(0.22)		(0.20)			
Sch N of minority students: 2nd quartile (ref: 1st)	-0.253			-0.194		-0.258
	(0.18)			(0.17)		(0.17)
Sch N of minority students: 3rd quartile (ref: 1st)	-0.0220			0.183		0.0684
	(0.18)			(0.16)		(0.17)
Sch N of minority students: 4th quartile (ref: 1st)	0.198			0.478*		0.318
	(0.25)			(0.19)		(0.18)
School N FRPL: 2nd quartile (ref: 1st)	0.107				0.130	
	(0.16)				(0.15)	
School N FRPL: 3rd quartile (ref: 1st)	0.152				0.229	
	(0.21)				(0.17)	
School N FRPL: 4th quartile (ref: 1st)	0.156				0.401	
	(0.30)				(0.23)	
Urban school (ref: suburban)	0.155					0.155
	(0.14)					(0.16)
Rural school (ref: suburban)	-0.0467					-0.102
	(0.18)					(0.14)
Teacher is female (ref: male)	0.116					
	(0.12)					
Teacher's age (centered on age 40)	0.00645					
	(0.0050)					
Teacher is black (ref: white, Asian)	-0.432*					
	(0.20)					
Teacher is Hispanic (ref: Non-Hispanic)	0.317					
	(0.24)					
Constant	-0.674	0.0620	-0.399	-0.192	-0.255*	-0.117
	(0.38)	(0.080)	(0.21)	(0.11)	(0.12)	(0.14)
Observations	675	735	733	735	723	735
R-squared	0.12	0.02	0.05	0.05	0.02	0.06

** $p < 0.01$, * $p < 0.05$. Standard errors in parentheses.

Note: The items used to construct the dependent summary scale range from 1 = not an problem [mixed classes] to 5 = is a serious problem [mixed classes]. Negative coefficients in this table thus represent more favorable ratings and positive coefficients represent negative ratings.

LE: Less than or equal to; GE: Greater than or equal to; FRPL: Free or reduced-price lunch

Source: Based on responses to the National Opinion Research Center's National Survey of Algebra Teachers, 2007.

Table C-10: Regressions of Teachers' Summary Ratings of Family Participation is a Problem in Algebra I on Grade Level and Class Size of the Target Class, and School And Teacher Demographic Variables, 2007

	Model					
Independent Variable	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
School is a middle or other school (ref: 9th to 12th grade high school)	-0.650*** (0.12)	-0.681*** (0.12)				
Class size LE 15 students (ref: 21–25)	-0.0983 (0.21)		-0.284 (0.28)			
Class size 16–20 students (ref: 21–25)	0.0749 (0.14)		-0.132 (0.19)			
Class size 26–30 students (ref: 21–25)	0.0953 (0.17)		-0.0917 (0.16)			
Class size 31–35 students (ref: 21–25)	0.316 (0.22)		0.0871 (0.25)			
Class size GE 36 students (ref: 21–25)	-0.415* (0.19)		-0.945** (0.32)			
School size: 1st quartile (ref: 2nd)	0.0318 (0.22)		0.137 (0.27)			
School size: 3rd quartile (ref: 2nd)	0.00348 (0.21)		0.00244 (0.22)			
School size: 4th quartile (ref: 2nd)	-0.0545 (0.21)		0.184 (0.19)			
Sch N of minority students: 2nd quartile (ref: 1st)	-0.301 (0.20)			-0.292 (0.18)		-0.358 (0.19)
Sch N of minority students: 3rd quartile (ref: 1st)	-0.0448 (0.17)			0.214 (0.16)		0.170 (0.19)
Sch N of minority students: 4th quartile (ref: 1st)	0.369 (0.22)			0.701*** (0.17)		0.655** (0.20)
School N FRPL: 2nd quartile (ref: 1st)	0.314* (0.15)				0.346* (0.17)	
School N FRPL: 3rd quartile (ref: 1st)	0.458* (0.19)				0.531** (0.19)	
School N FRPL: 4th quartile (ref: 1st)	0.543* (0.23)				0.830*** (0.20)	
Urban school (ref: suburban)	-0.129 (0.15)					-0.0898 (0.14)
Rural school (ref: suburban)	-0.200 (0.18)					-0.157 (0.16)
Teacher is female (ref: male)	-0.223* (0.10)					
Teacher's age (centered on age 40)	-0.00146 (0.0043)					
Teacher is black (ref: white, Asian)	-0.426 (0.26)					
Teacher is Hispanic (ref: Non-Hispanic)	-0.0157 (0.17)					
Constant	0.179 (0.31)	0.206** (0.075)	-0.138 (0.21)	-0.270* (0.12)	-0.526*** (0.13)	-0.155 (0.17)
Observations	673	733	731	733	721	733
R-squared	0.25	0.11	0.03	0.11	0.08	0.11

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. Standard errors in parentheses.

Note: The items used to construct the dependent summary scale range from 1 = not an problem [family help] to 5 = is a serious problem [family help]. Negative coefficients in this table thus represent more favorable ratings and positive coefficients represent negative ratings.

LE: Less than or equal to; GE: Greater than or equal to; FRPL: Free or reduced-price lunch

Source: Based on responses to the National Opinion Research Center's National Survey of Algebra Teachers, 2007.

APPENDIX D: Means and Confidence Intervals for Items in the National Survey of Algebra Teachers

Item	Low 95% CI	Mean	High 95% CI
Section 1: Your Algebra I Class			
1. How many students are in your Target Class?			
Number of students in Target Class less than 15	0.05	0.10	0.15
Number of students in Target Class 15–20	0.21	0.27	0.32
Number of students in Target Class 21–25	0.26	0.32	0.38
Number of students in Target Class 26–30	0.18	0.25	0.31
Number of students in Target Class 31–35	0.04	0.07	0.10
Scale = Proportion			
2. How many of the students in your Target Class:			
a. Are in the 7th grade	0.10	0.21	0.33
b. Are in the 8th grade	1.35	1.65	1.96
c. Are in the 9th grade	1.74	2.00	2.25
d. Are in the 10th grade	0.56	0.68	0.80
e. Are in the 11th grade	0.26	0.33	0.40
f. Are in the 12th grade	0.11	0.17	0.23
g. Are in special education (i.e., have an IEP)	0.53	0.61	0.69
h. Are currently enrolled in your school's bilingual program	0.23	0.34	0.44
Scale = Proportion			

