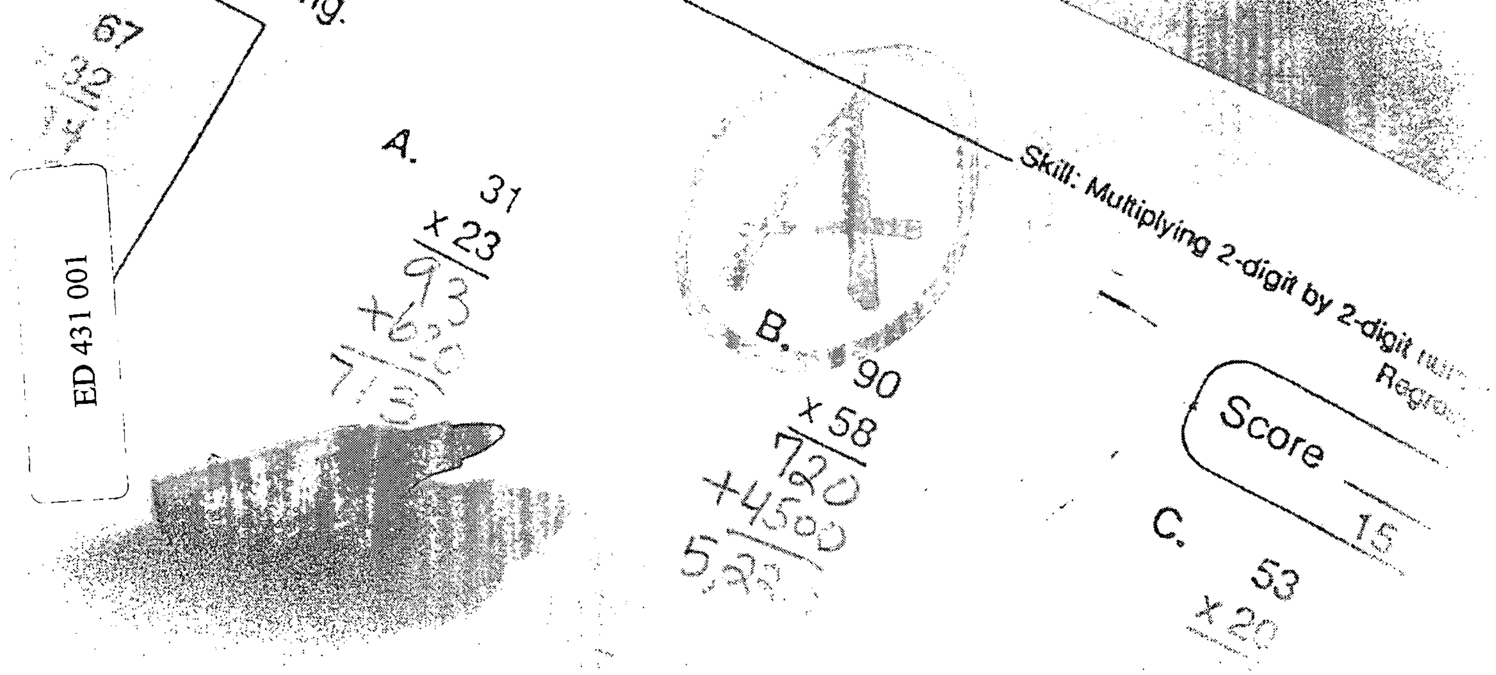
DOCUMENT RESUME

ED 431 001 TM 029 827

AUTHOR	Wilson, Linda Dager; Blank, Rolf K.
TITLE	Improving Mathematics Education Using Results from NAEP and TIMSS.
INSTITUTION	Council of Chief State School Officers, Washington, DC. State Education Assessment Center.
SPONS AGENCY	National Science Foundation, Arlington, VA.
ISBN	ISBN-1-884037-56-9
PUB DATE	1999-00-00
NOTE	54p.
CONTRACT	REC-98-03080
AVAILABLE FROM	Council of Chief State School Officers, Attn: Publications, One Massachusetts Avenue NW, Suite 700, Washington, DC 20001-1431; Tel: 202-336-7016; Web site: http://www.ccsso.org ($8, including shipping and handling).
PUB TYPE	Reports - Evaluative (142)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	Achievement Gains; *Educational Improvement; Educational Policy; *Educational Practices; Elementary Secondary Education; Foreign Countries; International Studies; *Mathematics Achievement; *Mathematics Education; National Surveys; *Test Results
IDENTIFIERS	*National Assessment of Educational Progress; State Mathematics Assessments; *Third International Mathematics and Science Study

ABSTRACT

The findings of the National Assessment of Educational Progress (NAEP) and the Third International Mathematics and Science Study (TIMSS) are approached from the perspective of mathematics educators and education decision makers at state and local levels. Some key findings are highlighted to show problems and successes in mathematics teaching and learning, and some educational practices and policies that appear to improve student performance are identified. One half of the states improved student NAEP mathematics scores from 1990 to 1996. Students improved their performances in the algebra content areas and in geometry, and while they scored well on whole number computation and operations, they were weak in the area of number sense and estimation. Measurement is a particular weakness in U.S. mathematics at all grade levels. Well-prepared teachers appear to make a difference in student achievement. The NAEP, TIMSS, and state assessment programs have important differences in their purposes, frameworks, types of items, and methods of reporting results. Based on the analyses of NAEP and TIMSS results, recommendations are offered for educators and decision makers. An appendix contains a map of some NAEP questions for grade 8. (Contains 29 figures and 59 references.) (SLD)



Improving Mathematics Education Using Results from NAEP and TIMSS

LINDA DAGER WILSON AND ROLF K. BLANK

COUNCIL OF CHIEF STATE SCHOOL OFFICERS · ONE MASSACHUSETTS AVENUE, NW, SUITE 700 · WASHINGTON, DC 20001-1431

COUNCIL OF CHIEF STATE SCHOOL OFFICERS

STATE EDUCATION ASSESSMENT CENTER

Improving Mathematics Education Using Results from NAEP and TIMSS

LINDA DAGER WILSON AND ROLF K. BLANK

Linda Dager Wilson is a professor of mathematics education at the University of Delaware. Rolf K. Blank is director of education indicators at the Council. *The development of this paper was supported by a grant to CCSSO from the National Science Foundation (REC 98-03080).*

THE COUNCIL OF CHIEF STATE SCHOOL OFFICERS

The Council of Chief State School Officers (CCSSO) is a nationwide, nonprofit organization composed of the public officials who head departments of elementary and secondary education in the states, the District of Columbia, the Department of Defense Education Activity, and five extra-state jurisdictions. CCSSO seeks its members' consensus on major educational issues and expresses their view to civic and professional organizations, federal agencies, Congress, and the public. Through its structure of standing and special committees, the Council responds to a broad range of concerns about education and provides leadership on major education issues.

Because the Council represents the chief education administrators, it has access to the educational and governmental establishment in each state and to the national influence that accompanies this unique position. CCSSO forms coalitions with many other education organizations and is able to provide leadership for a variety of policy concerns that affect elementary and secondary education. Thus, CCSSO members are able to act cooperatively on matters vital to the education of America's young people.

The State Education Assessment Center was established by chief state school officers to improve the information base on education in the United States, especially from a state perspective. The Center works to improve the breadth, quality, and comparability of data on education, including state-by-state achievement data, instructional data, indicators of quality in areas such as mathematics and science, and performance assessment of teachers and students. In collaboration with state education agencies, the federal government, and national and international organizations, the Center contributes to the development of a set of useful and valid measures of educational quality geared, when appropriate, to education standards.

Copies of this paper may be ordered for $8.00, including shipping and handling, from:

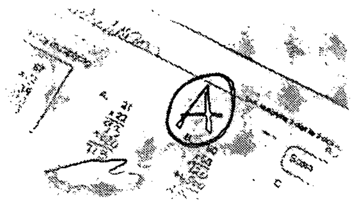
Council of Chief State School Officers
Attn: Publications
One Massachusetts Avenue NW
Suite 700
Washington, DC 20001-1431

Phone: (202) 336-7016
Fax: (202) 408-8072
www.ccsso.org

ISBN # 1-884037-56-9

TABLE OF CONTENTS

LIST OF FIGURES



LIST OF FIGURES

(CONTINUED)

Improving Mathematics Education Using Results from NAEP and TIMSS

LINDA DAGER WILSON AND ROLF K. BLANK

Recent results of national and international studies have focused the attention of educators, education policymakers and the public on the condition of mathematics and science education in our nation's schools. The 1997 findings from the Third International Mathematics and Science Study (TIMSS) provide the most comprehensive, in-depth information to date on the achievement of students in U.S. schools as compared to countries around the world. Also in 1997 the U.S. Department of Education released the results of the National Assessment of Educational Progress (NAEP) in mathematics that provided the first analysis of trends in mathematics learning in the 1990's state-by-state. With these two major studies, mathematics and science educators have available a wealth of information about the knowledge and performance of students as well as details about the characteristics of their teachers and schools.

Well-publicized reports and press releases from the sponsoring agencies of these studies have focused attention on U.S. students' relatively good performance in mathematics at the elementary level and the apparent decline in math proficiency as students move into middle school and high school. Mathematics curriculum in the U.S. was reported by TIMSS to be very broad in relation to other countries, covering too many topics at each grade with too little depth. Many of the headlines about NAEP mathematics results have focused on most students not meeting expected levels of achievement, only small improvements in student achievement over time, and widely differing achievement results from state to state.

Analyses and interpretation of these major studies are just beginning to be more generally available. Mathematics educators are unlikely to have sufficient useful information from the NAEP and TIMSS analyses so far to guide their efforts toward improvement of teaching and curriculum. In this paper, we approach the NAEP and TIMSS results from the perspective of mathematics educators and education decision-makers at state and local levels. We highlight some of the key findings that show problems and successes in mathematics teaching and learning in schools, and we pinpoint some of the educational practices and policies that appear to improve student performance.

Before presenting our analyses, what have been the main themes that have predominated the discussion of recent NAEP and TIMSS results for mathematics? What are the messages that many educators and the public are likely to have received, so far, about the performance of our students?

"... Most students are not meeting defined current national levels of 'proficiency' in mathematics on NAEP, and long-term trends in mathematics performance show little improvement."

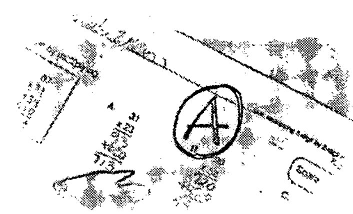
"... In the TIMSS mathematics results, U.S. students ranked high in 4th grade, below average in 8th grade, and almost last at grade 12."

"... Students in Midwest and Northeast states did much better on NAEP mathematics than other states because these states have fewer low-income students; NAEP scores appear to be more related to social and economic differences between students than differences in school quality, curriculum, or teaching."

"... Results on NAEP and TIMSS show U.S. students doing worse in mathematics than do the results from other large-scale assessments in mathematics, such as college entrance tests or standardized tests used in state and local testing programs."

The initial conclusions that readers draw from NAEP and TIMSS results may lead them to question the usefulness of the findings because the results appear to be based on quite different expectations or standards for mathematics education than those they know. Or, the findings may not appear to be useful because the results are reported in the aggregate at national or state levels and they are not reported with sample items, analysis of performance by content areas, or supporting data on teaching practices and curriculum. Thus, NAEP and TIMSS results often do not appear to be a resource for analyzing how mathematics education can be improved or for identifying policies or practices leading to higher student achievement. In fact, reports and supporting data on NAEP and TIMSS provide very useful materials for teachers, parents, policy makers, and administrators who are looking for assistance in raising the achievement level of students in mathematics. In this paper, we look more deeply at the performance of our students in these studies, we analyze areas of strength and weakness in U.S. mathematics performance on NAEP and TIMSS, and we identify important school and classroom factors that are related to higher achievement of U.S. students in mathematics.

Using NAEP and TIMSS Results: Key Findings



Some of the key findings follow from our detailed analysis of the recent results from NAEP and TIMSS as they apply to mathematics education in U.S. public schools:

- **One-half of states improved student scores on NAEP Mathematics from 1990 to 1996.**

The 1996 NAEP results reveal trends in the progress of mathematics for each participating state. In 1990, NAEP began reporting results at the state level and the assessment was changed to increase the focus on problem solving and to require more open-ended mathematics exercises. Only 15 percent of grade 8 students scored at the Proficient level in 1990. From 1990 to 1996, 27 states significantly improved the proportion of their students scoring at the Proficient level, and three states improved by 11 to 12 percentage points—Michigan, North Carolina, and Minnesota. In 1996, although mathematics progress was made in many states, other states did not improve, and as a nation only one fourth of grade 8 students reached the Proficient level.

- **U.S. students at grades 4 and 8 improved their performance in the algebra content area, and grade 8 students improved in geometry; our students scored well on whole number computation and operations, but they were weak in the area of number sense and estimation.**

It is not sufficient to only report that most students in grade 8 did not reach the Proficient level, or that students at grade 4 are above the international average or twelfth graders are below the international average. There are relative strengths and weaknesses within and between different content areas of mathematics, and careful analysis of the results can help to focus improvement efforts. For example, on TIMSS our fourth graders scored above the international average on questions involving patterns, relations and functions, and our eighth graders were right at the international average in this area. Relative to other content areas, NAEP results show improved performance in algebra since 1990 at all grade levels.

8

In the numbers area, our students did well on questions involving fundamental concepts of numbers, relationships between numbers, and properties of numbers, as well as in skills required for manipulating numbers and completing computations. Our students did poorly on questions requiring multi-step solutions, using new concepts, or applying number sense to new or unusual situations.

Measurement is a particular weakness in mathematics for U.S. students at all grade levels.

U.S. students scored below the international average in this content area at grades 4 and 8, and NAEP measurement scores were lower than the overall math averages at grades 8 and 12. The questions that were particularly difficult for students were those requiring unit conversions, calculations of volume and circumference, and estimation of measurements. We find that students are given too few opportunities to actually engage in the use of measuring instruments. Instead they are shown pictures of objects and the instruments chosen to measure some attribute of those objects. Students should have more hands-on experiences with measuring, including making decisions about which instrument might be appropriate for measuring a particular attribute. Emphasis should be placed on understanding the underlying concepts, rather than simply applying formulas.

Students scored higher on NAEP multiple-choice items than on open-ended items.

At all grade levels, student performance on multiple choice items was significantly better than performance on either regular constructed response items (i.e., open-ended) or extended constructed response items. Constructed response items often assess mathematical reasoning and conceptual understanding and they demand good student communications skills and flexibility to solve non-routine problems. As more teachers use constructed response items in class and students gain more experience in answering them, performance of our students on these items should improve.

Teacher reports on curriculum content reveal many math topics, but little focus.

Results from TIMSS teacher surveys show that Japanese eighth grade teachers spend most of their time teaching a few topics: geometry, congruence and similarity, functions, relations and patterns, and equations and formulas, and these four areas of the curriculum account for approximately 67 percent of teaching time in Japanese classrooms. In contrast, grade 8 teachers in the U.S. spread time very thinly among a wide range of topics. The majority of our teachers cover 16-18 different topics, with only one topic accounting for more than 8 percent of their teaching time.

Well prepared teachers of mathematics make a difference.

The group of states with the highest average scores on NAEP at grade 8 are well above the national average in proportion of teachers with a major or minor in mathematics. The group of states with the lowest NAEP scores are below or near the average level of preparation of their mathematics teachers. Analyses of teacher preparation by school characteristics show that students in high poverty schools and schools with high minority enrollments have higher proportions of under-prepared teachers than other schools.

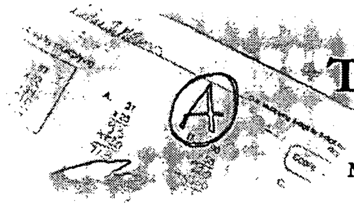
NAEP, TIMSS, and state assessment programs have important differences in their purposes, frameworks, types of items, and methods of reporting results.

Before interpreting the results of NAEP and TIMSS, or any large-scale assessments, it is essential to have an understanding of the purposes, design, and reporting scheme for the assessment. Each of these assessments was built from a different set of purposes and a different framework, and these contexts must be incorporated into any set of conclusions that might be drawn about the results.

Based on the results of our more detailed analyses of NAEP and TIMSS data, our paper offers a set of recommendations for educators and decision-makers. The analyses and interpretations of NAEP and TIMSS results in this paper will help mathematics educators understand specific details about the current quality and effectiveness of mathematics education in our nation's public schools. We try to go beyond the initial level of analysis about the findings, and we try to focus on some of the key findings as they reference results in states. The data we analyzed and the assessment items presented here are from reports released to the public (TIMSS: Beaton, et al., 1996, 1997, 1998; NCES, 1996a, 1997a, 1998a; NAEP: Reese, et al., 1997; Shaughnessy, et al., 1998).

We also address some of the barriers that educators have noted in trying to use the results from NAEP and TIMSS. Although these studies are well-known generally in some levels of the education community, educators and researchers have found difficulties in their use, because of: a) high degree of complexity in how the assessments are conducted, scored, and reported; b) overreliance on composite ratings and rankings of countries and states, c) methods of scoring and reporting NAEP and TIMSS which differ from those used with state and local tests; and d) results derived from state and national samples that do not appear to provide disaggregated data for analyzing issues of concern to teachers and schools. We contend that closer analyses of NAEP and TIMSS results reveal some striking messages that can be of use to all who are interested in improving mathematics education.

In the following sections, we first discuss key findings about mathematics learning from an in-depth examination of NAEP and TIMSS assessment results. We also present several examples of how analyzing variation in performance can be helpful to educators and decision makers. We illustrate the use of measures of opportunity-to-learn mathematics to help explain the wide differences in mathematics proficiency among our schools, classrooms, and states. Finally, to assist in analyzing results, we summarize some of the major differences in the purpose, design, and operation of NAEP, TIMSS, and state assessment programs in mathematics.



Trends in Mathematics Learning

WHAT NAEP AND TIMSS DATA REVEAL ABOUT MATHEMATICS TEACHING AND LEARNING IN THE U.S.

Recent results from NAEP and TIMSS can be valuable resources for in-depth analysis of mathematics education in U.S. schools. To analyze and interpret mathematics assessment results from NAEP and TIMSS it is important to keep in mind key aspects of the purposes and designs of these assessment studies, and how they differ. Some key differences and similarities between NAEP and TIMSS with regards to mathematics assessment are outlined in the table below.

The differences in purpose, design, and structure of NAEP and TIMSS indicate that direct comparison of results is difficult. The linking study of NAEP and TIMSS recently completed by the National Center for Education Statistics (NCES, 1998b) showed that caution should be taken with comparisons from the two studies. The linking study does show where states would have performed if their students had been in TIMSS. Similarities in the assessment frameworks which outline the content for the two studies provide a basis for validity of comparisons between the studies. A summary of the assessment frameworks and distribution of the test items across the frameworks are discussed in part four of the paper (page 34).

GAINS IN MATHEMATICS PROFICIENCY

NAEP mathematics assessment results are expressed as percentages of students reaching each of three "achievement levels," and they are reported using a NAEP scale score. The NAEP scale ranges from 0 to 500, and the same scale is used for reporting scores at grades 4, 8, and 12.

	NAEP	TIMSS
PURPOSE	Regular, periodic assessment and reporting of trends in student learning in schools for the nation and the states in core academic subjects.	Cross-national research and analysis of student achievement, curriculum, and teaching in mathematics and science education with 45 participating countries.
STUDENTS TESTED	National and state representative samples of students and their teachers, based on sampling at the school level.	National representative samples of students and their teachers, based on sampling at the school level.
TEST DEVELOPMENT	Based on NAEP mathematics assessment framework, with new and continuing items provided every four years	Based on TIMSS assessment framework; items written for TIMSS and reviewed by countries
FREQUENCY AND LEVEL (MATH)	NATIONAL: Four years (grades 4, 8, 12) STATE: Four years (grades 4, 8)	Main data collection 1995; by country: grades 4, 8, end of secondary school.
ITEMS	50 percent multiple choice, 50 percent open-ended or constructed response	75 percent multiple choice, 25 percent open-ended.
SCORING AND REPORTING	Three achievement levels and scale score by grade for nation and each state. Report of achievement results for subject, followed by report of supporting data on background and practices.	Scale score for each nation by grade, percent correct in content areas; Reports of student achievement by grade, curriculum analysis, and other studies.

NAEP achievement levels are descriptions of what students should know and be able to do in mathematics at each grade level. Three levels were defined for each grade level, under supervision of the National Assessment Governing Board—Basic, Proficient, and Advanced.

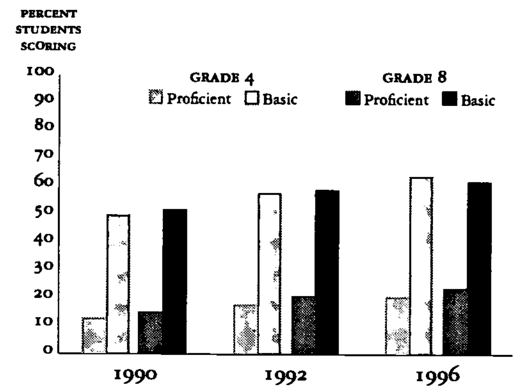
NAEP mathematics scores for 1996 reveal that students in grades 4 and 8 did make achievement gains in proficiency during the 1990s. Although only one quarter of students nationally scored at the Proficient level and just over half scored at the Basic level or higher, performance in mathematics improved. The data from 1996 also showed that results by state differed significantly. About half of the states made real gains in mathematics but the remainder showed little change.

With the 1990 NAEP assessment framework, the test was significantly changed to include open-ended exercises and a broader range of mathematics content. Also, in 1992 NAEP began reporting student results by achievement levels. Figures 1 and 2 provide a graphic summary of key gains since 1990.

- In 1996, 21 percent of grade 4 students performed at or above the Proficient levels and 62 percent of grade 4 students performed at or above the Basic level. Twelve states made significant improvement in the percentage of students at/above the Basic level as compared to 1992, and seven states improved the percentage of students at/above the Proficient level. Gains were largest at grade 4 mathematics in Texas, Indiana, and North Carolina, each improving over 8 to 10 percentage points.

- At grade 8 in 1996, 24 percent of students scored at or above the Proficient level and 62 percent were at/or above the Basic level. From 1990 to 1996, 27 states made significant improvement in the percentage of students at/above Proficient. Michigan, Minnesota, and North Carolina made the largest gains at grade 8, with each improving 11 to 12 percentage points.

FIGURE 1
NATIONAL MATHEMATICS TRENDS
ON MAIN NAEP 1990 TO 1996



Source: NAEP Mathematics Report Card, 1996

FIGURE 2
HIGHEST IMPROVING STATES
AT GRADES 4 AND 8
NAEP 1996

GRADE 4 IMPROVEMENT

State	Proficient Grade 4	Change 1992–96
TEXAS	25%	+10%
INDIANA	24	8
NORTH CAROLINA	21	8
CONNECTICUT	31	7
WEST VIRGINIA	19	7
TENNESSEE	17	7
COLORADO	22	5
NATION	21	+3

GRADE 8 IMPROVEMENT

State	Proficient Grade 8	Change 1990–96
MICHIGAN	28	+12
MINNESOTA	34	11
NORTH CAROLINA	20	11
WISCONSIN	32	9
CONNECTICUT	31	9
COLORADO	25	8
TEXAS	21	8
NATION	23	+9

Source: NAEP Mathematics Report Card, 1996

The data summarized in this paper are from state-by-state NAEP mathematics results for grades 4 and 8 in 1996, including trends since 1990 which are reported in the *NAEP Report Card in Mathematics* (Reese, et al., 1997) and in CCSSO's *State Indicators of Science and Mathematics Education* (Blank, et al., 1997a). Our analysis focuses on gains in mathematics performance for each state based on the NAEP achievement levels, and uses the percentage of students at/above the Proficient level as a key benchmark. The National Education Goals Panel has released a new report with state profiles of student achievement in mathematics and science using NAEP achievement levels and results from the NAEP-TIMSS linking study (1998b).

Figure 3 shows the percenrage of students that attained each of the achievement levels in 1996. Consistent with the scale score results, achievement levels have increased in all three grades since 1990 and 1992. In particular, the percentage of fourth, eighth, and twelfth graders performing at or above the Basic and Proficient levels has increased. However, the percentage of fourth and twelfth grade students achieving the Advanced level has not shown an increase since either 1990 or 1992. Additionally, approximately one third of all students are below the Basic level at all grades.

MATH GAINS IN INTERNATIONAL PERSPECTIVE

The international perspective offered by TIMSS confirms that U.S. mathematics education is still far from the goal of being "first in the world in mathematics and science achievement by the year 2000," as President Bush and 50 governors declared in 1989. Fourth graders scored above the international average on TIMSS, although students in seven countries—Singapore, Korea, Japan, Hong Kong, Netherlands, Czech Republic, and Austria—outperformed U.S. students (NCES, 1997a). U.S. eighth graders scored below the international average of the 41 TIMSS countries (NCES, 1996a).

The United States was the only TIMSS country for which fourth-grade results were above the average and eighth-grade results were below the average (see Figure 4). Eighth graders in Singapore, Korea, and Japan scored more than 100 points higher on the TIMSS scale than eighth graders in the U.S. This is a substantial difference, considering that the difference in performance between grades 7 and 8 is only 26 points in the U.S. (Mullis, 1998). Even for the best U.S. eighth grade students, the news is discouraging—only 5 percent would be included

FIGURE 3
MATHEMATICS ACHIEVEMENT LEVELS AND RESULTS
NAEP 1996

	Grade 4	Grade 8	Grade 12
BASIC Partial mastery of prerequisite knowledge and skills that are fundamental for proficient work at each grade	64%	62%	69%
PROFICIENT Solid academic performance for each grade assessed. Students reaching this level have demonstrated competency over challenging subject matter, including subject-matter knowledge, application of such knowledge to real-world situations, and analytical skills appropriate to the subject matter.	21%	24%	16%
ADVANCED Superior performance	2%	4%	2%

Source: NAEP Mathematics Report Card

in the top 10 percent of all eighth-grade students in the 41 TIMSS countries. For Singapore the corresponding number would be 45 percent.

While the news was not good at eighth grade, it was worse for twelfth grade. U.S. twelfth graders scored below the international average and the U.S. placed among the lowest of the 21 TIMSS nations in mathematics general knowledge in the final year of secondary school. When the eighth grade results are compared with the average of the 20 countries that participated in both eighth and twelfth grade TIMSS, the U.S. scores are similar to the international average. But at twelfth grade only two countries, Cyprus and South Africa, had scores below those of U.S. students. The average U.S. twelfth grade score was 461, while the highest score (Netherlands) was 560. An assessment of advanced mathematics was given to a sample of students taking advanced course work in mathematics. The performance of these advanced students was among the lowest of the 16 countries that participated, for all three content areas assessed.

LONG-TERM NAEP TRENDS

One of the strengths of the NAEP program in the U.S. is the capacity for analyzing trends in learning over a considerable period of time. As a context for examining results from the most recent national and international assessments in the content areas, we can look at what has happened over time in Figure 5. NAEP includes a portion of the assessment that has remained the same since 1973, yielding valuable trend data for those 23 years. At all age levels, the overall trend is of increased performance over time (Campbell et al., 1997).

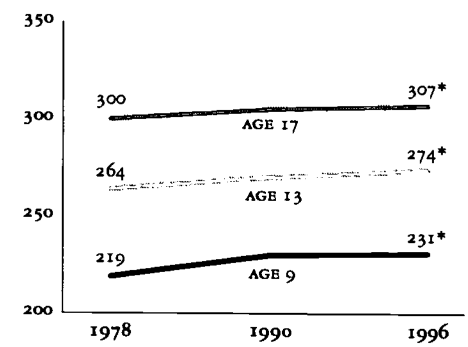
When this data is broken down by quartiles in Figure 6, the result is the same: every quartile has improved. For example, the lower quartile of student scores improved from 221 to 237—a statistically significant increase over the 18-year span from 1978 to 1996. The results show that for all students, regardless of achievement level, the NAEP results have improved over time. Moreover, grade 4 and 8 students have shown statistically significant improvements since both the 1990 and the 1992 assessments.

FIGURE 4
TIMSS MATHEMATICS

	Grade 4	Grade 8	Grade 12
INTERNATIONAL AVERAGE	529	513	500
U.S.	545	500	461
JAPAN	597	605	N/A
CANADA	532	527	519
ENGLAND	513	506	N/A
GERMANY	N/A	509	495

Note: TIMSS scale 0 to 800 each grade
Source: NCES, 1996a, 1997a, 1998a

FIGURE 5
LONG-TERM NATIONAL TRENDS, NAEP MATHEMATICS



** Significant change from 1978   Source: Campbell et al., 1997*

FIGURE 6
NAEP MATHEMATICS TRENDS BY QUARTILES

	1978	1990	1996
AGE 9			
UPPER	256	266	268*
MIDDLE TWO	221	231	232*
LOWER	178	190	191*
AGE 13			
UPPER	305	307	311*
MIDDLE TWO	266	271	275*
LOWER	221	234	237*
AGE 17			
UPPER	339	341	342*
MIDDLE TWO	302	305	308*
LOWER	260	268	270*

** Significant change from 1978. Scale range: 0 to 500*
Source: Campbell et al., 1997

GROWTH IN MATHEMATICS LEARNING FROM GRADE 4 TO GRADE 8

One of the concerns for mathematics educators arising from the TIMSS results was the finding that our students do worse at grade 8 mathematics compared to other nations than at grade 4. An approach to analyzing NAEP results called *cohort growth analysis* sheds light on the question of how much improvement U.S. students make over those four years of schooling. NAEP mathematics results by state for 1992 and 1996 can be analyzed to show the extent of improvement in the same cohort of students over time. Barton and Coley (1998) used the cohort analysis method to determine the extent of improvement in NAEP for the nation and by state from 1992 to 1996 by comparing the scores of grade 4 students in mathematics with scores four years later for grade 8 students. Thus, samples of the same cohort of students are compared. The advantage of this approach is to move closer to determining the effects of schooling in improving mathematics because comparable samples of schools and students are compared from one period to the next.

Figure 7 shows selected states with high, average, and below average growth in terms of number of NAEP scale points increase over four years—1992 to 1996. The states with highest growth, Nebraska and Michigan, improved scores by 57 points—Nebraska scores rose from 226 in 1992 to 283 in 1996, while Michigan increased from 220 in 1992 to 277 in 1996. The average growth for the nation was 52 points, and six states were at this level of growth. A total of 14 states were above average growth, and 14 were below average.

The growth analysis by state allows us to see that states whose mean score was below the national average at grade 8, such as Kentucky and Arkansas, have made the same amount of improvement in mathematics learning as Maine which is near the top in performance at grades 4 and 8. Several states with close to average scores, such as Michigan and North Carolina, were near the top in math growth from grade 4 to grade 8. Five scale points difference in growth (e.g., 57 points vs. 52 points) could be viewed as about three months difference in mathematics learning, since the average increase in NAEP scores is about 13 points per school year.

The NAEP trend data tell us that scores in mathematics have been improving, steadily but gradually, over the past 23 years. Yet the TIMSS results show that, after fourth grade, U.S. students lag far behind students around the world. Within this context, we now examine results in specific content areas of mathematics, along with sample items.

FIGURE 7
COHORT GROWTH FOR SELECTED STATES IN MATHEMATICS
NAEP 1992 TO 1996

High Growth in Math	NAEP Score Points Improvement*
NEBRASKA, MICHIGAN	57
NORTH DAKOTA, MINNESOTA	56
NORTH CAROLINA, COLORADO	55
Average Growth	
MAINE, MARYLAND, TEXAS, TENNESSEE, NEW YORK, KENTUCKY, ARKANSAS	52
NATION	52
Below Average Growth	
MISSISSIPPI, SOUTH CAROLINA, ALABAMA, LOUISIANA, HAWAII	48
GEORGIA	47

* *Difference between average grade 8 score (1996) and grade 4 score (1992)*
Source: Barton and Coley, Growth in NAEP (1998).

15

Content Area Strengths and Weaknesses

In order to more fully understand the results from NAEP and TIMSS, it is helpful to examine some of the specific content areas covered by both assessments. On NAEP there was little variation in results from one content area to another, although some content areas showed improvement. In particular, public school students in grades 4 and 8 showed significant improvement in algebra, and grade 8 students have improved in geometry from 1992 to 1996. Since the 1990 assessment, grade 8 scores have improved in all content areas. While these findings show good news for mathematics education, we must keep in mind that the overall scores continue to be low (which is especially highlighted by the TIMSS results). The distribution of NAEP mathematics scores is negatively skewed, showing that there is an abundance of low scores that pull the mean score down.

In spite of the small amount of variation on NAEP among content areas, we can learn more about the relative strengths and weaknesses by looking at NAEP and TIMSS results by content area. Doing so enables us to make more accurate statements about what U.S. students know and can do in mathematics. Because the NAEP and TIMSS frameworks are similar, it is possible to look within content areas and analyze results from both assessments. In this next section, we will focus on those content areas that showed weaknesses at grades 4, 8 or 12. A sample of items from both assessments is included within each content area discussion.

MEASUREMENT IS A WEAKNESS AT ALL GRADES

In this content strand of mathematics, students are expected to have a conceptual and procedural understanding of measurement units, the ability to use measurement tools and instruments, and the ability to solve problems related to perimeter, area, and volume. The NAEP assessment also measured students' ability to estimate absolute and relative measurements. At fourth grade, the emphasis for NAEP was on measurement of time, money, temperature, length, perimeter, area, weight/mass, and angles. Problems for students in grades 8 and 12 were more complex and involved volume and surface area in addition to the other topics. Some questions also dealt with reasoning with proportions, which are skills required in scale drawing and map reading.

The TIMSS data show that measurement is a particular weakness of U.S. elementary and middle grades students. U.S. students scored below the international average in this content area at both grades 4 and 8. On NAEP, students' average score in measurement (mean and median) was lower than the overall average at grades 8 and 12. So, relative to other content areas on NAEP, measurement was weak in those grades. The questions that were particularly difficult for students were those requiring unit conversions, calculations of volume and circumference, and estimation of measurements.