Continued on p. 9-62

Appendix D, continued

Item	Low 95% CI	Mean	High 95% CI
3. How many students in your Target Class:			
a. Come to class on time	3.49	3.57	3.65
b. Attend class regularly	3.39	3.46	3.54
c. Come to class prepared with appropriate supplies and books	2.79	2.92	3.05
d. Create serious behavior problems in your class	0.53	0.61	0.69
e. Regularly pay attention in class	2.70	2.82	2.93
f. Actively participate in class activities	2.57	2.69	2.80
g. Take notes	2.59	2.72	2.86
h. Have serious difficulties reading English	0.41	0.47	0.54
i. Care about what grade they receive	2.78	2.90	3.02
Scale: 0 = None 1 = Some 2 = About Half 3 = Most 4 = Nearly All			
4. Based on your experience with incoming Algebra I students in your Target Class, how would you rate students' background in each of the following areas of mathematics?			
a. Whole numbers and operations with whole numbers	1.77	1.86	1.95
b. Positive and negative integers and operations with positive and negative integers	2.46	2.58	2.69
c. Rational numbers and operations involving fractions and decimals	2.97	3.10	3.22
d. Ratios, percents, rates, and proportions	2.71	2.83	2.95
e. Solving word problems	3.14	3.26	3.38
f. The concept of variables	2.38	2.48	2.58

Continued on p. 9-63

Appendix D, continued

Item	Low 95% CI	Mean	High 95% CI
g. Manipulation of variables	2.72	2.82	2.92
h. Solving simple linear equations and inequalities	2.70	2.80	2.89
i. Plotting points and graphing lines on the four-quadrant coordinate plane	2.32	2.44	2.56
j. Measurement formulas of basic geometric shapes	2.71	2.81	2.92
k. Basic study skills and work habits necessary for success in math	2.90	3.00	3.10
l. Computation skills	2.42	2.53	2.64
m. Ability to use math in contexts that are identified as real world situations	2.84	2.94	3.04
n. Working independently	2.48	2.58	2.68
o. Working cooperatively with other students	2.22	2.32	2.41
Scale: 1 = Excellent 2 = Good 3 = Fair 4 = Poor			
5. On average how often do you use the following instructional materials and tools in your Target Class?			
a. Textbooks	2.76	2.92	3.07
b. Printed instructional materials other than textbooks	2.49	2.60	2.71
c. Teacher/colleague written instructional materials	1.96	2.11	2.25
d. Graphing calculators (the school's or their own)	1.29	1.53	1.78
e. Physical objects (manipulatives)	1.13	1.26	1.38
f. Computer-based instructional tools (software)	0.81	1.00	1.20
Scale: 0 = Never 1 = Less than once a week 2 = About once a week 3 = Several times a week 4 = Everyday			

Continued on p. 9-64

Appendix D, continued

Item	Low 95% CI	Mean	High 95% CI
6. Please indicate your level of agreement or disagreement with the statement: "Computer-based instructional tools (software) are helping Algebra I students in my Target Class."	3.16	3.33	3.51
Scale: 1 = Strongly agree 2 = Somewhat agree 3 = Neither agree nor disagree 4 = Somewhat disagree 5 = Strongly disagree			
8. Please indicate your level of agreement or disagreement with each of the following statements regarding the Algebra I textbook you use in your Target Class.			
a. The textbook includes the appropriate topics and content to teach the course.	1.67	1.77	1.87
b. The textbook appropriately sequences math concepts.	2.09	2.23	2.38
c. The textbook provides examples and lessons that are focused directly on the mathematics involved and that explain concepts clearly.	1.96	2.09	2.22
d. The textbook provides opportunities for the development of problem-solving skills.	2.02	2.16	2.31
e. The textbook provides adequate practice for each topic covered.	2.12	2.29	2.45
f. The textbook and the supporting materials which come with it provide the right mix of useful suggestions and problems for homework assignments.	2.08	2.24	2.39
g. The textbook provides adequate supplementary/support materials.	2.12	2.27	2.43
h. The textbook is clearly focused on Algebra I and contains very few if any distractions to the instructional focus (e.g., off-task activities pictures with no sense of purpose, etc.).	1.90	2.01	2.13

Continued on p. 9-65

Appendix D, continued

Item	Low 95% CI	Mean	High 95% CI
i. The textbook and the accompanying materials provide useful suggestions for meeting the needs of diverse learners.	2.57	2.73	2.89
Scale: 1 = Strongly agree 2 = Somewhat agree 3 = Neither agree nor disagree 4 = Somewhat disagree 5 = Strongly disagree			
9. About what percentage of your current Algebra I students in your Target Class do you anticipate will fail your course?			
None will fail	0.16	0.22	0.28
1–10% will fail	0.34	0.41	0.47
11–20% will fail	0.12	0.18	0.24
21–30% will fail	0.05	0.08	0.11
41–50% will fail	0.03	0.06	0.08
More than 50% will fail	0.02	0.03	0.04
Scale = Proportion			
10. On average, about how much time per day do you expect your Algebra I students in your Target Class to spend on assignments outside of class?			
None	0.01	0.04	0.07
1–15 minutes	0.10	0.14	0.17
16–30 minutes	0.46	0.53	0.60
31–45 min	0.18	0.24	0.30
46–60 minutes	0.02	0.04	0.06
More than 60 minutes	0.00	0.00	0.00
Scale = Proportion			

Continued on p. 9-66

Appendix D, continued

Item	Low 95% CI	Mean	High 95% CI
11. On average, about how many of your Algebra I students in your Target Class complete their outside-of-class assignments? Scale = 1 All or almost all 2 = About two-thirds 3 = About one-third 4 = None or almost none	1.87	1.97	2.06
12. On average how many minutes per week does your Algebra I Target Class meet? Scale = Minutes	116.96	118.24	119.52
13. Does your Algebra I Target Class meet everyday? Scale = Proportion	0.76	0.83	0.89
14. How long is each period during which you teach Algebra I? Scale = Minutes	58.85	61.74	64.63
15. Is this enough instructional time to adequately teach Algebra I to your Target Class? Scale = Proportion	0.71	0.76	0.82
Section 2: Your School and Algebra I			
1. Below is a list of factors that may cause problems in Algebra I instruction. For each factor please indicate whether it is not a problem, a minor problem, a moderate problem, or a serious problem in your school.			
a. Insufficient access to computers	1.68	1.86	2.04
b. Inadequate access to graphing calculators	1.58	1.70	1.81