FIGURE 8
"ODOMETER"
NAEP 1996, GRADE 8

A car odometer registered 41,256.9 miles when a highway sign warned of a detour 1,200 feet ahead. What will the odometer read when the car reaches the detour?

A. 42,456.9
B. 41,279.9
C. 41,261.3
D. 41,259.2
E. 41,257.1

OVERALL CORRECT	26%
PROFICIENT	50%

The Odometer item in Figure 8 is a multiple choice item that assesses the measurement strand. Using the conversion for feet to miles given in the problem, the student must convert the 1,200 feet given in the problem into miles, and then add this amount to the reading on the odometer. Using a calculator,

16

the student could compute 1,200 divided by 5,280 and then the sum of that quotient and 41,256.9. As an alternative, the student could use number sense and estimation to solve the problem, especially since the answer is in multiple-choice format. The student could estimate that 1,200 feet is about 1/5 of a mile, which would be 2/10, and add 0.2 to 41,256.9, choosing the correct answer of "E".

U.S. eighth graders did not do well on this item: only 26 percent got it correct. Another 37 percent chose option A, found by simply adding the number of feet to the odometer reading without first converting the distance to miles. Males performed significantly better than females on this item. Of students whose overall NAEP score was at the Proficient level, 50 percent got this item correct. The item touches on several weak spots in overall performance by students on NAEP: measurement, operations with decimals, and estimation.

One of the primary problems with the teaching of measurement is that, when it is taught, it is often done with proxies or simulations. Students are given too few opportunities to actually engage in the use of measuring instruments. Instead they are shown pictures of objects and the instruments chosen to measure some attribute of those objects. Students should have more hands-on experiences with measuring, including making decisions about which instrument might be appropriate for measuring a particular attribute. Emphasis should be placed on understanding the underlying concepts, rather than simply applying formulas. Measuring activities should emphasize the approximate nature of measurement, and the related issues of precision and error.

NUMBER SENSE AND ESTIMATION: STUDENTS HAVE DIFFICULTY APPLYING KNOWLEDGE TO SITUATIONS

The NAEP results show that U.S. students can do simple whole number computations but lack flexibility in applying them to new or unusual situations. The content strand of number sense and estimation covers basic arithmetic skills and concepts, which represent a significant part of the mathematics curriculum at most U.S. schools, particularly at the lower grades. The NAEP assessment reflected this emphasis, with 40 percent of the questions for students in grade 4, 25 percent of those in grade 8, and 20 percent of those in grade 12 falling within this content strand. For TIMSS, number sense and estimation was covered in sections on measurement, estimation and number sense for students in grade 4, and within fractions and number sense for students in grade 8.

NAEP scores on number sense were below the overall average at grades 4 and 12. While fourth graders were strong in whole number computation, they showed weaknesses in number sense items. Students scoring in the Basic achievement level on NAEP (64 percent for grade 4, 62 percent for grade 8, and 69 percent for grade 12) appeared to grasp many of the fundamental concepts of numbers, relationships between numbers, and properties of numbers, as well as to display the skills required for manipulating numbers and completing computations. Questions requiring multi-step solutions or involving new concepts tended to be more difficult. Additionally, questions requiring students to solve problems and communicate their reasoning proved challenging, and often it was the communication aspect that provided the most challenge. Internationally, U.S. students were below the international average at grade 4, while eighth graders were at the international average.

Number sense involves having an intuition about numbers; it entails being able to use a variety of strategies, including mental computation, to find solutions to problems, either in a context or context-free. Questions in this content strand required students to demonstrate an understanding of number properties and operations, to generalize from numerical patterns, and to verify results. These questions also assessed student understanding of numerical relationships as expressed in ratios, proportions, and percentages. Students at all grade levels were assessed on their ability to reason mathematically and to communicate the reasoning they used to solve problems involving number sense, properties, and operations.

FIGURE 9
"SAM'S LUNCH"
NAEP 1996, GRADE 4

Sam can purchase his lunch at school. Each day he wants to have juice that costs 50¢, a sandwich that costs 90¢, and fruit that costs 35¢. His mother has only $1.00 bills. What is the least number of $1.00 bills that his mother should give him so he will have enough money to buy lunch for five days?

OVERALL CORRECT	17%
PROFICIENT	44%

The regular constructed-response item for fourth graders in Figure 9 measures number sense and operations with decimals. Students could use many different strategies to solve this problem. One possibility is to add the cost of juice, sandwich, and fruit, multiply by 5, and then round up to the nearest dollar. Or the cost of each food item could be multiplied by 5 and then a total sum found. Another strategy would be to estimate the number of dollars needed each day (2), then see that the change from four days (.25 times 4) makes another dollar, so the total would be 9. Regardless of the approach, the successful student must take into account the cost of the food over five days and the notion of "least" number of dollars.

This item was scored using a three-point scoring guide that allowed for partial credit. This question was difficult for most students. Ten percent did not respond to the question, and half of the students responded incorrectly. Only 17 percent of fourth graders received the highest score on this item. For students whose overall NAEP score was Proficient, 44 percent got the highest score on this item.

In the twelfth grade item shown in Figure 10, students must consider a percent of a percent to find the number of serious bicycle accidents that involve fatal head injuries. The correct answer (20 percent) is found by taking 80 percent of 25. The answer could also be found by using number sense, and estimating that 20 percent is about 80 percent of 25.

FIGURE 10
"BICYCLE ACCIDENTS"
TIMSS, GRADE 12

Experts say that 25% of all serious bicycle accidents involve head injuries and that, of all head injuries, 80% are fatal. What percent of all serious bicycle accidents involve fatal head injuries?

A. 16% B. 20% C. 55% D. 105%

INTERNATIONAL AVERAGE	64%
U.S.	57%

Twelfth grade students in the U.S. scored below the international average on this item. Only 57 percent chose the correct response, compared to the international average of 64 percent. In The Netherlands, for example, 83 percent of the students got the right answer.

While U.S. students seem to be fairly strong in basic whole number computation, they seem to lack the flexibility to apply those skills to new or unusual situations. The emphasis in the curriculum should move beyond basic paper and pencil computations with numbers to include such topics as computational estimation and mental computation. Both are skills that are strongly needed with the increased use of technology to perform the algorithms that used to be done with a pencil and paper. In a world where problem solving and reasoning are highly valued, mental strategies and the ability to judge the reasonableness of an answer are more important than ever.

18

GEOMETRY: FOURTH GRADERS STRONG, BUT EIGHTH AND TWELFTH GRADERS WEAK

The questions classified under this content strand centered around a conceptual understanding of geometric figures and their properties. Fourth-grade students were asked to demonstrate an understanding of the properties of shapes and to visualize shapes and figures under simple combinations and transformations, as well as to write verbal descriptions of the properties of geometric figures. Eighth grade students were asked about concepts related to properties of angles and polygons, such as symmetry, congruence and similarity and the Pythagorean Theorem. They also had to apply reasoning skills to make and validate conjectures about combinations and transformations of shapes. Twelfth graders were expected to demonstrate knowledge of more sophisticated geometric concepts and to use more sophisticated reasoning processes. Some questions involved proportional reasoning or coordinate geometry.

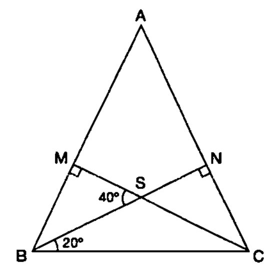
Fourth graders on TIMSS scored above the international average in geometry, with only two nations scoring significantly higher. By eighth grade, however, 25 countries scored higher than the U.S., and our students were below the international average. The advanced U.S. twelfth graders scored the lowest of all countries taking the test. On NAEP, fourth graders were relatively strong in geometry, compared with other content areas. The geometry results were lower at eighth grade, and then higher at twelfth grade.

The twelfth grade item in Figure 11 represents a fairly standard geometry proof. To gain full credit, students had to exhibit some understanding of angle sums in a triangle, isosceles triangles, and possibly other concepts, such as vertical angles and supplementary angles. They also had to be able to justify each statement in the proof. The international average for this item (getting it at least partially correct) was 48 percent, while only 19 percent of U.S. advanced twelfth graders reached this level of achievement.

FIGURE 11
"GEOMETRIC PROOF"
TIMSS, GRADE 12

In the $\triangle ABC$, shown below, the altitudes BN and CM intersect at point S. The measure of the $\angle MSB$ is 40° and the measure of $\angle SBC$ is 20°. Write a proof of the following statement:

"$\triangle$ ABC is isosceles."



ANSWER:

INTERNATIONAL AVERAGE	48%
U.S.	19%

Geometry needs to become a more visible part of the middle grades curriculum. Too often it is ignored until high school, and delays students' experience in a highly valuable and applicable part of mathematics. Geometric thinking and spatial visualization are linked to many other areas of mathematics, such as algebra, fractions, data, and chance. Teachers should spend more time developing geometric concepts (concretely and with many different representations) and principles (in varied settings) and not merely focus on practice involving algorithmic procedures.

PROPORTIONALITY: STRONG AT GRADE 4, WEAK AT GRADES 8 AND 12

Proportions are relationships among quantities that are related by multiplication. To say that quantity a is to quantity b as quantity c is to quantity d is to describe a relationship between a, b, c, and d that is multiplicative (for example, a times c is equal to b times d). Ratios and proportions are an important part of mathematics learning, usually starting in the upper elementary and middle grades. The TIMSS framework separates proportionality into a separate content category at grades 4 and 8, while NAEP embeds notions of proportional reasoning within the other strands, particularly number sense.

Results from TIMSS show that, like the other content areas, U.S. students fared better at grade 4 than at the higher grades. On items measuring fractions and proportionality, fourth graders scored slightly higher than the international average, with six countries scoring higher. Eighth graders scored below the international average, and 18 countries scored higher.

FIGURE 12
"AMOUNT PAID"
TIMSS, GRADE 8

Peter bought 70 items and Sue bought 90 items. Each item cost the same and the items cost $800 altogether. How much did Sue pay?

INTERNATIONAL AVERAGE	38%
U.S.	23%

To be successful on the eighth grade item shown in Figure 12, a student needs to find the portion of the $800 total that accounts for Sue's items. This could be accomplished by computing the total number of items (160) and calculating that Sue's portion is 90/160, therefore her portion of the total bill would be that amount times 800. The international average for this short constructed-response item was 38 percent. Only 23 percent of eighth graders in the U.S. got this item correct. In Singapore, 83 percent of the eighth graders got this item correct.

Figure 13 shows a twelfth grade, regular constructed-response item that measures number sense, proportions, and operations. Students' responses were scored using a three-point scoring guide that allowed for partial credit. To earn a "satisfactory" score, or full credit, the student would have to show that Martin's mixture had a stronger cherry flavor. This would entail showing that Martin's ratio of 5 ounces of syrup to 42 ounces of water had a higher syrup to water ratio than Luis' mixture of 6 ounces to 53 ounces of water. In the sample response, the student showed this by dividing 5 by 42 and 6 by 53 and showing that one quotient was larger than the other. Other possible strategies would include setting up a proportion and "cross multiplying," or finding a common denominator and comparing numerators. Only 23 percent of U.S. twelfth grade students were able to reach the satisfactory level on this item, including 60 percent for those whose overall score was Proficient.

FIGURE 13
"CHERRY SYRUP"
NAEP 1996, GRADE 12

Luis mixed 6 ounces of cherry syrup with 53 ounces of water to make a cherry-flavored drink. Martin mixed 5 ounces of the same cherry syrup with 42 ounces of water. Who made the drink with the stronger flavor?

Give mathematical evidence to justify your answer.

OVERALL CORRECT	23%
PROFICIENT	60%

Proportional reasoning is a challenging concept for many students, and one that should be an important part of the middle grades curriculum. To be able to understand and work with proportional situations requires a multiplicative way of thinking about relationships. Teachers need to enhance the development of the topic as it is typically presented in textbooks, with hands-on experiences. Initial activities should focus on the development of meaning, postponing efficient procedures until such understandings are internalized.

ALGEBRA RESULTS HIGHER THAN OTHER AREAS, BUT OVERALL KNOWLEDGE AT A LOW LEVEL

The algebra strand extends from work with simple patterns at grade 4 to basic algebra concepts at grade 8 to sophisticated analysis at grade 12. Students are expected to use algebraic notation and thinking in meaningful contexts to solve mathematical and real-world problems.

Fourth grade algebra questions on TIMSS were categorized as those involving patterns, relations, and functions. In this content area fourth graders were again above the international average, with only four nations scoring higher. Eighth graders scored just at the international average, and twelfth graders were lower. From the NAEP results we can say that, relative to the other content areas, algebra results were higher at all three grade levels. However, the overall level of algebra knowledge is relatively low. In particular, 64 percent of fourth graders, 62 percent of eighth graders, and 69 percent of twelfth graders scored well on basic algebraic representations and simple equations, as well as finding simple patterns. But only 24 percent of eighth graders and 16 percent of twelfth graders demonstrated knowledge of linear equations, algebraic functions and trigonometric identities expected for their grade level. Only four percent of eighth graders and two percent of twelfth graders demonstrated the ability to identify and generalize complex patterns and solve real-world problems expected for their level.

The item shown in Figure 14 asks eighth graders a conceptual question about an algebraic equation. Rather than asking students to simply manipulate symbols, this question probes a student's understanding of a variable expression. The student must be aware that $n-1$, n, and $n+1$ are representations of the consecutive whole numbers in the problem, and that n would necessarily represent the middle of the numbers.

FIGURE 14
"MEANING OF EQUATION"
TIMSS, GRADE 8

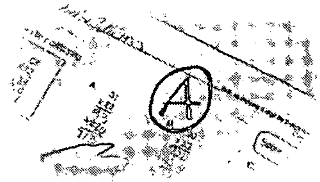
Brad wanted to find three consecutive whole numbers that add up to 81. He wrote the equation $(n\text{-}1) + n + (n+1) = 81$. What does n stand for?

A. The least of the three whole numbers.
B. The middle whole number.
C. The greatest of the three whole numbers.
D. The difference between the least and greatest of the three whole numbers.

INTERNATIONAL AVERAGE	37%
U.S.	32%

The international average for this item was low at 37 percent, yet the number of U.S. eighth graders choosing the correct response was below the average, at 32 percent. In Japan, 62 percent of the students chose the correct response.

Algebra has traditionally been considered a high school course, but more and more students are taking a first formal course in eighth grade. There is a need for algebraic thinking to be introduced in the early elementary grades, so that algebra becomes a natural way of expressing what is sensible in arithmetic. With this sort of foundation, the first formal courses in algebra would be a smoother continuation of ideas that had evolved naturally.



A Closer Look at Item Types on NAEP

The 1996 NAEP in mathematics included three types of items: multiple choice, regular constructed response, and extended constructed response. The regular constructed response items required that students provide their own answer, rather than selecting from a given set of options. The extended constructed response items, however, involved longer responses and were designed to assess higher levels of problem solving, reasoning, and mathematical communication. Typically, students are asked to explain their thinking or justify their solutions. These latter items were first included on NAEP in 1992.

The results by item type from 1992 and 1996 show wide differences. At all three grade levels, student performance on multiple choice items was significantly better than performance on constructed response items. Figure 15 shows the mean percent correct for multiple choice and regular constructed response, and the percent satisfactory or better on the extended constructed response (Dossey et al, 1993; Silver et al., 1998). Clearly students perform better on multiple choice questions than on either type of constructed response. The results are similar for 1992 and 1996. It is interesting to note that eighth graders did comparatively better on the regular constructed response items than either fourth or twelfth graders.

The most obvious conclusion from Figure 15 is that overall performance on all of the item types is too low: slightly more than half of the responses to multiple choice items are correct, but performance on the constructed response questions was worse. When students were asked to produce an answer rather than choose from a set of possible answers, their performance declined considerably. When asked to produce an extended response, their performance was abysmal.

Perhaps the tasks are too difficult

What is it about the extended constructed response items that makes the performance levels so poor? Some may wonder whether the tasks are too difficult. Let's take a look at one fourth grade released item that was on both the 1992 and the 1996 NAEP.

The task in Figure 16 shows two figures, both drawn on the same grid. Students are first asked to tell how the two figures are alike, and then to list ways that they are different. A typical response might say that they are alike because they both have four sides, or that they have the same length or base, or that they both have little squares inside. It is also true that both figures have the same height and the same area, and

FIGURE 15
RESULTS BY ITEM TYPE
NAEP 1992, 1996

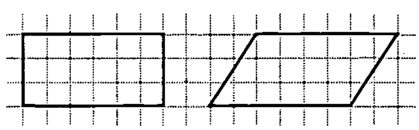
	Multiple Choice (% Correct)	Constructed Response (% Correct)	Extended Constructed Response (% Satisfactory or Better)
1992			
GRADE 4	50%	42%	16%
GRADE 8	56	53	8
GRADE 12	56	40	9
1996			
GRADE 4	54	38	17
GRADE 8	55	49	9
GRADE 12	60	34	12

Source: NAEP Mathematics, 1992, 1996

they both have parallel sides. For differences, the student might notice that one figure has four equal angles, while the other does not, and that those equal angles are right angles. The figures also have different perimeters. They belong to different classes of four-sided figures, in that one is a rectangle and one is a parallelogram.