Continued on p. 9-67

Appendix D, continued

Item	Low 95% CI	Mean	High 95% CI
c. Poor quality or out-of-date textbooks	1.43	1.59	1.75
d. Class sizes are too large	1.84	1.97	2.10
e. Too little coordination or articulation between classes in the mathematics curriculum	1.62	1.75	1.87
f. Some teachers are inadequately prepared to teach Algebra I	1.32	1.41	1.49
g. Lack of teacher planning time	1.63	1.74	1.85
h. Students with different abilities and interests taking the same math classes	2.40	2.53	2.66
i. Too little parent/family support	2.61	2.74	2.87
j. Inadequate opportunities for professional learning	1.55	1.66	1.77
k. Inadequate administrative support	1.52	1.64	1.75
Scale: 1 = Not a problem 2 = Minor problem 3 = Moderate problem 4 = Serious problem			
2. Does your school offer different levels of Algebra I to groups of students based on ability?	0.39	0.47	0.54
Scale = Proportion			
3. How many CLASS PERIODS do you teach a WEEK? (Exclude study halls and homeroom periods.)			
Scale = Number of Periods	17.58	18.86	20.15
4. Is your Algebra I class part of block scheduling at your school?	0.26	0.34	0.41
Scale = Proportion			
5. On average how many minutes are you scheduled during the school day to prepare for classes?	55.69	59.29	62.89
Scale = Minutes			

Continued on p. 9-68

Appendix D, continued

Item	Low 95% CI	Mean	High 95% CI
6. On average how much time do you spend outside of the regular school day preparing for your Algebra I classes?	47.82	52.11	56.39
Scale = Minutes			
7. To what grades are you currently teaching Algebra I? (Check all that apply)			
% 7th grade	.04	.07	.10
% 8th grade	0.31	0.38	0.46
% 9th grade	0.51	0.58	0.65
% 10th grade	0.37	0.43	0.50
% 11th grade	0.22	0.28	0.33
% 12th grade	0.12	0.17	0.21
Scale = Proportion			
8. How do you rate the remedial help in your school for students who are struggling in Algebra I?			
a. Availability of tutoring or other remedial assistance	2.35	2.52	2.69
b. Quality of tutoring or other remedial assistance	2.26	2.42	2.58
Scale: 1 = Excellent 2 = Good 3 = Fair 4 = Poor			
Section 3: Your Views of Mathematics Education			
1. How important is a solid foundation in each of the following areas to students' success in Algebra I?			
a. Whole numbers and operations with whole numbers	4.58	4.65	4.72
b. Positive and negative integers and operations with positive and negative integers	4.71	4.77	4.83

Continued on p. 9-69

Appendix D, continued

Item	Low 95% CI	Mean	High 95% CI
c. Rational numbers and operations involving fractions and decimals	4.52	4.59	4.67
d. Ratios, percents, rates, and proportions	4.09	4.19	4.28
e. Solving problems involving whole numbers fractions and decimals	4.45	4.51	4.58
f. The concept of variables	4.53	4.61	4.69
g. Manipulation of variables	4.46	4.55	4.64
h. Solving simple linear equations and inequalities	4.34	4.44	4.53
i. Plotting points and graphing lines on the four-quadrant coordinate plane	4.25	4.35	4.44
j. Measurement formulas of basic geometric shapes	3.32	3.45	3.58
k. Basic study skills and work habits necessary for success in math	4.66	4.72	4.78
l. Computation skills	4.46	4.54	4.61
m. Ability to use math in contexts that are identified as real world situations	4.01	4.10	4.20
n. Working independently	4.26	4.34	4.42
o. Working cooperatively with other students	3.92	4.02	4.12
Scale: 1 = Not at all important 2 = Slightly important 3 = Moderately Important 4 = Very Important 5 = Extremely Important			
3. In your opinion are the local district expectations for student proficiency with Algebra I	1.92	1.97	2.02
Scale: 1 = Too low 2 = About right 3 = Too high			
4a. How well do you feel your preservice teacher education program prepared you to teach Algebra I?	1.94	2.06	2.17

Continued on p. 9-70

Appendix D, continued

Item	Low 95% CI	Mean	High 95% CI
4b. How well do you feel your professional development opportunities have helped you to teach Algebra I?	1.96	2.05	2.14
Scale: 1 = Very well 2 = Adequately 3 = Less than adequately 4 = Very poorly			
5. Does your district have teachers at the K–8 level who are mathematics specialists (even if they are called something else)?	0.36	0.45	0.55
a. Do these teachers work with classes of students?	0.51	0.63	0.74
b. Do these teachers provide support to other teachers?	0.76	0.84	0.93
c. Are these teachers specifically qualified or trained to be mathematics specialists?	0.51	0.70	0.88
Scale = Proportion			
6. Are students required to pass Algebra I in order to graduate high school in your district?	0.85	0.88	0.92
Scale = Proportion			
7. How do you rate the state or local school district mathematics standards and math tests that they currently use for Algebra I?			
a. Content standards for Algebra I	2.05	2.17	2.29
b. Assessments of Algebra I outcomes	2.39	2.52	2.64
Scale: 1 = Excellent 2 = Good 3 = Fair 4 = Poor			
Section 4: Teacher Background			
1. What is your sex?	0.60	0.66	0.72
Scale = Proportion Female			

Continued on p. 9-71

Appendix D, continued

Item	Low 95% CI	Mean	High 95% CI
2. Are you Hispanic or Latino?	0.04	0.06	0.08
Scale = Proportion			
3. Which of the following best describes your Hispanic origin or descent?			
Mexican/a or Chicano/a	0.50	0.50	0.50
Puerto Rican	0.05	0.05	0.05
Cuban	0.08	0.08	0.08
Other Hispanic	0.18	0.18	0.18
Scale = Proportion	0.83	0.83	0.83
4. What is your racial background?			
American Indian or Alaska Native	0.00	0.02	0.04
Native Hawaiian or other Pacific Islander	0.00	0.00	0.01
Asian	0.01	0.03	0.04
Black or African-American	0.01	0.04	0.06
White	0.88	0.91	0.94
Scale = Proportion			
5. What is your age?	39.46	41.11	42.75
Scale = Age			
6. What is your employment status in this school system?			
Regular full-time teacher	0.94	0.97	0.99
Regular part-time teacher	0.00	0.02	0.04

Continued on p. 9-72

Appendix D, continued

Item	**Low 95% CI**	**Mean**	**High 95% CI**
Long-term substitute teacher	0.00	0.01	0.02
Other			
Scale = Proportion			
7. Counting this year how many years in total have you taught at either the elementary or secondary level? Please also note the number of years in total.			
a. Elementary (K–6)	1.06	2.07	3.08
b. Secondary (7–12)	10.99	12.15	13.31
c. Total (K–12)	11.51	12.77	14.02
Scale = Number of Years			
8. Counting this year how many years in total have you taught in this school?	6.93	8.00	9.08
Scale = Number of Years			
9. How many years of experience do you have teaching Algebra I?	8.55	9.49	10.44
Scale = Number of Years	1.07	1.15	1.23
10. In which subject area have you taught the most during this school year?			
Math	0.86	0.92	0.97
Science	-0.01	0.05	0.10
English	0.00	0.02	0.04

Continued on p. 9-73

Appendix D, continued

Item	Low 95% CI	Mean	High 95% CI
Social Studies/History	0.00	0.00	0.00
Other	0.00	0.02	0.03
Scale = Proportion			
11. What type of teaching certification do you currently hold?			
Regular or standard state certificate	0.78	0.82	0.87
Probationary certificate	0.01	0.02	0.03
Provisional or temporary certificate	0.07	0.11	0.14
Waiver or emergency certificate	0.00	0.01	0.02
Other	0.02	0.04	0.06
Scale = Proportion			
12. Which of the following best describes your national certification status?			
I have achieved certification by the National Board for Professional Teaching Standards.	0.08	0.12	0.17
I am currently working on National Board certification but have not achieved it.	0.02	0.04	0.06
I am not working on National Board certification.	0.79	0.84	0.88
Scale = Proportion			
13. Under the *No Child Left Behind Law (NCLB)* are you considered to be a highly qualified teacher of:			
a. High school mathematics	0.77	0.83	0.89
b. Middle school mathematics	0.91	0.94	0.98

Continued on p. 9-74

Appendix D, continued

Item	**Low 95% CI**	**Mean**	**High 95% CI**
14. What is the highest academic degree you hold?			
Bachelor's	0.45	0.51	0.57
Master's	0.35	0.41	0.46
Education specialist or professional diploma based on at least one year of work past Master's-degree level	0.04	0.06	0.09
Doctorate	0.00	0.01	0.01
Professional degree (e.g., M.D., L.L.B., J.D., D.D.S.)	0.00	0.01	0.01
Scale = Proportion			
15. In what YEAR did you receive your highest college degree?	1992.16	1993.70	1995.24
Scale = Year			
16. What was your major field of study for your bachelor's degree?			
Education	0.14	0.20	0.25
English	0.00	0.01	0.02
History	0.00	0.02	0.03
Mathematics	0.38	0.44	0.49
Natural/Physical science	0.02	0.07	0.12
Foreign language	0.00	0.00	0.01
Other	0.22	0.27	0.31
Scale = Proportion			

Continued on p. 9-75

Appendix D, continued

Item	Low 95% CI	Mean	High 95% CI
17. What was your minor field of study for your bachelor's degree?			
Education	0.10	0.15	0.20
English	0.00	0.01	0.01
History	0.02	0.06	0.10
Mathematics	0.25	0.33	0.41
Natural/Physical science	0.05	0.10	0.15
Foreign language	0.02	0.05	0.08
Other	0..24	0.30	0.37
Scale = Proportion			
18. If you have earned a graduate degree, what was your major field of study for your highest graduate degree?			
Education	0.43	0.50	0.58
Mathematics	0.09	0.15	0.21
Natural/Physical science	0.00	0.01	0.02
Other	0.26	0.33	0.41
Scale = Proportion			
19. How skillful would you say you are at helping students master Algebra I?	1.27	1.33	1.40
Scale: 1 = Very skillful 2 = Somewhat skillful 3 = Sometimes less skillful than I would like to be 4 = Much Less Skillful than I would like to be			

Continued on p. 9-76

Appendix D, continued

Item	**Low 95% CI**	**Mean**	**High 95% CI**
20. What do you find most challenging in teaching Algebra I successfully?			
Explaining material to struggling students	0.01	0.03	0.05
Handling accelerated students	0.00	0.01	0.03
Teaching procedures	0.00	0.00	0.01
Explaining concepts (e.g., why procedures work, what ideas mean)	0.00	0.04	0.09
Using diagrams or models effectively	0.00	0.01	0.02
Interpreting students' errors and difficulties	-0.01	0.01	0.04
Working with unmotivated students	0.55	0.62	0.68
Working with advanced students	0.00	0.01	0.02
Helping students whose home language is other than Standard English	0.01	0.01	0.02
Making mathematics accessible and comprehensible to all of my students	0.08	0.11	0.15
Other	0.10	0.14	0.17
Scale = Proportion			

APPENDIX E: NSAT Questionnaire

National Survey of Algebra Teachers

Sponsored by:
The U.S. Department of Education
National Mathematics Advisory Panel

Conducted by:
NORC
at the University of Chicago

The National Survey of Algebra Teachers seeks to obtain information from Algebra I teachers about their views on students' preparation, curriculum and instruction.

Participation of teachers is voluntary and no negative consequences will attend a decision not to participate. Responses to this data collection will be used only for statistical purposes. The reports prepared for this study will summarize findings across the sample and will not associate responses with a specific district, school, or individual. We will not provide information that identifies you or your district to anyone outside the study team, except as required by law.

You may use either pen or pencil.
Please clearly circle your answers.
If you need to change an answer, please make sure the old answer is either completely erased or clearly crossed out.

The time required to complete this form varies according to individual circumstances, but the average time is estimated to be 25 minutes. If you have any comments regarding this time estimate, please write to: U.S. Department of Education, The National Mathematics Advisory Panel, Washington, D.C. 20202-4651. If you have any specific questions or comments regarding this study, please contact Lekha Venkataraman of NORC at 1-866-696-4580.

Thank you for taking the time to complete this questionnaire.

OMB No: 1875-0243

Expiration Date: 09/30/2007

Section 1: Your Algebra I Class

In this section of the survey we would like for you to report on ONE specific class, which we will call your Target Class. When you see this referred to in a question, please report on this ONE class, even if it is not typical of the Algebra I classes you teach.

How to determine your Target Class

Your Target Class is the first Algebra I class you teach on Mondays. If you do not teach an Algebra I class on Monday, your Target Class is the first Algebra I class you teach on the following day.

Please answer the following questions regarding your Target Class.