FIGURE 16
"TWO GEOMETRIC SHAPES"
NAEP 1996, GRADE 4

Think carefully about the following question. Write a complete answer. You may use drawings, words, and numbers to explain your answer. Be sure to show all of your work.



In what ways are the figures above alike? List as many ways as you can. In what ways are the figures above different? List as many ways as you can.

MINIMAL	31%
PARTIAL	29%
SATISFACTORY	11%

The scoring rubric for this task essentially counts the number of correct reasons, both for similarities and differences. To earn a Minimal response, the student needs to identify one correct reason, or give a nonspecific response. A Partial response requires two reasons, while a Satisfactory response requires three in some combination of similarities or differences. The Extended response requires some combination of four similarities or differences.

Nearly one third of fourth graders had responses to this item that were either incorrect or off task (or there was no response at all). Most fourth graders were only able to identify one or two reasons why the figures were either alike or different, and only about 11 percent reached the Satisfactory or Extended level.

Why the poor showing? This task does not present a complex problem solving situation. It presents two simple geometric figures and asks how they are alike or different. The scoring rubric does not differentiate between students who responded about the figures based solely on appearances ("one looks slantier than the other") versus those who used more sophisticated geometric terminology, yet most students were still unable to come up with more than two characteristics. Perhaps we should consider other possible causes for the poor performance on this and other extended constructed response tasks.

Maybe they didn't try hard enough

Among the many reasons for the relatively poor showing on the NAEP extended constructed response tasks, one must consider motivation. On the sample item shown above, 8.5 percent of the responses were off task or omitted. The issue of motivation is more clearly seen with responses to the twelfth grade items, comparing the number of "no response" or "off task" papers in twelfth grade to those in the eighth or fourth grade. For example, on the released extended constructed response tasks from 1996, the twelfth grade results showed a range of 25–30 percent of the papers in the combined categories of "off task" or "omit." In contrast, at both fourth and eighth grades, the number of "off task" or "omit" responses on the released tasks ranged from about 6 percent to about 13 percent. Because NAEP is designed to give results at the national or state level, there are no student-level scores given. Nor are there any consequences for students based on their results. It appears that twelfth graders, aware of the lack of consequences, put less effort into these tasks than did either eighth or fourth graders.

Maybe they didn't get to the item

Another, perhaps related, reason for the poor showing on the extended constructed response tasks is the position of the tasks on the test. Each student is given a block of items to work, and this includes a combination of multiple-choice, regular constructed response, and one extended constructed response items. The extended constructed response item is always positioned at the end of the block. It is reasonable to assume that, for some students, there is insufficient time to work the extended constructed response item. The number of "not reached" responses to the fourth grade item was only 2.5 percent. Again, the statistics for "not reached" are higher for twelfth grade students than either eighth or fourth. When combined with lower motivation, this could account for many students either not trying at all, or not giving their best efforts to these items. After all, these are items that demand much more effort than either choosing from a set of given responses, or finding a single numerical answer.

Perhaps they have never done this kind of mathematics before

Another consideration regarding the poor showing on these items is the opportunity to learn the content embedded in the item. On the 1996 NAEP, a disproportionate number of extended constructed response items came from the content areas of data and geometry. Both of these areas have been traditionally weak in the U.S. curriculum, especially at fourth and eighth grades. Both are topics found in the last chapters of traditional elementary textbooks, and often not reached (or valued) by elementary teachers. On the comparison of geometric figures described above, typical instruction in the elementary grades might include simple definitions for certain shapes, but students are not often asked to analyze the geometric features of those shapes. In fact, a qualitative analysis of a sample of student responses showed that students had much more difficulty with the "difference" question than the "likeness" question, which may be related to the frequency with which students are asked to look for contrasts, rather than comparisons.

These items demand a different kind of reasoning

Aside from lack of motivation, insufficient time, or opportunity to learn, it is important to consider the more substantive reasons for low scores on these items. That is, these items assess higher order thinking skills, such as reasoning, making connections, analyzing, making conjectures, and writing explanations or proofs. Such tasks are not only inherently more difficult, but often students have not had many opportunities to engage in this kind of activity in mathematics class. When students spend most of their class time learning how to perform relatively simple procedures, rather than learning to understand concepts and engaging in higher order thinking, it is not surprising that they do poorly on these types of tasks. Fourth graders who may have learned to identify shapes will have a more difficult time when asked to analyze the geometric properties of those shapes.

These tasks require good communication skills

Nearly all extended constructed response tasks demand that students be able to communicate mathematical ideas clearly, precisely, and concisely. Responses might include a variety of types of communication, such as equations, graphs, tables, diagrams, charts, and words. Scoring of these responses almost always involves how clear the explanation is, or how clearly the student can demonstrate understanding of the concepts. Once again, students need practice in these types of activities. They need to be in classrooms where both verbal and written explanations are valued and form an integral part of the class. From the earliest grades, students need to gain an understanding of what constitutes a valid mathematical argument, and they need to practice explaining their ideas to others.

Problems are often non-routine, and students do not have readily available algorithms or procedures for solving them

For nearly twenty years mathematics educators have seen from the NAEP results that students have difficulty with any non-routine problems that require more than one step to solve or involve some analysis or thinking. In an interpretive report of the 1980 NAEP mathematics assessment, Carpenter et al. noted,

> Part of the cause of students' difficulty with non-routine problems may lie in our overemphasis on one-step problems that can be solved by simply adding, subtracting, multiplying, or dividing. ...Instruction that reinforces this simplistic approach to problem solving may contribute to students' difficulty in solving unfamiliar problems. (1981, p. 146)

Certainly most extended constructed response tasks fit the description of non-routine, multi-step problems. It is also clear that little progress has been made in students' abilities to solve such problems successfully. Similar comments about the need for instruction in such tasks were made after the 1992 assessment (Kenney & Silver, 1997).

SUMMARY AND RECOMMENDATIONS

The poor performance on the extended constructed response items may be due to any one of, or the combination of, the reasons given here. While we might explain some of the results by test-related issues, such as motivation or the placement of the items at the end of the block, those were clearly not the major reasons for the low scores on the fourth grade item we analyzed. The lessons to be learned from these results are clear: students in mathematics classes need more opportunities to work non-routine problems, to use higher-order thinking skills, and to communicate their mathematical ideas.

Too often, we find that national, state, and local assessment results are reported and interpreted only with mean scores, percentages of all students, or other statistics of central tendency. Large-scale studies such as NAEP, TIMSS, or state assessments are reported in a way that gives little idea of differences among groups of students, types of schools, or different curricula. In the 1990s, NAEP results have typically been reported using state-level averages and percentages, (e.g., the percent of students at/above the Proficient level). A common use of state NAEP results is to compare where one state ranks, on average, against other states in the same region or the nation. This kind of one-statistic analysis is often too narrow, not informative, and not useful for mathematics educators or decision-makers. The analysis of results can lead to misleading interpretations. Careful study of differences within the target population can increase the usefulness of the results.

In the following section we look at four examples of how to analyze differences in NAEP results at national and state levels. The examples discussed in this paper focus on variations in NAEP results that are in our judgement likely to have specific policy and program interest to states.

DIFFERENCES IN STATE NAEP RESULTS BY STATE CONTEXT

Public analysis and discussion about NAEP results has often focused on attributing high or low results to differences in characteristics of the state, such as social, cultural, or economic characteristics. In Figure 17 we show NAEP results for high and low performing states by three variables commonly cited as being reasons for test score differences related to state context: amount of money spent on education, students living in poverty, and adult education level.

FIGURE 17
STATE MATH PROFICIENCY BY MEASURES OF STATE CONTEXT
NAEP 1996, GRADE 8

High-Performing States	% At/Above Proficient	Children in Poverty (5–17 YEARS OLD)	Expenditures per Pupil (ADJUSTED COL)	Education of Adults (% H.S. GRADS)
MINNESOTA	34	16%	$5,738	82%
NORTH DAKOTA	33	14	5,234	77
WISCONSIN	32	14	6,588	79
MONTANA	32	18	5,653	81
CONNECTICUT	31	18	7,279	79
Low-Performing States				
MISSISSIPPI	7	33%	$4,358	64%
LOUISIANA	7	34	4,875	68
ALABAMA	12	24	4,604	67
ARKANSAS	13	22	4,804	66
NEW MEXICO	14	29	4,749	75

Source: Blank, Manise, et al., 1997; (Poverty—Current Pop. Survey, 1994; Expenditures—NCES, 1994–95; Adult Education—Bureau of Census, 1990)

This kind of high/low analysis can tell us if there is a likely correlation between the characteristics of the state and student performance. The state context also informs educators about some of the barriers or disadvantages that schools and districts must address in improving mathematics in classrooms.

The data show clearly that states with high NAEP math performance results have better conditions for education. Compared to low-performing states, high-performing states have 10 to 20 percent fewer children living in poverty, 10 to 15 percent more adults with high school diplomas, and $1,000 to $1,500 higher expenditures per pupil. The difference between the extremes on NAEP student performance results shows that state context is related to student achievement in mathematics.

VARIATION IN MATHEMATICS PROFICIENCY FROM TOP STUDENTS TO BOTTOM STUDENTS

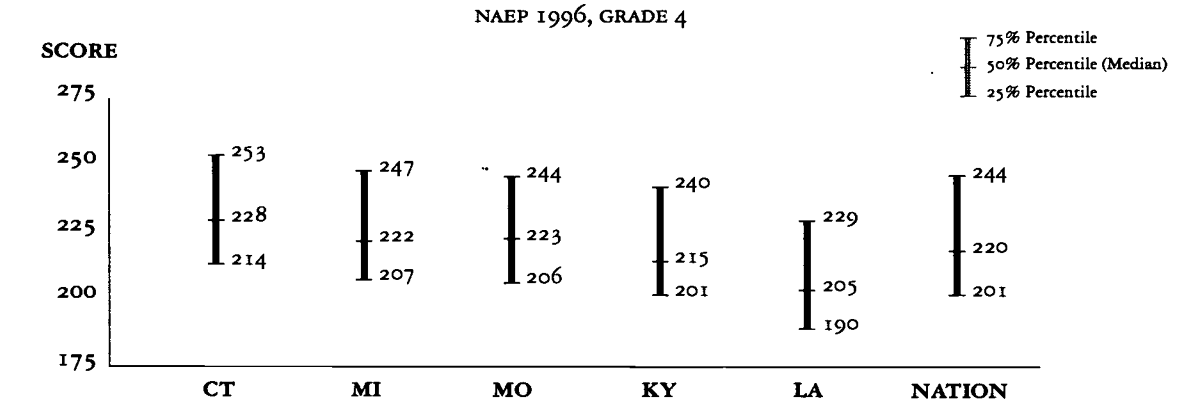
A way of improving our analysis of student achievement is to examine the extent of variation between the students performing at high and low levels, and the degree to which students at each end of the distribution are improving their learning. Educators and policymakers need first to see the total range of performance and then examine where the average falls in the range. An average score may hide poor performance by a specific portion of students.

State percentile scores are a useful approach for analyzing variation. Percentiles give the range of performance—differences between the students learning the most mathematics and those learning the least. To obtain a picture of how much score variation there is among states we examine NAEP results from five states in Figure 18. The states were selected according to the percentage of students scoring at/above Proficient in grade 4: Connecticut (the top performing state), Michigan (11th from top), Missouri (21st), Kentucky (31st), and Louisiana (41st). The percentile breaks on NAEP are reported using the NAEP scale (0 to 500). For grade 4 students a score of 249 and higher is at/above the Proficient level, and 214 is at/above the Basic level (Reese, et al., 1997, p.10).

The variation nationally from bottom quartile to top quartile at grade 4 mathematics is 43 points. This difference shows significant variation in mathematics learning and performance. Variation in performance of grade 4 students was similar in all five states for 1996. Thirteen points on the NAEP scale represent about one year of education, based on differences between grade 4 and grade 8 results, which means a difference of over 40 points can be viewed as greater than three years of math education.

It should be noted that the differences in math performance within each state are much greater than differences between states. For example, the

FIGURE 18
SCORE VARIATION WITHIN SELECTED STATES
NAEP 1996, GRADE 4



Source: NAEP Cross-State Compendium

difference in average scale scores between Connecticut and Louisiana is only 23 points.

NAEP results can address the effect of more grades of schooling on differences in student math performance. Above, we observed that math performance on NAEP did improve in some states, particularly at grade 8. With the data on variation, we can ask whether more schooling tends to increase or decrease the variation in math performance. The results show that variation does increase from grade 4 to grade 8 (NAEP *Cross State Compendium*, 1996): Connecticut increased from 39 points (grade 4) to 48 points (grade 8); Michigan from 40 points to 49; Missouri from 38 to 43; Kentucky from 39 to 42; and Louisiana from 39 to 43.

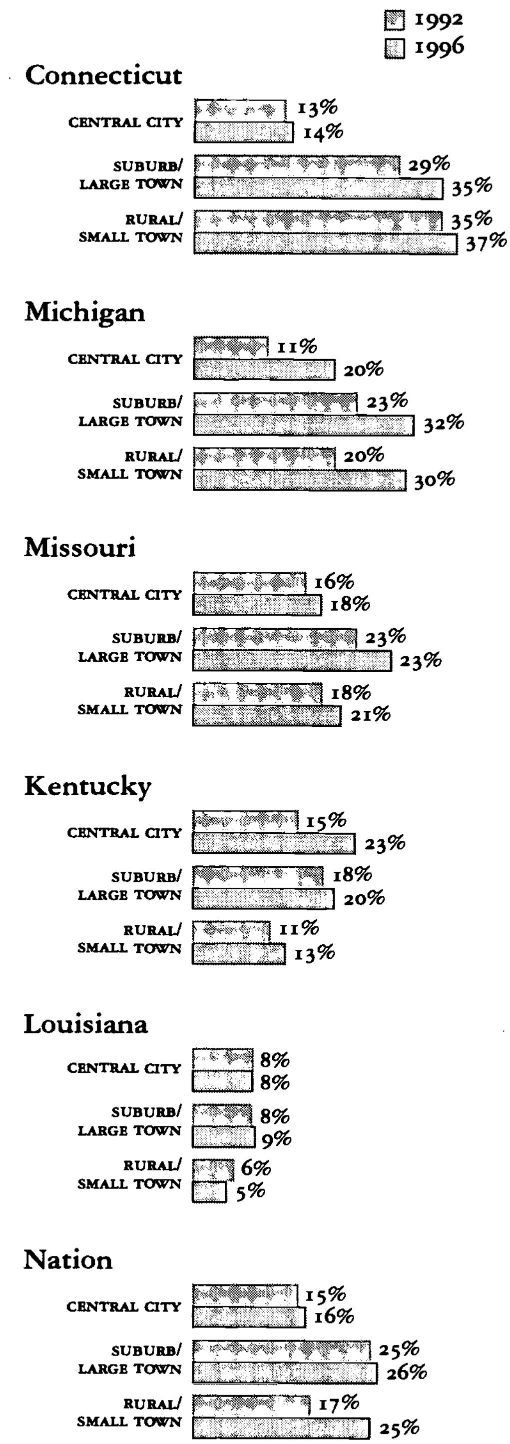
VARIATION BY TYPE OF COMMUNITY OR SCHOOL LOCATION

A second kind of analysis of differences in performance in mathematics is by type of community or location of school. From 1990 to 1996, a total of 27 states had significant improvement in the percentage of grade 8 students at/above Proficient. We can go further with the NAEP data and answer the question of how scores, and improvement, differ by the community characteristics of schools. For example, we can examine the percentage of students reaching the Proficient level by schools in central cities, suburbs, and rural areas.

In Figure 19, we show variation in 1992 and 1996 NAEP results by school location. Students in central city schools do less well in math in each state. In four of five states, students in rural or small town schools score slightly less well than suburban children, with the exception of rural/small town Connecticut schools.

The state and community location analysis shows significant improvement in suburbs/large towns in Michigan (9 percentage points) and Connecticut (6 percentage points). Math performance improved in central cities in Michigan (9 percentage points) and Kentucky (8 points). Michigan also showed the

FIGURE 19
DIFFERENCES IN SELECTED STATES' MATHEMATICS PROFICIENCY BY SCHOOL LOCATION
NAEP GRADE 8, 1992 TO 1996



Source: NAEP Cross-State Compendium

most improvement in rural and small town schools (10 points). Nationally, the most improvement at grade 8 math was in rural/small town schools.

The NAEP *Mathematics Cross-State Compendium* report (Shaughnessy, et al., 1998) shows that other states making significant improvement in central city schools were North Carolina (11 points) and Indiana (7 points). States with significant improvement in rural schools were Rhode Island (9 points) and Wisconsin (8 points).