1. How many students are in your Target Class?

1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐
Less than 15 students	15–20 students	21–25 students	26–30 students	31–35 students	More than 35 students

2. How many of the students in your Target Class: *(Please circle one per line)*

		None	Some	About half	Most	Nearly all
2a.	Are in the 7th grade	0	1	2	3	4
2b.	Are in the 8th grade	0	1	2	3	4
2c.	Are in the 9th grade	0	1	2	3	4
2d.	Are in the 10th grade	0	1	2	3	4
2e.	Are in the 11th grade	0	1	2	3	4
2f.	Are in the 12th grade	0	1	2	3	4
2g.	Are in special education (i.e., have an IEP)	0	1	2	3	4
2h.	Are currently enrolled in your school's bilingual program	0	1	2	3	4

3. How many students in your Target Class: *(Please circle one per line)*

		None	Some	About half	Most	Nearly all
3a.	Come to class on time	0	1	2	3	4
3b.	Attend class regularly	0	1	2	3	4
3c.	Come to class prepared with appropriate supplies and books	0	1	2	3	4
3d.	Create serious behavior problems in your class	0	1	2	3	4
3e.	Regularly pay attention in class	0	1	2	3	4
3f.	Actively participate in class activities	0	1	2	3	4
3g.	Take notes	0	1	2	3	4
3h.	Have serious difficulties reading English	0	1	2	3	4
3i.	Care about what grade they receive	0	1	2	3	4

4.	**Based on your experience with in-coming Algebra I students in your Target Class, how would you rate students' background in each of the following areas of mathematics? *(Please circle one per line)***				
		Excellent	**Good**	**Fair**	**Poor**
4a.	Whole numbers and operations with whole numbers	1	2	3	4
4b.	Positive and negative integers and operations with positive and negative integers	1	2	3	4
4c.	Rational numbers and operations involving fractions and decimals	1	2	3	4
4d.	Ratios, percents, rates, and proportions	1	2	3	4
4e.	Solving word problems	1	2	3	4
4f.	The concept of variables	1	2	3	4
4g.	Manipulation of variables	1	2	3	4
4h.	Solving simple linear equations and inequalities	1	2	3	4
4i.	Plotting points, and graphing lines on the four-quadrant coordinate plane	1	2	3	4
4j.	Measurement formulas of basic geometric shapes	1	2	3	4
4k.	Basic study skills and work habits necessary for success in math	1	2	3	4
4l.	Computation skills	1	2	3	4
4m.	Ability to use math in contexts that are identified as real world situations	1	2	3	4
4n.	Working independently	1	2	3	4
4o.	Working cooperatively with other students	1	2	3	4

5.	**On average, how often do you use the following instructional materials and tools in your Target Class? *(Please circle one per line)***					
		Never	**Less than once a week**	**About once a week**	**Several times a week**	**Everyday**
5a.	Textbooks	0	1	2	3	4
5b.	Printed instructional materials other than textbooks	0	1	2	3	4
5c.	Teacher/colleague written instructional materials	0	1	2	3	4
5d.	Graphing calculators (the school's or their own)	0	1	2	3	4
5e.	Physical objects ("manipulatives")	0	1	2	3	4
5f.	Computer-based instructional tools (software)	0	1	2	3	4

6. Please indicate your level of agreement or disagreement with the statement "Computer-based instructional tools (software) are helping Algebra I students in my Target Class." ***(check one)***

1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
Strongly agree	Somewhat agree	Neither agree nor disagree	Somewhat disagree	Strongly disagree

7. What is the name of the textbook you primarily use in your Algebra I Target Class? If you do not use a textbook please write N/A in the space provided.

7a. Title ☐

7b. Author ☐

7c. Publisher ☐

7d. Date of Publication ☐

8. Please indicate your level of agreement or disagreement with each of the following statements regarding the Algebra I textbook you use in your Target Class. ***(Please circle one per line)***

		Strongly Agree	Agree	No Opinion	Disagree	Strongly disagree
8a.	The textbook includes the appropriate topics and content to teach the course.	1	2	3	4	5
8b.	The textbook appropriately sequences math concepts.	1	2	3	4	5
8c.	The textbook provides examples and lessons that are focused directly on the mathematics involved and that explain concepts clearly.	1	2	3	4	5
8d.	The textbook provides opportunities for the development of problem-solving skills.	1	2	3	4	5
8e.	The textbook provides adequate practice for each topic covered.	1	2	3	4	5
8f.	The textbook and the supporting materials which come with it, provide the right mix of useful suggestions and problems for homework assignments.	1	2	3	4	5
8g.	The textbook provides adequate supplementary/support materials.	1	2	3	4	5
8h.	The textbook is clearly focused on Algebra I and contains very few if any distractions to the instructional focus (e.g. off task activities, pictures with no sense of purpose, etc.).	1	2	3	4	5
8i.	The textbook and the accompanying materials provide useful suggestions for meeting the needs of diverse learners.	1	2	3	4	5

9. About what percentage of your current Algebra I students in your Target Class do you anticipate will fail your course? *(check one)*

1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐	8 ☐
None	1–10 %	11–20%	21–30%	31–40%	41–50%	More than 50%	No answer

10. On average, about how much time per day do you expect your Algebra I students in your Target Class to spend on assignments outside of class? *(check one)*

1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
None	1–15 mins	16–30 mins	31–45 mins	46–60 mins	More than 60 mins	No answer

11. On average, about how many of your Algebra I students in your Target Class complete their outside-of-class assignments? *(check one)*

1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
All or almost all	About two-thirds	About one-third	None or almost none	Not applicable/ no homework

12. On average how many minutes per week does your Algebra I Target Class meet?

☐ ***(FILL IN MINUTES)***

13. Does your Algebra I Target Class meet everyday?

1 ☐ Yes 2 ☐ No

14. How long is each period during which you teach Algebra I?

☐ ***(FILL IN MINUTES)***

15. Is this enough instructional time to adequately teach Algebra I to your Target Class?

1 ☐ Yes 2 ☐ No

Section 2: Your School and Algebra I

1. Below is a list of factors that may cause problems in Algebra I instruction. For each factor, please indicate whether it is not a problem, a minor problem, a moderate problem or a serious problem in your school. ***(Please circle one per line)***

		Not a problem	Minor problem	Moderate problem	Serious problem
1a.	Insufficient access to computers	1	2	3	4
1b.	Inadequate access to graphing calculators	1	2	3	4
1c.	Poor quality or out-of-date textbooks	1	2	3	4
1d.	Class sizes are too large	1	2	3	4
1e.	Too little coordination or articulation between classes in the mathematics curriculum	1	2	3	4
1f.	Some teachers are inadequately prepared to teach Algebra I	1	2	3	4
1g.	Lack of teacher planning time	1	2	3	4
1h.	Students with different abilities and interests taking the same math classes	1	2	3	4
1i.	Too little parent/family support	1	2	3	4
1j.	Inadequate opportunities for professional learning	1	2	3	4
1k.	Inadequate administrative support	1	2	3	4