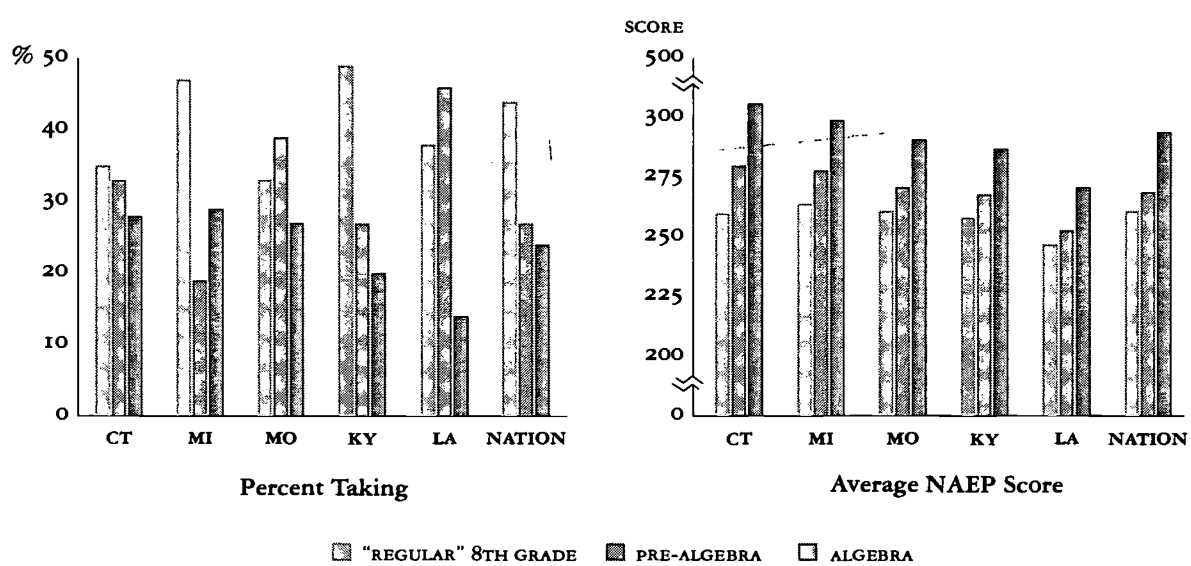
The data lead to several questions that states and districts could ask in order to further analyze their performance: What efforts did some states, such as Michigan, Kentucky and North Carolina, make in central cities to raise scores as much as in suburban and rural schools? What could educators in central cities in Connecticut learn from their colleagues in suburban and rural schools that are scoring very well on NAEP?

VARIATION IN MATH PERFORMANCE BY CURRICULUM

A key finding reported from analysis of the Second International Mathematics Study (SIMS) was that schools in the U.S. really offer three or four different mathematics curricula. A primary reason for the wide variation in math performance on achievement tests is the differentiated curriculum provided to students (McKnight, et al., 1987). Analyses of recent NAEP results show that high mathematics proficiency is best explained by the level of mathematics courses students have completed (Mullis et al., 1993, Reese, et al., 1997; Jones, L.V., et al., 1986).

Figure 20 shows differences in NAEP 1996 math performance according to the percentage of students that took algebra, pre-algebra, or "regular" eighth grade math.

FIGURE 20
EIGHTH GRADE MATHEMATICS CURRICULUM IN SELECTED STATES
BY NAEP SCORE, 1996



Source: NAEP Cross-State Compendium

BEST COPY AVAILABLE

29

Grouping students by prior mathematics achievement has been debated by educators and researchers for years. The data for these selected states show clearly that student achievement is predicted in each state by the course students take, particularly when considering algebra vs. pre-algebra. We should also note that the proportion of students taking these different course levels differs significantly among states. Kentucky has 49 percent of students taking regular eighth grade math and 20 percent taking algebra, as compared to 35 percent of Connecticut students taking regular math and 28 percent taking algebra.

Research on patterns of student achievement has demonstrated that instructional time and course taking in math and science varies widely across U.S. schools, and that they are correlated with the socioeconomic status of students in the schools (Goodlad, 1984; Horn & Hafner, 1992; Oakes, 1990; Weiss, 1994; NCES, 1997c). New research on secondary curricula show that schools with highly differentiated (or tracked) secondary course offerings have the lowest achievement among economically disadvantaged students (Lee, et al., 1995.)

The level of mathematics curriculum that students reach in high school makes a major difference in achievement for all groups of students. The table in Figure 21 shows the NAEP achievement scale score by three student ethnicity groups and by the level of high school mathematics attained by graduation.

Course-level is a strong predictor of NAEP scores for all student groups—white, black, and Hispanic students. The NAEP trends data back to 1973 show that all ethnic groups have doubled their enrollment in higher level math by graduation, e.g., algebra 2 and pre-calculus. In 1973, only 28 percent of black students took algebra 2 by graduation as compared to 45 percent in 1996. At the same time we should note that even though NAEP scores of all groups have gone up significantly since 1973, there is still a significant gap in achievement scores between white students and black and Hispanic students at all course levels. CCSSO's biennial report on *State Indicators of Science and Mathematics Education* (Blank, et al., 1997a) provides a detailed analysis of disparity in NAEP results by student race/ethnicity and trends since 1990.

FIGURE 21
HIGH SCHOOL MATHEMATICS COURSE COMPLETED BY RACE/ETHNICITY
BY NAEP SCORE, 1996

	White		Black		Hispanic	
	Percent Taking	Scale Score	Percent Taking	Scale Score	Percent Taking	Scale Score
Algebra 1	11%	287	18%	273	16%	*
Geometry	15	304	16	280	19	*
Algebra 2	53	320	45	299	41	306
Pre-calculus/calculus	13	342	8	*	9	*

** Sample not sufficient for reliable estimate.*
Source: NAEP Trends in Academic Progress, 1997

BEST COPY AVAILABLE

30

Opportunity to Learn Mathematics

TEACHERS, CURRICULUM, INSTRUCTIONAL PRACTICES

The international studies conducted under the International Association for Evaluation of Education Achievement (IEA), have provided survey models and research designs for analyzing students' "opportunity to learn" mathematics in classrooms and schools. The in-depth analysis of results of the Second International Mathematics Study (SIMS), such as those by McKnight, et al., (1987) and Travers (1985), provided important research findings on the dominant role of the "implemented curriculum" in explaining their mathematics performance. The analysis showed the wide variation in classroom mathematics curriculum among and often within U.S. schools, and showed that curriculum differences were more central to explaining differences in achievement than other variables such as total instructional time, homework, class size, or teacher degrees and courses.

Recent research by Porter (1995) and McDonnell, Burstein, et al. (1995) tested the validity of a variety of measures of opportunity to learn including content coverage, instructional practices, and materials, and demonstrated the relationship of these measures to achievement. The critical role of teacher preparation and knowledge in opportunity for learning was outlined by Stevens (1993) and Darling-Hammond (1995).

Analyses of data from the National Education Longitudinal Study (Hoffer, et al., 1995; Elliott, 1998) show the significant differences in access to learning opportunities among our students. New analyses of NAEP mathematics results by state (Raudenbush, 1998; Education Trust, 1998) show that achievement results are related to differences in math curriculum, teacher preparation, and teaching practices both within and between states, and these opportunity differences are related to student minority and economic status. Grissmer and Flanagan (1998) have conducted new analyses of long-term NAEP trends to explain improved performance of black students and low-income students and found that lower class size made a significant difference for these student populations.

In our analysis, we focus on three kinds of critical variables in opportunity to learn mathematics:

- Teacher preparation in mathematics;
- Implemented mathematics curriculum; and
- Teaching practices in mathematics classrooms.

Measures in these areas have been used in prior studies of opportunity to learn. They are measures that can be analyzed with data available in NAEP and TIMSS, and they indicate conditions that can be affected by education policies and decisions about practice.

WELL-PREPARED TEACHERS OF MATHEMATICS MAKE A DIFFERENCE

National professional standards in mathematics and science, as well as many state standards, call for change in teaching and classroom practices to emphasize active learning by students, deep understanding of concepts, and developing skills in problem solving and reasoning (NCTM, 1989, 1991; AAAS, 1993; NRC, 1995). The standards for teaching in mathematics and science de-emphasize teacher lectures, memorizing facts and terminology, and curriculum aimed at briefly covering many topics.

One implication of states establishing challenging content standards in mathematics is that teachers need in-depth knowledge and understanding of their discipline, and skills in a variety of classroom practices that actively engage students. A problem in measuring teacher knowledge and skill in relation to standards is the lack of precision in the data. Most of the information about teachers and teaching are from self-report surveys by teachers or administrative records (e.g., degrees, transcripts, etc.). The videotape study of eighth grade mathematics classrooms in the U.S., Germany and Japan led by Stigler (NCES, 1997d) has revealed great differences in the approach and methods of teachers, the role of students in the classes, and the use of materials and technology in class.

For analyzing differences in performance of U.S. students in mathematics, we would ideally prefer measures of the *quality of teaching* in classrooms and the quality of teacher preparation. However, research has shown that traditional measures of differences among teachers in the amount of course preparation in mathematics and science are related to student performance among U.S. schools and classrooms. Although these measures are less useful for international comparisons, they are useful for explaining differences within the U.S.

An indicator often used in national and state-by-state reports is the percentage of teachers that hold an undergraduate or graduate major in the teaching field they are assigned to teach. Research has consistently shown a positive relationship between the amount of course work preparation of U.S. teachers in science and mathematics and student learning in those fields (Shavelson et al., 1989). A recent analysis of data from the Longitudinal Study of American Youth showed that each additional mathematics course taken by mathematics teachers above the average for teachers translates into two to four percent higher student achievement (Monk, 1993).

We can compare levels of teacher preparation in mathematics with NAEP mathematics student proficiency. Figure 22 shows the 10 states with highest NAEP performance at grade 8 (from 34 to 28 percent at/above Proficient), and the 10 states with lowest performance. For each of these states we report the percentage of secondary math teachers (grades 7–12) with a major or minor in mathematics/math education (based on the 1994 *Schools and Staffing Survey*, NCES, 1996b).

The data on teachers major or minor in mathematics show extensive variation among states—from 95 percent of secondary teachers well prepared in mathematics to just over 60 percent. Nationally, 80 percent of all mathematics secondary teachers (i.e., teaching math one or more period) have a major or minor in math; and, 72 percent of secondary teach-

FIGURE 22
MATHEMATICS PREPARATION OF GRADE 7–12 TEACHERS
IN HIGH- VS. LOW-PERFORMING STATES ON NAEP 1996

High-Performing States, NAEP, Gr. 8	7–12 Teachers with Math Major/Minor	Low-Performing States, NAEP, Gr. 8	7–12 Teachers with Math Major/Minor
MINNESOTA	93%	MISSISSIPPI	82%
NORTH DAKOTA	92	LOUISIANA	82
WISCONSIN	83	ALABAMA	87
MONTANA	90	ARKANSAS	80
CONNECTICUT	88	WEST VIRGINIA	89
NEBRASKA	87	SOUTH CAROLINA	71
IOWA	92	NEW MEXICO	81
MAINE	71	TENNESSEE	69
ALASKA	57	KENTUCKY	79
MASSACHUSETTS	75	HAWAII	62
NATION	80		

Note: *Standard errors for state estimates are from 2 to 6%.*
Sources: *NAEP 1996 Mathematics Report Card; Schools and Staffing Survey, 1994*

ers with their main assignment in math have a major in that field. (Note: the national figure of 20 percent without adequate math preparation represents more than 46,000 secondary teachers of math.)

The levels of teacher preparation in mathematics show a difference between the two groups of states. Seven of the ten high-performing states had percentages significantly above the 80 percent national average of teachers with a major or minor in mathematics or math education.

The low-performing states on NAEP had fewer well prepared teachers in mathematics. With the exception of Alabama and West Virginia, the low-performing NAEP states were near or below the national average for well-prepared teachers in mathematics.

Quality of teacher preparation in mathematics varies significantly *within* states by school. Other national data analyses (Ingersoll and Gruber, 1996; Weiss, 1994) have shown that schools with a high proportion of low-income or minority students have significantly lower percentages of math teachers with a major or minor in their field.

NEED FOR IMPROVED MEASURES OF TEACHER PROFESSIONAL DEVELOPMENT

Professional standards for teaching mathematics (NCTM, 1991) recommend that teachers have adequate course work preparation in the content areas they will be teaching. In addition the professional organizations recommend ongoing professional development in the subject content and methods of teaching their assigned field and grade level. One question on the NAEP teacher questionnaire addressed the number of hours of professional development in math or math education received by teachers during the past year.

In Figure 23 we compare the extent of professional development in math for two groups of states—states with highest improvement in NAEP scores at grade 4 (1992 to 1996) and states with highest improvement at grade 8 (1990 to 1996). Nationally, 28 percent of fourth grade teachers had 16 or more hours of professional development in mathematics education, and 48 percent of eighth grade teachers had 16 or more hours.

The percentage of grade 4 teachers with high levels of professional development varies widely by state.

FIGURE 23
IMPROVING STATES
BY TEACHER PROFESSIONAL DEVELOPMENT IN MATH
NAEP 1996

Grade 4 High Improvement States	Prof. Devel. Math 16+ Hours	Grade 8 High Improvement States	Prof. Devel. Math 16+ Hours
CONNECTICUT	22%	MINNESOTA	50
TEXAS	46	TEXAS	64
INDIANA	13	WISCONSIN	40
COLORADO	21	CONNECTICUT	47
NORTH CAROLINA	19	NEBRASKA	36
WEST VIRGINIA	20	MICHIGAN	44
TENNESSEE	19	COLORADO	42
NATION	28	NATION	48

Note: *Standard errors of state estimates are from 1 to 3 percent.*
Source: *NAEP 1996 Mathematics Cross-State Compendium*

For example, only 13 percent of Indiana's grade 4 teachers and 19 percent of North Carolina teachers received 16 hours or more of math development, while 46 percent of Texas teachers received this much development over a one year period. Only one of the high improving states at grade 4 had more than the national average amount of mathematics professional development (more than 28 percent of teachers). At grade 8, only two of the improving states had more than 48 percent of teachers receiving over 16 hours professional development.

The amount of professional development being received by teachers does not have a consistent relationship to student achievement averages for states that showed most improvement in NAEP scores. It might be expected that in states with more professional development in mathematics we would find greater improvement on NAEP mathematics than in states with less professional development. But this hypothesis is demonstrated not to hold up with the data available at the state level. A majority of the states that have shown high improvement are below the national average with respect to the percentage of teachers receiving more math professional development (i.e., 16 or more hours in the past year).

Our view is that focusing on the amount of time that teachers spend in professional development is inadequate because it does not analyze the quality or effectiveness of professional development provided. We also examined other questions pertaining to professional development in NAEP such as training with NCTM standards but none covered the topic well. States, districts, or schools may offer professional development or in-service experiences for a variety of reasons, including state law or district teacher contracts. To adequately analyze professional development, mathematics educators need to monitor its content, the methods used to work with teachers, the continuity and followup, and the impact on the teachers. The items on amount of exposure are too general and encompassing of different methods of teacher development and thus are unlikely to show a clear relationship of teacher development to student achievement.

OTHER DATA SOURCES ON PROFESSIONAL DEVELOPMENT

TIMSS did not include questions on either amount of professional development or what was studied. An alternate national and state-level data source concerning professional development is the Schools and Staffing Survey (SASS) which is normally conducted by NCES every four years. The 1994 SASS included questions about the content of professional development activities, i.e., what was intended for teachers to learn. Teachers reported on how much time they spent on selected topics during the previous year. The national percentages (NCES, 1996b) for elementary and secondary teachers are shown in Figure 24:

FIGURE 24
TOPICS AND TIME OF PROFESSIONAL DEVELOPMENT, K–12 TEACHERS

Topic	1–8 hrs	> 9 hrs.
Content study in a subject	15%	15%
Teaching methods	37	27
Methods of student assessment	40	12
Use of education technology	35	15

Source: SASS, 1994

These topics of professional development and time are reported by state and by elementary vs. secondary teachers in CCSSO's *State Education Indicators with a Focus on Title I* (Blank, et al., 1997c). The next SASS survey will provide data on professional development during the 1999–2000 school year.

CURRICULUM CONTENT DATA REVEAL MANY MATH TOPICS, LITTLE FOCUS

In the 1990's almost all states have developed new content standards or curriculum frameworks for core academic subjects, including mathematics (Blank, et al., 1997b; CCSSO, *Key State Policies*, 1998a). The state standards consistently call for focusing mathematics content on a smaller number of areas, developing mathematical abilities across content

34

areas, and planning the curriculum across the grades. As these standards are being applied in writing local curriculum, designing professional development, and developing student assessment programs, educators and policy makers are likely to need consistent, reliable information and feedback on the actual subject content that is taught in math and what students are expected to know and be able to do. CCSSO is currently studying the implementation of state reforms and standards in a project involving 11 states that employs "surveys of enacted curriculum" to measure classroom curriculum content and teaching practices. A field study of the surveys showed that the data can be collected from teachers, analyzed, and reported in ways that are useful to educators (CCSSO, SCASS Science Project, 1997). Gamoran, Porter and colleagues completed a study of effects of different high school mathematics instruction using a similar survey approach (1997).