2. Does your school offer different levels of Algebra I to groups of students based on ability?

1 ☐ Yes 2 ☐ No 3 ☐ Don't know

3. How many CLASS PERIODS do you teach a WEEK? (Exclude study halls and homeroom periods.)

☐ ***(Please enter a number)***

4. Is your Algebra I class part of block scheduling at your school?

1 ☐ Yes 2 ☐ No

5. On average, how many minutes are you scheduled during the school day to prepare for classes?

[ ] ***(FILL IN MINUTES)***

6. On average how much time do you spend outside of the regular school day preparing for your Algebra I classes?

[ ] ***(FILL IN MINUTES)***

7. To what grades are you currently teaching Algebra I? *(Check all that apply)*

1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
7th grade	8th grade	9th grade	10th grade	11th grade	12th grade	Special Education

8. How do you rate the remedial help in your school for students who are struggling in Algebra I? *(Please circle one per line)*

	Excellent	Good	Fair	Poor
8a. Availability of tutoring or other remedial assistance	1	2	3	4
8b. Quality of tutoring or other remedial assistance	1	2	3	4

Section 3: Your Views of Mathematics Education

1. How important is a solid foundation in each of the following areas to students' success in Algebra I? *(Please circle one per line)*

		Not at all important	Slightly important	Moderately important	Very important	Extremely important
1a.	Whole numbers and operations with whole numbers	1	2	3	4	5
1b.	Positive and negative integers and operations with positive and negative integers	1	2	3	4	5
1c.	Rational numbers and operations involving fractions and decimals	1	2	3	4	5
1d.	Ratios, percents, rates, and proportions	1	2	3	4	5
1e.	Solving problems involving whole numbers, fractions, and decimals	1	2	3	4	5
1f.	The concept of variables	1	2	3	4	5
1g.	Manipulation of variables	1	2	3	4	5
1h.	Solving simple linear equations and inequalities	1	2	3	4	5
1i.	Plotting points, and graphing lines on the four-quadrant coordinate plane	1	2	3	4	5
1j.	Measurement formulas of basic geometric shapes	1	2	3	4	5
1k.	Basic study skills and work habits necessary for success in math	1	2	3	4	5
1l.	Computation skills	1	2	3	4	5
1m.	Ability to use math in contexts that are identified as real world situations	1	2	3	4	5
1n.	Working independently	1	2	3	4	5
1o.	Working cooperatively with other students	1	2	3	4	5

2. Please provide a brief description of any changes you would like to see in the curriculum leading up to Algebra I in your district.

3. In your opinion, are the local district expectations for student proficiency with Algebra I: *(Please check one)*

1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
Too low	About right	Too high	I do not know the expectations	There are no district expectations

4a. How well do you feel your preservice teacher education program prepared you to teach Algebra I?

1 ☐	2 ☐	3 ☐	4 ☐
Very well	Adequately	Less than adequately	Very poorly

4b. How well do you feel your professional development opportunities have helped you to teach Algebra I?

1 ☐	2 ☐	3 ☐	4 ☐
Very well	Adequately	Less than adequately	Very poorly

5. Does your district have teachers at the K–8 level who are "mathematics specialists" (even if they are called something else)?

1 ☐ Yes ↓ 2 ☐ No ➔ **skip to question 6** 3 ☐ Not sure ➔ **skip to question 6**

		Yes	No	Not Sure
5a.	Do these teachers work with classes of students?	1	2	3
5b.	Do these teachers provide support to other teachers?	1	2	3
5c.	Are these teachers specifically qualified or trained to be mathematics specialists?	1	2	3

6. **Are students required to pass Algebra I in order to graduate high school in your district?**

1 ☐ Yes 2 ☐ No 3 ☐ Don't know

7. **How do you rate the state or local school district mathematics standards and math tests that they currently use for Algebra I? *(Please circle one per line)***

		Excellent	Good	Fair	Poor	Not applicable—no standards defined
7a.	Content standards for Algebra I	1	2	3	4	5
7b.	Assessments of Algebra I outcomes	1	2	3	4	5

Section 4: Teacher Background

1. What is your sex?

1 ☐ Male   2 ☐ Female

2. Are you Hispanic or Latino?

1 ☐ Yes ➔ **If Yes, answer question 3**

2 ☐ No ➔ **If No, skip to question 4**

3. Which of the following best describes your Hispanic origin or descent? *(Please check all that apply)*

1 ☐ Mexican/a or Chicano/a

2 ☐ Puerto Rican

3 ☐ Cuban

4 ☐ Other Hispanic, *Specify* ____________

4. What is your racial background? *(Please check all that apply)*

1 ☐ American Indian or Alaska Native

2 ☐ Native Hawaiian or other Pacific Islander

3 ☐ Asian

4 ☐ Black or African American

5 ☐ White

5. What is your age?

☐ ***(FILL IN AGE)***

6. What is your employment status in this school system?

1 ☐ Regular full-time teacher

2 ☐ Regular part-time teacher

3 ☐ Long-term substitute teacher

4 ☐ Other, *Specify* ____________

7. Counting this year how many years in total have you taught at either the elementary or secondary level? Please also note the number of years in total.

7a.	Elementary (K–6)	☐	***Number of Years***
7b.	Secondary (7–12)	☐	***Number of Years***
7c.	Total (K–12)	☐	***Number of Years***

8. Counting this year, how many years in total have you taught in this school?

☐ ***Number of Years***

9. How many years of experience do you have teaching Algebra I?

☐ ***Number of Years***

10. In which subject area have you taught the most during this school year?

1 ☐	2 ☐	3 ☐	4 ☐	5 ☐
Math	Science	English	Social Studies/ History	Other, *please specify* ________

11. What type of teaching certification do you currently hold?

1 ☐ Regular or standard state certificate
2 ☐ Probationary certificate
3 ☐ Provisional or temporary certificate
4 ☐ Waiver or emergency certificate
5 ☐ Other, *please specify* ____

12. Which of the following best describes your national certification status? *(Check one)*

1 ☐ I have achieved certification by the National Board for Professional Teaching Standards.
2 ☐ I am currently working on National Board Certification but have not achieved it.
3 ☐ I am not working on National Board Certification.

13. Under the No Child Left Behind Law (NCLB) are you considered to be a "highly qualified" teacher of:

		Yes	No	Not Applicable
13a.	high school mathematics	1	2	3
13b.	middle school mathematics	1	2	3

14. What is the highest academic degree you hold?

1 ☐ Less than a Bachelor's degree
2 ☐ Bachelor's
3 ☐ Master's
4 ☐ Education specialist or professional diploma based on at least one year of work past Master's degree level
5 ☐ Doctorate
6 ☐ Professional degree (e.g., M.D. L.L.B., J.D., D.D.S.)