The most comprehensive and detailed data on "implemented curriculum" have been collected and analyzed in international studies. TIMSS measured student achievement in 41 countries based on mathematics and science assessment frameworks developed by consensus of the participating countries (NCES, 1996a, 1997a). A highlight of the TIMSS results reported in the U.S. has been the "Videotape Classroom Study" of mathematics teaching in grade 8 (NCES, 1997d). The video and accompanying compact disk provides the opportunity to analyze teaching practices by content topic, teaching approach, and student activity, and it thus offers a qualitative dimension for research on how curriculum is taught. This level of detail would be very difficult and costly for states or districts to replicate on a large scale.

Therefore, we focus on data from TIMSS teacher surveys. This approach can be used more readily by states and districts to assist in analyzing achievement results in mathematics education. TIMSS data collection included surveys with teachers and students that had a goal of collecting reliable, comparable data on the content of curriculum in math and science classrooms across the participating countries. Teachers completed a survey that asked which of the 35 curriculum content topics in mathematics were taught during that year and the average number of periods the topic or topics were taught.

The results of the survey show variation across the 41 participating countries in the content of the actual curriculum taught, as well as the degree of variation in curriculum within a country. In Figure 25, we compare data on implemented curriculum in mathematics at grade 8 in Japan and the U.S. We show the percentage of teachers that covered each topic, and the average percentage of time per year spent on the topic.

As the data show, Japanese eighth grade teachers spend most of their time teaching a few topics: geometry, congruence and similarity, functions, relations and patterns, and equations and formulas. In fact, those four areas of the curriculum account for approximately 67 percent of the time they spend teaching. In contrast, teachers in the U.S. spread time very thin among a wide range of topics. The majority teach 16–18 different topics, with only one topic accounting for more than eight percent of their teaching time. That topic is fractions, which only about a quarter of Japanese teachers teach, accounting for only two percent of their time. These results point to why some have accused the U.S. mathematics curriculum, especially in the middle grades, of being "a mile wide and an inch deep." With so many different topics taught during that year, it is unrealistic to expect that any given topic will be treated at more than a superficial level.

FIGURE 25
MATHEMATICS CURRICULUM TOPICS, TIMSS, GRADE 8

	Japan		U.S.	
	Teachers Covered Topic	Percent of Time	Teachers Covered Topic	Percent of Time
Meaning of Whole Numbers	19%	1%	87%	6%
Fractions	26	2	98	17
Percentages	19	1	92	6
Number Concepts	26	2	8	1
Number Theory	14	1	95	6
Estimation/Number Sense	44	1	93	4
Measurement Units/Processes	25	1	88	5
Measurement Estimation/Error	40	2	61	2
Perimeter, Area and Volume	27	2	92	5
Geometry, 1D & 2D	81	14	90	5
Symmetry/Transformations	21	1	56	2
Congruence/Similarity	98	23	72	3
3D Geometry	22	1	51	2
Ratio/Proportion	66	3	98	5
Slope/Trigonometry	31	3	38	1
Functions, Relations, Patterns	81	12	63	3
Equations/Formulas	94	18	90	8
Data/Statistics	84	2	75	3
Probability/Uncertainty	2	0	66	2
Sets/Logic	10	0	49	2
Other advanced content	49	4	48	6

Source: TIMSS Teacher Questionnaire, Population 2 (Schmidt & Cogan, unpublished data, 1998)

TEACHING PRACTICES DATA INDICATE MATHEMATICS CHANGE BUT ALSO TRADITIONAL PRACTICES

The NCTM curriculum standards (1989) and teaching standards (1991) recommend approaches to instruction that increase students' conceptual understanding, and their abilities to communicate mathematically, to reason and solve problems with mathematics, to make connections between math learning and real-world problems, and to learn skills and procedures. Many states have completed their own standards and curriculum frameworks in mathematics and science that suggest teaching strategies or provide examples of classroom practices that are consistent with challenging content standards (Blank, et al., 1997b). We have selected data from the NAEP mathematics teacher survey that are intended to be used in analyzing teaching in relation to standards. In Figure 26, we show data on practices in five selected states covering the range of state performance. State-by-state statistics on these teaching practices are available in CCSSO's *Science and Mathematics Indicators* report (Blank, et al., 1997a, pp. 47, 49).

The statistics on teaching practices for the five states at different levels of student achievement do not show any overall relationship between these teaching practices, consistent with NCTM standards and state NAEP scores. There is considerable variation in these practices among the five states, and in comparison to the national averages. Connecticut

FIGURE 26

MATHEMATICS CLASSROOM PRACTICES IN SELECTED STATES

NAEP 1996, GRADE 8

	Develop Reasoning/ Analytical Ability *(Percent)*[1]	Discuss Solutions to Math Problems *(Daily)*	Write About Math *(Weekly/more)*	Use a Calculator *(Daily)*	NAEP Score Average
CONNECTICUT	59 %	44 %	42 %	60 %	280
MICHIGAN	48	48	39	78	277
MISSOURI	46	43	22	71	273
KENTUCKY	49	41	55	60	267
LOUISIANA	44	44	25	19	252
NATION	52	48	32	57	271

Notes: [1] *Percent = Teachers' response "a lot", as compared to "some" or "none". Standard errors of state estimates are from 1 to 4 percent.*

Source: *Teacher questionnaire; NAEP 1996 Mathematics Cross-State Compendium*

teachers report higher emphasis on "developing reasoning and analytical ability," Kentucky and Connecticut teachers include more writing answers to math problems, and Michigan and Missouri have greater use of calculators in class. The variation shown among the 50 states is consistent with the 50-state data (Blank, et al., 1997a).

Develop reasoning and analytical ability; Discuss solutions to math problems

These two practices address the problem solving and reasoning theme of the NCTM standards for mathematics education. Figure 26 shows that nationally, half the grade 8 students have teachers that report emphasis on teaching to develop reasoning and almost half the students discuss math problems almost every day. Although these practices are quite prevalent, many teachers also do not emphasize these approaches to teaching math. Other data on grade 4 classroom practices show that 35 percent of teachers report that students discuss problems with other students almost every day. In 12 of 45 states participating in NAEP 1996, over 50 percent of students in grade 4 reported they discussed solutions to math problems with other students in class once per week or more. Often these students may be working with other students in small groups.

The fairly high frequency of students reporting they discuss solutions to math problems with other students may be somewhat surprising, given the common perception of U.S. math instruction as teacher-centered or individuals working on their own math assignments in class. These findings may indicate there is change taking place in methods of math instruction in our schools.

Write about solving math problems

One third of eighth grade students across the nation reported that they write about how to solve math problems in class once a week or more. Applying mathematics to real-life needs and problems is a major emphasis of NCTM standards. Many states have recommended in their standards that instruction should develop students' abilities to communicate mathematically, such as by writing about how to solve a math problem. Writing about math in grade 8 classes varied from 23 percent in Indiana, Utah, and West Virginia to 58 percent in Kentucky and 50 percent in California.

Writing about solving mathematics problems is used slightly more often in grade 4 math classes than in grade 8, according to NAEP surveys with students. At grade 4, 37 percent of students report they write about solving math problems once a week or more, compared to 32 percent of eighth grade classes.

USE OF CALCULATORS IN CLASS

Nationally, 57 percent of students reported they use calculators *almost every day* in grade 8 mathematics class. Twenty states had over half their students using calculators in math class almost every day. Over 76 percent of grade 8 students use calculators at least once a week, which increased from 53 percent in 1992. States with the greatest calculator use at grade 8 in 1996 were: Alaska, Department of Defense Schools, Iowa, Michigan, Minnesota, Missouri, Montana, Nebraska, Oregon, Utah, and Wisconsin.

In 1992, only 18 percent of grade 4 students across the United States were reported by their teachers as using calculators in math class once per week or more. From 1992 to 1996 the rate increased to 34 percent. Eight states had over 50 percent of their fourth-grade students using calculators in class at least once per week in 1996.

The TIMSS teacher questionnaire in mathematics includes items about how calculators are used in class. These data may be helpful for educators to analyze in relation to teaching practices. The questions in TIMSS ask for frequency of use of calculators for the following activities: (a) checking answers; (b) tests and exams; (c) routine commutation; (d) solving complex problems; and (e) exploring number concepts.

CLASSROOM ACTIVITIES ANALYZED IN TIMSS

The students in the TIMSS study reported a wide range of classroom activities according to the responses "most lessons," "some lessons," or "never." Some of the class activities reported most often by U.S. students are shown in Figure 27.

The data on classroom activities indicate that in most lessons, teachers show students how to solve problems and students do individual "seat work" at both grades 4 and 8. Half of grade 8 students report they begin homework in class, but only one fourth are asked to apply everyday life to math.

The TIMSS teacher questionnaire asked for details about what students are expected to do in class. Teachers in most classes at both levels expect students to be able to explain the reasoning behind an idea in mathematics. Only about a third of the teachers expect students to solve problems and only one in ten expects students to work on problems that do not have an immediate solution.

FIGURE 27

CLASSROOM ACTIVITIES USED IN "MOST LESSONS"
TIMSS

Class Activity	Most Lessons Grade 4	Most Lessons Grade 8
Teacher shows how to do problems	73%	78%
Individual worksheets or textbook use	55	59
Have a quiz or test	48	39
Apply everyday life in solving problems	37	23
Begin homework in class	36	50

Source: TIMSS Student Questionnaire

TEACHER EXPECTATIONS OF STUDENTS IN "MOST LESSONS"
TIMSS

Expectation	Most Lessons Grade 4	Most Lessons Grade 8
Explain reasoning behind an idea	71%	67%
Work on problems with no immediate solution	7	12
Write equations to solve exercises/problems	28	38
Practice computational skills	70	59

Source: TIMSS Teacher Questionnaire

CONDITIONS FOR TEACHING
A KEY OPPORTUNITY TO LEARN MEASURE

A problem that is often raised by teachers with regard to improving instruction or support for their role as teachers is the need for appropriate textbooks and materials. The issue is often not only the adequacy of the materials themselves but the teacher's views on the degree of support they receive from their school or district when they request assistance. In the NAEP mathematics teacher questionnaire a question is asked about teacher views about the availability of instructional materials and resources. The NAEP teachers were asked, "how well are you provided with instructional materials and resources you need to teach?"

The results shown in Figure 28 indicate the availability of materials and resources to teachers at grade 4 for mathematics. The availability of resources does not explain whether teachers are providing quality teaching, but it does show evidence about one indicator of good teaching and high student performance. In all but one of the high-performing states, the percentage of students whose teachers receive all or most of the materials they need is significantly above the national average. The results indicate that teachers in high-performing states have more satisfaction with the availability of materials and resources than teachers in low-performing states. Four of five low-performing states are at or below the national average with respect to this statistic.

NAEP teacher questionnaires provide state-level information about a number of school conditions that may be important for teaching mathematics, including availability of a curriculum specialist, support from parents as aides, student absenteeism, and preparation time for teachers. We have focused on data about materials and resources as one key indicator which appears related to student performance and could be helpful to mathematics educators for continuing to track and monitor the important elements of high quality teaching.

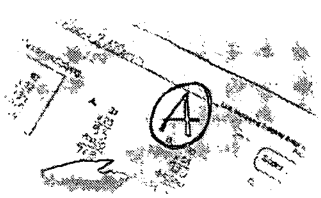
FIGURE 28
PERCENT OF STUDENTS WHOSE TEACHERS RECEIVE INSTRUCTIONAL MATERIALS AND RESOURCES THEY NEED
NAEP 1996, GRADE 4

High-Performing States	Receive Materials	
	Most/All	Some/None
CONNECTICUT	75%	25%
MINNESOTA	72	28
MAINE	67	33
WISCONSIN	80	20
NEW JERSEY	73	27
TEXAS	78	22
NATION	66	34
Low-Performing States		
MISSISSIPPI	68%	32%
LOUISIANA	59	41
CALIFORNIA	58	42
ALABAMA	64	36
SOUTH CAROLINA	66	34

Note: Standard errors from 2 to 4 percent
Source: NAEP 1996 Mathematics Cross-State Compendium

SUMMARY ON OPPORTUNITY TO LEARN

The measures for studying opportunity to learn mathematics in NAEP and TIMSS show the differences between the two data sources. TIMSS was designed to provide many different measures of differences in classrooms and teaching in mathematics and science as well as student learning in those subjects. NAEP was primarily designed to report on student achievement, and the number and types of measures of what happens in classrooms are limited. We did find that NAEP data are useful for examining differences across states on some key characteristics sensitive to policy and program change such as teacher preparation, use of calculators, and classroom materials. Alternatively, TIMSS data provide much greater depth and detail about the content of mathematics that is taught and the different types of teaching practices that are used. With TIMSS, we can see nation to nation variations, but it is not possible to report state differences. It is possible to analyze school characteristics that are important to educators, such as location, background of students, and school and class size.



Major Differences Between NAEP and TIMSS and State Assessments

To provide context for our analysis of NAEP and TIMSS results, and to relate these results to the experience and knowledge of math educators, we can look at important differences in the assessments—including their purposes, structure, and development. Common understanding about how the tests are developed and how the results are reported will help educators and policymakers to better use these large-scale test results and compare the findings with results from their state and local mathematics results. Too often we begin to read and analyze statistical findings without clear understanding and perspective on the sources of the data and what meaning can be taken from them.

PURPOSE OF THE ASSESSMENTS

The purposes need to be clearly stated and understood. Differing purposes for a student assessment study or program produce different frameworks for the content of the tests, different scope and size, different methods of selecting respondents, and varying methods of testing and collecting other research data.

NAEP. The National Assessment of Educational Progress provides a regular, periodic report on the extent of knowledge and skills of students in America's schools and to track the progress of learning—NAEP is the "Nation's Report Card." Since the inception of the NAEP under federal legislation and support in 1969, it has regularly reported on student performance in core academic subjects including mathematics, reading, science, history, arts. From the outset, NAEP tests were written from a framework developed independently from local curricula or textbooks and independently from state guidelines or standards for student learning. Subject specialists, educators, policymakers, parents and other advisers develop the NAEP "assessment framework" which provides guidance for the development of the tests, scoring, and reporting of results. A representative sample of the nation's students is assessed in grades 4, 8, and 12, and starting in 1990 representative samples of students in each state focusing on grades 4 and 8.

NAEP provides national and state-level indicators of student learning in math, disaggregation of summary scores by content topic (e.g., number sense, measurement) and a wide range of indicators of student background, teacher preparation, instruction, and school and classroom conditions. NAEP does not provide diagnostic data or curriculum analysis directly to teachers, schools, or local districts. A strength of the NAEP is its capacity for providing a high-quality comprehensive assessment of math learning which is independent of specific local curriculum textbooks, or state policies and standards.

The NAEP assessment results and supporting questionnaires from students and teachers are based on a sample of 2,000 students per state at each assessed grade. The data do not provide a way for states to analyze student achievement for each school and district. The results, however, are still extremely valuable as indicators. NAEP results provide a way to monitor state progress in student achievement; to assess education received by specific groups of students; and, very important, to determine the relationship of student achievement to characteristics of schools, classroom practices, and teachers, by state.

TIMSS. The purpose of the Third International Mathematics and Science Study was to measure the extent of student learning in mathematics and science in the 41 participating countries, and to determine the key factors in explaining differences in student learning. The study design and U.S. data collection were supported by the National Science Foundation and the U.S. Department of Education. Student testing and data collection in schools were completed during the 1994–95 school year. The TIMSS research design included a study of countries' intended, or written, curricula, the implemented curriculum, the methods of instruc-

40

tion, and key variables about teachers, schools, and students. All of the participating countries had to agree to the consensus assessment framework for mathematics and science from which the tests were written, as well as the types of items to be used and the final test instruments. A representative sample of schools and classrooms was selected at each of the three grade levels of the study with the goal of providing fair and comparable samples of students for each country. In addition to student tests, TIMSS data collection included teacher questionnaires, school questionnaires, curriculum document analysis, a videotaped classroom study, and case studies.

A main strength of TIMSS as compared to NAEP is the time and effort devoted to analysis of the curriculum and methods of instruction in mathematics and science. A detailed methodology was developed for coding, categorizing and analyzing each nation's curriculum standards, guides and textbooks. For the U.S., the analysis was based on a representative sample of documents from states. The videotaping of instruction in a sample of 8th grade mathematics classrooms in three countries (U.S., Japan, Germany) was a second key feature of the TIMSS analysis. And, third, the TIMSS questionnaires for teachers provide much greater detail and depth about what teachers cover in math and science and what instructional practices they use to teach specific content. The main TIMSS study did not provide results by states in the U.S., but several states (Colorado, Minnesota, Missouri, and Oregon) chose to conduct the TIMSS study and their results can be compared to the U.S. and other countries.

HOW DO THE PURPOSES OF NAEP AND TIMSS COMPARE TO STATE ASSESSMENTS?

Each state selects or develops their own assessment of learning in mathematics, generally under state law or mandate. In 1996–97, 45 states administered a state mathematics test to almost all students at one or more grade levels (CCSSO, *Key State Policies*, 1998a). State assessments have a variety of stated purposes, according to state directors, including accountability for schools and districts, instructional improvement, monitoring student progress and certifying students for graduation (CCSSO, SSAP 1998b). The high priority purposes that distinguish state tests are accountability, since state law requires testing and reporting the results by school and district, and in some cases by classroom or student. State assessments are given annually and except in two states include all students in selected grades. The great majority of states now produce a report for each school and district in the state, and many also provide reporting by classrooms and individual student (CCSSO, 1998c). States report their test results within six to nine months of students taking the test.