15. In what YEAR did you receive your highest college degree?

☐ ***YYYY***

16. What was your major field of study for your bachelor's degree?

1 ☐ Education
2 ☐ English
3 ☐ History
4 ☐ Mathematics
5 ☐ Natural/Physical science
6 ☐ Foreign language
7 ☐ *Other specify:* ________________________

17. What was your minor field of study for your bachelor's degree?

1 ☐ Education
2 ☐ English
3 ☐ History
4 ☐ Mathematics
5 ☐ Natural/Physical science
6 ☐ Foreign language
7 ☐ *Other specify:* ________________________
8 ☐ Not applicable

18. If you have earned a graduate degree, what was your major field of study for your highest graduate degree?

1 ☐ Education
2 ☐ English
3 ☐ History
4 ☐ Mathematics
5 ☐ Natural/Physical science
6 ☐ Foreign language
7 ☐ *Other specify:* ________________________
8 ☐ Not applicable

19. How skillful would you say you are at helping students master Algebra I?

1 ☐ Very skillful
2 ☐ Somewhat skillful
3 ☐ Sometimes less skillful than I would like to be
4 ☐ Much less skillful than I would like to be

20. What do you find most challenging in teaching Algebra I successfully? *(Please check one)*

1 ☐ Explaining material to struggling students
2 ☐ Handling accelerated students
3 ☐ Teaching procedures
4 ☐ Explaining concepts (e.g., why procedures work, what ideas mean)
5 ☐ Using diagrams or models effectively
6 ☐ Interpreting students' errors and difficulties
7 ☐ Working with unmotivated students
8 ☐ Working with advanced students
9 ☐ Helping students whose home language is other than Standard English
10 ☐ Making mathematics accessible and comprehensible to all of my students
11 ☐ Other, please specify: ____________________

Thank you!

APPENDIX A: Presidential Executive Order 13398

Federal Register / Vol. 71, No. 77 / Friday, April 21, 2006 / Presidential Documents 20519

Presidential Documents

Executive Order 13398 of April 18, 2006

National Mathematics Advisory Panel

By the authority vested in me as President by the Constitution and the laws of the United States of America, it is hereby ordered as follows:

Section 1. *Policy.* To help keep America competitive, support American talent and creativity, encourage innovation throughout the American economy, and help State, local, territorial, and tribal governments give the Nation's children and youth the education they need to succeed, it shall be the policy of the United States to foster greater knowledge of and improved performance in mathematics among American students.

Sec. 2. *Establishment and Mission of Panel.* (a) There is hereby established within the Department of Education (Department) the National Mathematics Advisory Panel (Panel).

(b) The Panel shall advise the President and the Secretary of Education (Secretary) consistent with this order on means to implement effectively the policy set forth in section 1, including with respect to the conduct, evaluation, and effective use of the results of research relating to proven-effective and evidence-based mathematics instruction.

Sec. 3. *Membership and Chair of Panel.* (a) The Panel shall consist of no more than 30 members as follows:

(i) no more than 20 members from among individuals not employed by the Federal Government, appointed by the Secretary for such terms as the Secretary may specify at the time of appointment; and

(ii) no more than 10 members from among officers and employees of Federal agencies, designated by the Secretary after consultation with the heads of the agencies concerned.

(b) From among the members appointed under paragraph(3)(a)(i) of this order, the Secretary shall designate a Chair of the Panel.

(c) Subject to the direction of the Secretary, the Chair of the Panel shall convene and preside at meetings of the Panel, determine its agenda, direct its work and, as appropriate to deal with particular subject matters, establish and direct the work of subgroups of the Panel that shall consist exclusively of members of the Panel.

Sec. 4. *Report to the President on Strengthening Mathematics Education.* In carrying out subsection 2(b) of this order, the Panel shall submit to the President, through the Secretary, a preliminary report not later than January 31, 2007, and a final report not later than February 28, 2008. Both reports shall, at a minimum, contain recommendations, based on the best available scientific evidence, on the following:

(a) the critical skills and skill progressions for students to acquire competence in algebra and readiness for higher levels of mathematics;

(b) the role and appropriate design of standards and assessment in promoting mathematical competence;

(c) the processes by which students of various abilities and backgrounds learn mathematics;

(d) instructional practices, programs, and materials that are effective for improving mathematics learning;

20520 Federal Register / Vol. 71, No. 77 / Friday, April 21, 2006 / Presidential Documents

(e) the training, selection, placement, and professional development of teachers of mathematics in order to enhance students' learning of mathematics;

(f) the role and appropriate design of systems for delivering instruction in mathematics that combine the different elements of learning processes, curricula, instruction, teacher training and support, and standards, assessments, and accountability;

(g) needs for research in support of mathematics education;

(h) ideas for strengthening capabilities to teach children and youth basic mathematics, geometry, algebra, and calculus and other mathematical disciplines;

(i) such other matters relating to mathematics education as the Panel deems appropriate; and

(j) such other matters relating to mathematics education as the Secretary may require.

Sec. 5. *Additional Reports.* The Secretary may require the Panel, in carrying out subsection 2(b) of this order, to submit such additional reports relating to the policy set forth in section 1 as the Secretary deems appropriate.

Sec. 6. *General Provisions.* (a) This order shall be implemented in a manner consistent with applicable law, including section 103 of the Department of Education Organization Act (20 U.S.C. 3403), and subject to the availability of appropriations.

(b) The Department shall provide such administrative support and funding for the Panel as the Secretary determines appropriate. To the extent permitted by law, and where practicable, agencies shall, upon request by the Secretary, provide assistance to the Panel.

(c) The Panel shall obtain information and advice as appropriate in the course of its work from:

(i) officers or employees of Federal agencies, unless otherwise directed by the head of the agency concerned;

(ii) State, local, territorial, and tribal officials;

(iii) experts on matters relating to the policy set forth in section 1;

(iv) parents and teachers; and

(v) such other individuals as the Panel deems appropriate or as the Secretary may direct.

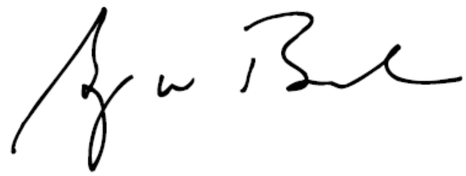
(d) Members of the Panel who are not officers or employees of the United States shall serve without compensation and may receive travel expenses, including per diem in lieu of subsistence, as authorized by law for persons serving intermittently in Government service (5 U.S.C. 5701–5707), consistent with the availability of funds.

(e) Insofar as the Federal Advisory Committee Act, as amended (5 U.S.C. App.) (the "Act"), may apply to the administration of any portion of this order, any functions of the President under that Act, except that of reporting to the Congress, shall be performed by the Secretary in accordance with the guidelines issued by the Administrator of General Services.