Most state assessment programs in mathematics do not have the degree of emphasis on multiple methods of assessment found in NAEP, where 50 percent of the test score is based on open-ended or constructed response questions. Recent information shows that 14 states do have some mathematics exercises requiring open-ended, extended responses and 12 states ask for short-answer responses (CCSSO, SSAP, 1998b). State programs do not typically include teacher, student, and school questionnaires, as in NAEP and TIMSS, that allow for detailed analysis or explanation of assessment results, although more states are now collecting some supporting data on instructional practices received by students.

Neither NAEP nor TIMSS have specific consequences or "high stakes" for students, teachers or schools. NAEP and TIMSS results are used by some states as an important reference point about student learning and the teacher classroom, and school background data have been used by many states.

Participation in NAEP assessments raises issues for schools. Some states report that participation in NAEP by the state or cooperation of selected schools is a problem because the administration time adds to their own high-stakes tests. Also, questions have been raised about the motivation of students to do well on NAEP and other special assessments such as TIMSS, particularly for students in grade 8 or 12 who might be aware that their scores do not reflect on their own performance or that of the school.

ASSESSMENT FRAMEWORKS

An "assessment framework" as used with NAEP and TIMSS provides guidelines for item construction and test development which define the subject content and expected student abilities or capacities.

NAEP. The NAEP Assessment Framework is developed under the National Assessment Governing Board, a federally-supported body with appointed members representing policymakers, educators, researchers, and constituents. The Mathematics Framework for 1996 (NAGB, 1996) was written by an advisory group of 15 mathematics educators and mathematicians.

The NAEP frameworks after 1990 were strongly influenced by the NCTM mathematics standards (NCTM, 1989, CCSSO, 1988), and the recent NAEP assessments incorporate more open-ended and constructed response items, in large part to match the content and expected student performance set in the assessment framework. Federal funding support has increased to match the costs of developing a high-quality test, including the development, piloting and scoring processes.

The NAEP Mathematics Assessment for 1996 included items and problems from five framework mathematics strands and two domains that cross the strands:

CONTENT STRANDS:
Number sense, Measurement, Geometry, Data/Statistics, and Algebra/Functions

DOMAINS:
Mathematical Abilities (Conceptual understanding, Procedural knowledge, Problem solving); and Mathematical Power (Reasoning, Connections, Communications)

(Reese, et al., 1997, p. 2)

TIMSS. The TIMSS Assessment Framework was developed by an international panel of scholars and educators in mathematics and science. Draft frameworks and content category descriptions were reviewed by oversight committees representing participating countries. Countries were asked to determine if their curriculum and instruction matches the Framework. Items to match the TIMSS framework were drafted and submitted by any participating country, and items were reviewed by all participating countries to determine validity of the test in relation to their curricula. The TIMSS framework for mathematics had eight Content categories and five Performance expectations categories:

CONTENT:
Numbers, Measurement, Geometry, Proportionality, Functions/relations/equations, Data representation/probability/statistics, Elementary analysis, and Validation and structure

PERFORMANCE EXPECTATIONS:
Knowing, Using routine procedures, Investigating and problem solving, Mathematical reasoning, and Communicating

(Robitaille, et al., 1993)

FIGURE 29
DISTRIBUTION OF TEST ITEMS
BY CONTENT STRAND
GRADE 8: NAEP, 1996 AND TIMSS

Strand	% of Items NAEP	TIMSS
Number Sense, Properties, Operations	25%	
Proportionality/ Ratios		7%
Fractions and Number Sense		34%
Measurement		
Measurement	12%	12%
Geometry and Spatial Sense	20%	
Geometry		15%
Data Analysis, Statistics, and Probability	15%	
Data Representation, Analysis and Probability		14%
Algebra and Functions	25%	
Algebra		18%

Sources: Reese, 1997; NCES, 1996a

The content strands for the two frameworks were quite similar. However, the distribution of items to content strands differed significantly. The TIMSS math test for grade 8 had slightly less algebra than NAEP and slightly less geometry. TIMSS placed greater emphasis in testing fractions and proportionality and ratios. The two tests had about the same emphasis on data, statistics, and probability.

Another important difference between the mathematics tests was the types of items. On the NAEP math assessment in 1996, 50 percent of the questions were multiple choice and 50 percent were constructed response, including short answer questions and problems requiring extended written responses and explanations. The NAEP items were written and scored to count toward more than one content strand, and each item was coded for the domain of math abilities or power that was tested. On the TIMSS mathematics test, 75 percent of items were multiple choice and one-fourth were constructed response, including short answer and extended response.

STATE FRAMEWORKS. Recent reports on state policies and reform initiatives (Zucker et al., 1998; CCSSO, *Key State Policies*, 1998a) indicate that states are actively working to "align" their assessment programs in mathematics and science to the new content standards developed in almost every state since 1993. A recent review of the main categories and structure of state standards and frameworks in a CCSSO study (Blank et al., 1997b) reveals that most states do have content topics and expectation for students that are modeled after the NCTM curriculum standards, and generally include the five content strands and knowledge domains outlined in NAEP and TIMSS. However, we cannot say how the assessment frameworks or the actual assessments in mathematics conducted by states match to what is set out for NAEP and TIMSS.

States may want to examine the assessment frameworks used by NAEP and TIMSS. The assessment frameworks allow the public, educators, and policymakers to determine how the assessment will be written—usually providing much more specificity than either state standards or local curriculum.

The process of aligning assessments to state standards is an important task being carried out by many states, and the assessment framework step provides a way to focus the test development effort toward each priority of the standards, both in breadth and depth of coverage. In fact, alignment of assessments to standards has gained a new level of importance as state content standards have been approved as the means for assuring accountability of schools and districts. More detailed, systematic approaches to alignment analysis that can be applied to state standards are being tested with states by CCSSO (Webb, 1997) and by Achieve.

REPORTING ASSESSMENT RESULTS

NAEP assessment results are initially released to the public, press, states, and others in the NAEP Report Card. It includes the test results for the nation and states for three grade levels and a report for each state for the grades tested, (e.g., grades 4 and 8 in 1996). The Report Card is released by NCES about one year after the test administration. The results are disaggregated by student demographic characteristics. Subsequent reports, such as the NAEP 1996 *Mathematics Cross-State Data Compendium* (Shaughnessy, et al., 1997), have provided data by state for all of the questionnaire items and further disaggregation of test results. Other research reports and analyses for NAEP are supported and produced by NCES.

TIMSS test results were released to the public and participating countries for math and science separately by grade level: eighth grade in Fall 1996, fourth grade in spring 1997, and twelfth grade in Fall 1997. The international reports (e.g., Beaton, et al., 1996) were released simultaneously with a report focusing on U.S. results from NCES (e.g., 1996a). Separate reports were produced and released in the TIMSS curriculum documents analysis, *Many Visions, Many Aims* (Schmidt, McKnight, et al., 1996), video study, and case studies (see NCES website for reports list, www.ed.gov/NCES). Data from questionnaires and test item results were made available on the Internet and by compact disc.

State assessment programs are typically instruments for accountability of the education system. Test results are reported by school for purposes of accountability in 37 states (CCSSO, SSAP, 1998). These reports are aimed at examining the degree of education progress of each school. State tests also have implications for individual students. In 21 states high school students must pass a state assessment or "exit exam" prior to graduation (CCSSO, *Key State Policies*, 1998). High priorities for states are rapid scoring of the items, analysis and reporting that can be provided to policymakers and the public efficiently, and methods for reporting and tracking progress of schools and districts.

In 11 states, school-level test scores are used in the school accreditation process and 10 states provide school awards or recognition (CCSSO, SSAP, 1998). Additionally, over 20 states have planned program interventions or sanctions for schools that do not show improvement in student achievement over a multi-year period. State assessment programs place emphasis on rapid turnaround so data can be used by educators at all levels, and they are concerned about holding down costs per student particularly because state programs emphasize testing the universe of students. A new trend in state assessment scoring and reporting is to indicate the specific subject content or standard for learning that has been assessed and how schools, classrooms, and students performed against each standard.

USE OF ACHIEVEMENT/ PROFICIENCY LEVELS IN REPORTING

NAEP achievement levels are descriptions of what students should know and be able to do in mathematics at each grade level. Under supervision of the National Assessment Governing Board, three levels were defined for each grade level—Basic, Proficient, and Advanced. A group of 75 educators, citizens and mathematicians participate in the level-setting process by rating the kinds of assessment exercises given in NAEP for what students should know and be able to do. The NAEP achievement levels are set prior to the test and they are established separately from the items and student scores for a given year's assessment. Each of the achievement levels is defined more narrowly for each grade level, in terms of the particular mathematical concepts and skills that apply. For example, below is the definition of the Basic level for grade 8:

> Eighth grade students performing at the Basic level should exhibit evidence of conceptual and procedural understanding in the five NAEP content strands. This level of performance signifies an understanding of arithmetic operations—including estimation—on whole numbers, decimals, fractions, and percent.
>
> Eighth graders performing at the Basic level should complete problems correctly with the help of structural prompts such as diagrams, charts, and graphs. They should be able to solve problems in all NAEP content strands through the appropriate selection and use of strategies and technological tools–including calculators, computers, and geometric shapes. Students at this level also should be able to use fundamental algebraic and informal geometric concepts in problem solving.
>
> As they approach the Proficient level, students at the Basic level should be able to determine which of the available data are necessary and sufficient for correct solutions and use them in problem solving. However, these eighth graders show limited skill in communicating mathematically (Reese, et al., 1997).

Several major advantages are offered in using the NAEP levels. First, NAEP scores are more understandable and interpretable by educators and the public when reported according to written standards for what is expected of students at a given grade level. We know what percentage of students meet expected standards and what percentage are still below the standard. The achievement levels are widely used in reporting and analyzing NAEP results. CCSSO incorporated the NAEP levels in reporting state mathematics and science indicators starting in 1993, and other organizations such as the National Education Goals Panel have used the percentage of students meeting these levels as key indicators.

The NAEP scale is a composite of the five content strands measured in the mathematics assessment. Student responses to each question are analyzed to determine the percentage of students responding correctly (for multiple-choice) and the percentage of students responding in each of the score categories (for regular and extended constructed response). NAEP uses item response theory (IRT) scaling methods to produce an overall composite mathematics score for the nation and each state by grade level, and scale scores are produced for each math content strand. TIMSS used IRT scaling to produce overall mathematics and science scale scores for each participating country. The IRT scaling method produces a score by averaging the responses of each student, taking into account the difficulty of each item. NAEP results are reported on a composite scale from 0 to 500 that allows comparisons of scores by state, type of school, or a variety of other disaggregations. TIMSS results are reported for each grade level on a composite scale from 0 to 800. TIMSS also reported each country's "average percent correct" for each separate content category on the Mathematics test, such as fractions and proportionality.

IRT scaling and matrix sampling of items for students allow NAEP and TIMSS to test a wide range of mathematics content by including many more items in the total assessment pool. NAEP emphasizes reporting of student performance for the nation, by state, and a for variety of subpopulations, in relation to proficiency levels for expected mathematics knowledge and skills. Scale scores provide efficient comparisons across groups but also can be linked to expected mathematics knowledge and ability. In 1996 the NAEP mathematics report provided a unique display of scale score results by showing the kinds of math problems a student could do at any given scale score. The graphic is reproduced in the Appendix.

Large-scale student tests typically have been scored and reported using a "norm-referenced" approach in which student scores are mainly interpreted relative to the performance of other students, other groups of students, or other states or types of schools. The "norms" for such a test are generally set by testing a national representative sample of students. Any group or individual can be compared to this sample score.

States have moved toward use of achievement, or proficiency, levels for state assessment programs. As of the 1996–97 school year, 39 of the 45 states with state mathematics assessments had defined performance levels for scoring and reporting (CCSSO, SSAP, 1998). The states move toward use of performance levels is encouraged by requirements for accountability under federal Title I law. By 2000 all states will need to show the relationship between their content standards for mathematics and reading and the state assessment and each state will need to report scores using three or more reporting levels to track progress. The intended focus of Title I programs and accountability is to improve the performance of schools that serve low-income students—the focus of Title I funds.

A significant advantage of performance levels is placing focus on the proportion of students in a school, district, or state that have met a set level of expected performance in the subject area, and then monitoring the extent to which the percentage of students meeting the level increases over time. This is a "growth-based" model for analyzing assessment results rather than an analysis based on absolute score or relative scores. The state can determine whether students are accomplishing expected knowledge and skills, and whether performance schools is improving. A focus on *growth* over time, using set performance standards for all students and schools redirects priority and function to an accountability system and assessment results and away from the simple use of scores to rank or rate schools, districts, or students within the system.

SAMPLING VS. ALL STUDENTS

NAEP and TIMSS results are based on scores from a representative sample of students in each participating entity—state or nation. Both use matrix sampling and up to eight different test booklets, thus allowing better coverage of the full mathematics framework. NAEP and TIMSS are much more likely to cover the content and expected abilities and skills called for in the assessment framework than traditional test designs and many of the current state assessments. The scoring and scaling of results provides weighting of scores for items with different complexity, difficulty, and item design, e.g., constructed response vs. multiple-choice.

With matrix sampling, students are administered different combinations of test items with the versions matched on items difficulty, content, and design. For example, the NAEP scores that are produced for a content strand such as algebra are an aggregation of student answers across all the different test versions. Matrix sampling increases the validity and reliability of state and national scores in relation to the assessment framework. The main disadvantage is that individual student, classroom, and school results cannot be produced and scores cannot be compared at these levels.

Implications For Mathematics Education

From the analysis given here of the results from the most recent administration of NAEP, the state level NAEP, and TIMSS, we can find several important messages. We have seen that there has been some improvement over time in the mathematics achievement of U.S. students, and a number of states are showing growth in student achievement. Yet the international picture is sobering, and reminds us that we are far from the goal of enabling all of our students to experience a quality education in mathematics. Some of the specific results noted in this analysis point to several major areas that need improvement. We can categorize those areas as: changing the emphasis on *what is taught*, attending to *how it is taught*, and improving *the preparation of teachers*.

WHAT IS TAUGHT: THE NEED FOR MORE DEPTH IN CONTENT

While basic computational skills seem to be fairly strong, the curriculum at all grades needs to put more emphasis on measurement, number sense, geometry, and proportionality. The critical factor in all of this is that the topics should not be taught in a shallow, "scattershot" approach. Fewer topics should be taught at more depth, at all levels, and with less repetition. This is especially true for the middle grades. Performance expectations need to be set higher at all levels.

Many traditional textbook series in the U.S. have placed too much emphasis on arithmetic skills as the primary target of the curriculum in grades K–8. Newer curricula, such as those developed with funding from the National Science Foundation, integrate mathematics topics from across the curriculum, emphasizing conceptual understanding. Many of these materials embed the mathematics in real contexts, where the required skills and procedures are learned as tools for solving non-routine problems.

Finally, NAEP results show clearly that the mathematics course that students receive is directly related to their performance. Students that are taught a more challenging curriculum in middle and high school mathematics reach higher levels of achievement .

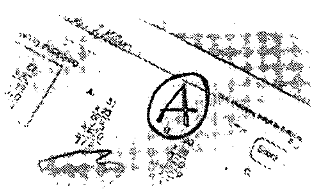
HOW IT IS TAUGHT: THE NEED FOR MORE MATHEMATICAL PROCESSES AND DEPTH OF UNDERSTANDING

Students are reasonably successful with basic, one-step problems and routine procedures. What they need are more experiences with non-routine problem solving situations and more opportunities to apply their skills in real-world contexts. Teachers need to choose more tasks for students that require mathematical reasoning, making conjectures, and justifying their answers. Students need more opportunities to learn to communicate mathematically, through listening, speaking, arguing, writing, reading, and explaining. Communication skills should be central to the activities in every mathematics classroom, and not simply relegated to the ubiquitous direction of "show your work." In addition, the emphasis should be on understanding concepts, rather than memorizing procedures.

TEACHER PREPARATION

If teachers are to choose the kinds of tasks described above for their students, that is, tasks that demand solving contextualized, novel problems, mathematical reasoning, and mathematical communication, the teachers' knowledge demands are much greater than if the teacher merely demonstrates to students how to carry out routine procedures. Teachers must have sufficient depth and breadth in content knowledge to feel comfortable facilitating discussions about mathematical concepts. Often such discussions take both teachers and students into new and perhaps unfamiliar mathematical territory. A teacher whose own background and knowledge of mathematics is shallow or uneven will not feel confident about such explorations, and will tend to retreat to more familiar and less demanding tasks. Therefore, the professional development of teachers is a major concern in bringing about these changes in the way mathematics is taught.

The quality of professional development opportunities is key. Rather than attending one-day "make and take" workshops that are designed only to offer "fun" activities, teachers need ongoing, serious work in developing breadth and depth in content knowledge and pedagogical skills. They need opportunities to examine student work on extended constructed response tasks, and to discuss with other teachers how to incorporate more communication and higher order thinking into their courses. On a regular basis teachers should be discussing teaching and learning with their colleagues, by visiting other classes, examining student products, and viewing videos of classrooms. The released tasks and scoring guides from NAEP and the toolkit from TIMSS can be valuable sources of materials for professional development.