(f) This order is not intended to, and does not, create any right or benefit, substantive or procedural, enforceable by any party at law or in equity against the United States, its departments, agencies, entities, officers, employees, or agents, or any other person.

Federal Register / Vol. 71, No. 77 / Friday, April 21, 2006 / Presidential Documents 20521

Sec. 7. *Termination.* Unless hereafter extended by the President, this Advisory Panel shall terminate 2 years after the date of this order.



THE WHITE HOUSE,
April 18, 2006.

[FR Doc. 06–3865
Filed 4–20–06; 8:45 am]
Billing code 3195–01–P

APPENDIX B: Rosters of Panel Members, U.S. Department of Education Staff, and Consultants

Panelists

- Larry R. Faulkner (Chair), President, Houston Endowment Inc.; President Emeritus, University of Texas at Austin
- Camilla Persson Benbow (Vice Chair), Patricia and Rodes Hart Dean of Education and Human Development, Peabody College, Vanderbilt University
- Deborah Loewenberg Ball, Dean, School of Education and William H. Payne Collegiate Professor, University of Michigan
- A. Wade Boykin, Professor and Director of the Graduate Program, Department of Psychology, Howard University
- Douglas H. Clements, Professor, Graduate School of Education, University at Buffalo, State University of New York (Began with the Panel March 19, 2007)
- Susan Embretson, Professor, School of Psychology, Georgia Institute of Technology (Began with the Panel March 19, 2007)
- Francis "Skip" Fennell, Professor of Education, McDaniel College
- Bert Fristedt, Morse-Alumni Distinguished Teaching Professor of Mathematics, University of Minnesota, Twin Cities (Began with the Panel March 19, 2007)
- David C. Geary, Curators' Professor, Department of Psychological Sciences, University of Missouri
- Russell M. Gersten, Executive Director, Instructional Research Group; Professor Emeritus, College of Education, University of Oregon
- Nancy Ichinaga, Former Principal, Bennett-Kew Elementary School, Inglewood, California (Served with the Panel through May 29, 2007)
- Tom Loveless, The Herman and George R. Brown Chair, Senior Fellow, Governance Studies, The Brookings Institution
- Liping Ma, Senior Scholar, The Carnegie Foundation for the Advancement of Teaching
- Valerie F. Reyna, Professor of Human Development, Professor of Psychology, and Co-Director, Center for Behavioral Economics and Decision Research, Cornell University

- Wilfried Schmid, Dwight Parker Robinson Professor of Mathematics, Harvard University

- Robert S. Siegler, Teresa Heinz Professor of Cognitive Psychology, Carnegie Mellon University

- James H. Simons, President, Renaissance Technologies Corporation; Former Chairman, Mathematics Department, State University of New York at Stony Brook

- Sandra Stotsky, Twenty-First Century Chair in Teacher Quality, University of Arkansas; Member, Massachusetts State Board of Education

- Vern Williams, Mathematics Teacher, Longfellow Middle School, Fairfax County Public Schools, Virginia

- Hung-Hsi Wu, Professor of Mathematics, University of California at Berkeley

Ex Officio Members

- Irma Arispe, Assistant Director for Life Sciences and Acting Assistant Director for Social and Behavioral Sciences, Office of Science and Technology Policy, Executive Office of the President (Began with the Panel May 30, 2007)

- Daniel B. Berch, Associate Chief, Child Development and Behavior Branch and Director, Mathematics and Science Cognition and Learning Program, National Institute of Child Health and Human Development, National Institutes of Health

- Joan Ferrini-Mundy, Division Director, Division of Research on Learning in Formal and Informal Settings, National Science Foundation (On an Intergovernmental Personnel Act Assignment from Michigan State University. Began with the Panel January 16, 2007)

- Diane Auer Jones, Deputy to the Associate Director for Science, White House Office of Science and Technology Policy (Served with the Panel through May 23, 2007)

- Thomas W. Luce, III, Assistant Secretary for Planning, Evaluation, and Policy Development, U.S. Department of Education (Served with the Panel through November 1, 2006)

- Kathie L. Olsen, Deputy Director, National Science Foundation, (Served with the Panel through January 11, 2007)

- Raymond Simon, Deputy Secretary, U.S. Department of Education

- Grover J. “Russ” Whitehurst, Director, Institute of Education Sciences, U.S. Department of Education

U.S. Department of Education Staff

- Tyrrell Flawn, Executive Director, National Mathematics Advisory Panel, U.S. Department of Education
- Ida Eblinger Kelley, Special Assistant, National Mathematics Advisory Panel, U.S. Department of Education
- Jennifer Graban, Deputy Director for Research and External Affairs, National Mathematics Advisory Panel, U.S. Department of Education
- Marian Banfield, Deputy Director of Programs and Special Projects, National Mathematics Advisory Panel, U.S. Department of Education

Additional support was provided by the following: Anya Smith, Director of Special Events and the Events Team, Office of Communications and Outreach, U.S. Department of Education; Holly Clark, Management and Program Analyst, Office of Innovation and Improvement, U.S. Department of Education; Mike Kestner, Math and Science Partnership Program, Office of Elementary and Secondary Education, U.S. Department of Education; Kenneth Thomson, Presidential Management Fellow, Office of Planning, Evaluation, and Policy Development, U.S. Department of Education; and Jim Yun, Math and Science Partnership Program, Office of Elementary and Secondary Education, U.S. Department of Education.

Consultants

- Alina Martinez, Abt Associates, Inc., Project Director
- Ellen Bobronnikov, Abt Associates, Inc.
- Fran E. O'Reilly, Abt Associates, Inc.
- Mark Lipsey, Vanderbilt University
- Pamela Flattau, Institute for Defense Analyses Science and Technology Policy Institute, Project Director
- Nyema Mitchell, Institute for Defense Analyses Science and Technology Policy Institute
- Kay Sullivan, Institute for Defense Analyses Science and Technology Policy Institute
- Jason Smith, Widmeyer Communications, Project Director

- Sara Appleyard, Widmeyer Communications
- Phyllis Blaunstein, Widmeyer Communications
- Alix Clyburn, Widmeyer Communications
- Jessica Love, Widmeyer Communications

Other Contractual Support

Laurie Bozzi
M.C. Bradley
Chris Chapman
Michael D. Coyne
Phil Daro
Steven Michael Estrada
Kristin Flannagan
Karen Fuson
Steven A. Hecht
Mary K. Hoard
Jeremy Luallen
Elias Morrel-Samuels
Natalia Rojkovskaia
Alyssa Rulf Fountain
Lauren Sher
Sarah Y. Siegel
Fran Stancavage
Jennifer Wiener
Anne Wolf