Conclusion

Educators and policymakers are constantly faced with findings from yet another national, state, or international study that calls the public's attention to the relative failings or successes of our schools. State assessment programs, NAEP, and TIMSS represent major investments in education research and public accountability for K–12 education. They bring a powerful focus on central questions about the health of our education system. Teachers and local educators, especially, may be suspicious of the findings of these studies because they receive significant attention for short periods of time without sustained followup for educators, and because they often are not accompanied with details on how the studies were conducted or with relevant information that can be applied to day-to-day teaching and learning.

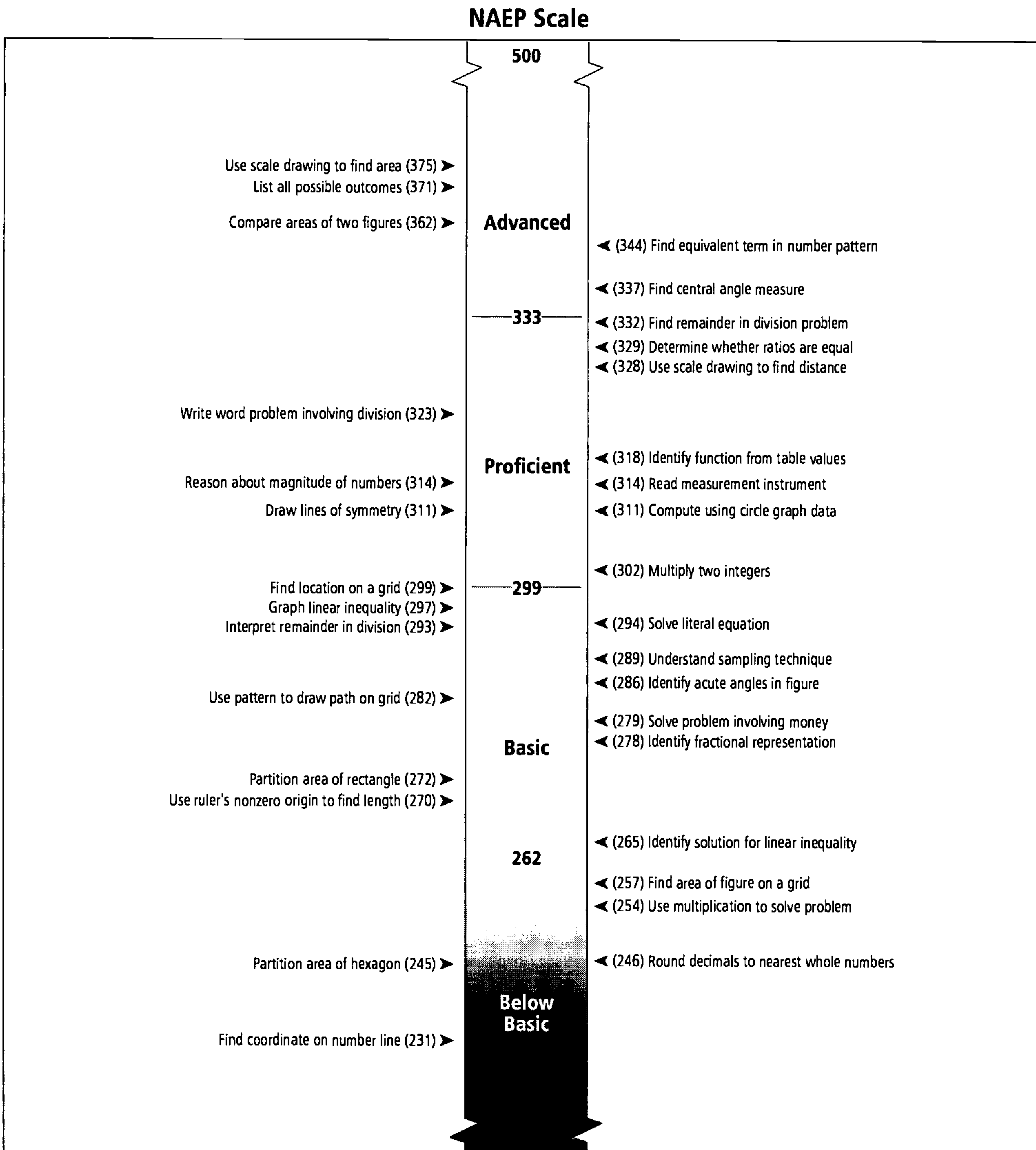
This paper has demonstrated that a more complete and detailed picture of mathematics education in the U.S. can be obtained by studying the details and meaning behind the averages and overall trends typically reported about NAEP and TIMSS. We have shown how educators' decisions about what and how to teach makes a great difference in student performance. We also have shown how averages and national and state summary statistics can mask both improvements and problems. For example, our research shows that improvement on NAEP mathematics assessment shown in student knowledge with number sense and operations must be weighed against poor proficiency in measurement and geometry. Similarly, the difficulties U.S. students have shown in answering non-routine problems requiring open-ended responses should be faced in light of the significant gains by students at all achievement levels in basic operations and procedures. The analysis and interpretation of student difficulties with the selected NAEP and TIMSS exercises are the kind of item analysis that will help highlight instructional improvements. The sample exercises provide examples of the kinds of broad examination of student knowledge and abilities that are offered in these assessments, and illustrate ways that classroom assessments and local accountability tests can be improved.

The analysis and comparison of U.S. results on NAEP mathematics and TIMSS help educators see key differences between the achievement tests and accompanying data for interpreting the findings. We have shown some differences between the assessment frameworks and the composition of items. Also we have examined the use of achievement levels and scale scores for reporting NAEP assessments, and shown some of the advantages and disadvantages of how the results are reported. It is important for state and local educators to be able to compare how their own tests are constructed and reported in relation to NAEP and TIMSS. The results are not generally directly comparable, but it is critical to see how the methods and emphases of these tests can be explained in relation to other results used by educators.

We have illustrated some methods of analyzing and disaggregating NAEP mathematics results to examine educational progress for different student populations and schools and classrooms with varying characteristics. Patterns of performance vary widely at one time, and extent of improvement over time differs. Expertise in monitoring trends for key target groups is required to make effective use of sample-based data. Issues of opportunity-to-learn are critical in the current era of standards-driven education reform. For all students to attain to challenging standards for mathematics requires analysis of differences in opportunities in curriculum, teaching, and teacher preparation. Adequate measures of how students are offered opportunities in mathematics will be needed to plan K–12 programs aimed toward high standards. NAEP and TIMSS offer excellent resources of measures of mathematics opportunity to learn. We observed limitations of each study, and also identified fruitful measures that could be incorporated into assessment programs and evaluation studies at local and state levels.

Finally, we have offered a list of improvements that we believe are implied by the combined results of NAEP and TIMSS. We have presented specific suggestions in three critical areas of mathematics education: curriculum, teaching, and professional development. There are some clear messages from these two assessments that point to specific areas in need of attention. We believe that all U.S. students deserve the chance to experience a quality mathematics curriculum that is taught in a way that promotes understanding. To achieve this goal will mean making some changes to what mathematics is taught, how it is taught, and how teachers are supported.

Appendix:
Map of Selected Questions on the NAEP Mathematics Scale for Grade 8



Note : Position of questions is approximate and an appropriate scale range is displayed for grade 8.

Source: NAEP 1996 Mathematics Report Card

50

References

American Association for the Advancement of Science (1993). *Benchmarks for Science Literacy*. New York: Oxford University Press.

Barton, P.E. & Coley, R.J. (1998). *Growth in School: Achievement Gains from the Fourth to the Eighth Grade*. Princeton, NJ: Educational Testing Service, Policy Information Center.

Beaton, A.E., Mullis, I.V.S., Martin, M.O., Gonzalez, E.J., Kelly, D.L., & Smith, T.A. (1996, 1997, 1998). *Mathematics Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study (TIMSS), Mathematics Achievement in the Primary School Year, Mathematics and Science Achievement in the Final Year of Secondary School*. Chestnut Hill, MA: Boston College.

Blank, R.K. & Langesen, D. (1997A) *State Indicators of Science and Mathematics Education 1997. State-by-State Trends and New Indicators from the 1995-96 School Year*. Washington, DC: CCSSO, State Education Assessment Center.

Blank, R.K., Langesen, D., Bush, M., Sardina, S., Pechman, E., & Goldstein, D. (1997B). *Mathematics and Science Content Standards and Curriculum Frameworks. States Progress on Development and Implementation*. Washington, DC: CCSSO, State Education Assessment Center.

Blank, R.K., Manise, J.G., & Brathwaite, B.C. (1997C). *State Education Indicators with a Focus on Title I*. Washington, DC: CCSSO, State Education Assessment Center.

Campbell, J.R., Voelkl, K.E., & Donahue, P.L. (1997). *NAEP 1996 Trends in Academic Progress*. Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Carpenter, T., Corbitt, M.., Kepner, H., Lindquist, M., & Reys, R. (1981). *Results from the Second Mathematics Assessment of the National Assessment of Educational Progress*. Reston, VA: National Council of Teachers of Mathematics.

Council of Chief State School Officers, State Education Assessment Center (1988). *Final Report: Assessing Mathematics in 1990 by the National Assessment of Educational Progress*. Washington, DC: Author.

Council of Chief State School Officers, SCASS Science Education (1997). *How is Science Taught in Schools? A Five-State Preliminary Study on the Use of Enacted Curriculum Surveys in Science*. Washington, DC: Author.

Council of Chief State School Officers, State Education Assessment Center (1998A). *Key State Education Policies on K-12 Education*. Washington, DC: Author.

Council of Chief State School Officers, State Education Assessment Center (1998B). *Annual Survey of State Student Assessment Programs, 1996-97 School Year*. Washington, DC: Author.

Council of Chief State School Officers, State Education Assessment Center (1998C). *State Education Accountability Reports and Indicator Reports: Status of Reports Across the States 1998*. Washington, DC: Author.

Darling-Hammond, L. (1995). The Role of Teacher Expertise and Experience in Students' Opportunity to Learn, pp. 19-24, *Strategies for Linking School Finance and Students' Opportunity to Learn*, edited by Patricia Brown. Washington, DC: National Governors' Association.

Dossey, J., Mullis, I., & Jones, C. (1993). *Can Students Do Mathematical Problem-Solving? Results from Constructed Response Questions in NAEP's 1992 Mathematics Assessment*. Washington: DC: U.S. Department of Education, National Center for Education Statistics.

Education Trust (1998) *Education Watch, 1998 State and National Data Book, Vol. II*. Washington, DC: Author.

Elliott, M. (1998). School Finance and Opportunities to Learn: Does Money Well Spent Enhance Students' Achievement? *Sociology of Education,* Vol. 71, Number 3.

Gamoran, A., Porter, A.C., Smithson, J., & White, P.A. (1997). Upgrading High School Mathematics Instruction: Improving Learning Opportunities for Low-Achieving, Low-Income Youth. *Educational Evaluation and Policy Analysis,* Vol. 19, Number 4.

Goodlad, J.I. (1984). *A Place Called School: Prospects for the Future*. New York: McGraw-Hill.

Grissmer, D. & Flanagan, A. (1998) Improving the Data and Methodologies in Educational Research. Paper presented at USED/Rand conference on Analytical Issues in the Assessment of Student Achievement. Washington, DC: RAND.

Hoffer, T.B., Rasinski, K.A., & Moore, W.M. (1995). *Social Background Differences in High School Mathematics and Science Course Taking and Achievement.* Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Horn, L. & Hafner, A. (1992). *A Profile of American Eighth-Grade Mathematics and Science Instruction.* Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Husen, T. (Ed.) (1967). *International Study of Achievement in Mathematics: A Comparison of Twelve Countries* (Vols. 1 and 2). New York: John Wiley.

Ingersoll, R.M. & Gruber, K (1996) *Out-of Field Teaching and Educational Quality.* Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Jones, L.V., Davenport, E.C., Btyson, 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.

Kenney, P. & Silver, E. (1997). *Results from the Sixth Mathematics Assessment of the National Assessment of Educational Progress.* Reston, VA: National Council of Teachers of Mathematics.

Lee, V.E., Bryk, A.S., & Smith, J.B. (1993). The Effects of High School Organization Influences the Equitable Distribution of Learning in Mathematics and Science. *Sociology of Education*, 70 (2). Washington, DC: The American Sociological Association.

McDonnell, L. (1995). Opportunity to Learn as a Research Concept and a Policy Instrument, *Educational Evaluation and Policy Analysis*, 17:305-22.

McKnight, C.C., Crosswhite, F.J., Dossey, J.A., Kifer, E., Swafford, J.O., Travers, K.J., & Cooney, T.J. (1987). *The Underachieving Curriculum: Assessing U.S. School Mathematics from an International Perspective.* Champaign, IL: Stipes Publishing.

Monk, D.H. (1993). *Subject Area Preparation of Secondary Mathematics and Science Teachers and Student Achievement.* Ithaca, NY: Cornell University, Department of Education.

Mullis, I.V.S., Dossey, J.A., Owen, E.H., & Phillips, G.W. (1993). *NAEP 1992 Math Report Card for the National and the States.* Washington, DC: U.S. Department of Education, National Center for Education Statistics.

National Assessment of Governing Board (1996). *Mathematics Framework for 1996 NAEP.* Washington, DC: U.S. Department of Education.

National Center for Education Statistics (1996a). *Pursuing Excellence: A Study of U.S. Eighth-Grade Mathematics and Science Teaching, Learning, Curriculum, and Achievement in International Context. Initial Findings from the Third International Mathematics and Science Study.* Washington, DC: U.S. Department of Education.

National Center for Education Statistics (1996B). *SASS by State, 1993-94 Schools and Staffing Survey.* Washington, DC: U.S. Department of Education.

National Center for Education Statistics (1997A). Pursuing Excellence: A Study of U.S. Fourth-Grade Mathematics and Science Achievement in International Context. Initial Findings from the Third International Mathematics and Science Study. Washington, DC: U.S. Department of Education.

National Center for Education Statistics (1997B). *Toward Better Teaching: Professional Development in 1993-94.* Washington, DC: U.S. Department of Education.

National Center for Education Statistics (1997B). *Condition of Education.* Washington, DC: U.S. Department of Education.

National Center for Education Statistics (1997D). *Videotape Classroom Study.* Washington, DC: U.S. Department of Education.

National Center for Education Statistics (1998A). *Pursuing Excellence: A Study of U.S. Twelfth-Grade Mathematics and Science Achievement in International Context. Initial Findings from the Third International Mathematics and Science Study.* Washington, DC: U.S. Department of Education.

National Center for Education Statistics (1998B). *Linking the National Assessment of Educational Progress and the Third International Mathematics and Science Study: Eighth-Grade Results.* Washington, DC: U.S. Department of Education.

National Council of Teachers of Mathematics (1989). *Curriculum and Evaluation Standards for School Mathematics.* Reston, VA: Author.

National Council of Teachers of Mathematics (1991). *Professional Standards for Teaching Mathematics*. Reston, VA: Author.

National Education Goals Panel (1998). *Mathematics and Science Achievement State by State, 1988*. Washington, DC: U.S. Government Printing Office.

National Research Council (1995). *National Science Education Standards*. Washington, DC: National Academy Press.

Oakes, J. (1990). *Multiplying Inequalities: The Effects of Race, Social Class, and Tracking on Opportunities to Learn Mathematics and Science*. Santa Monica, CA: The Rand Corporation.

Porter, A.C. (1995). *Developing Opportunity-to-Learn Indicators of the Content of Instruction: Progress Report*. Madison, WI: Wisconsin Center for Education Research.

Raudenbush, S.W. (1998) Synthesizing Results from the Trial State Assessment. Paper presented at USED/Rand conference on Analytical Issues in the Assessment of Student Achievement. Ann Arbor, MI: University of Michigan, School of Education.

Reese, C.M., Miller, K.E., Mazzeo, J., & Dossey, J.A. (1997). *NAEP 1996 Mathematics Report Card for the Nation and the States*. Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Robitaille, D.F., Schmidt, W.H., Raizen, S., McKnight, C., Britton, E., & Nicol, C. (1993). *Curriculum Frameworks for Mathematics and Science*. Vancouver, Canada: The Third International Mathematics and Science Study.

Rock, D., Braun, H.I., & Rosenbaum, P.R. (1985). *Excellence in High School Education: Cross-Sectional Study, 1980-1982*. (Final Report). Princeton, NJ: Educational Testing Service.

Schmidt, Cogan, et. al. (1998). Unpublished data from TIMSS, Teacher Questionnaire, Populations 1 and 2.

Schmidt, W.H., McKnight, C.C., Valverde, G.A., Houang, R.T., & Wiley, D.E. (1996). *Many Visions, Many Aims. A Cross-National Investigation of Curricular Intentions in School Mathematics*. Kluwer Academic Publishers.

Sebring, P.A. (1987). Consequences of Differential Amounts of High School Coursework: Will the New Graduation Requirements Help? *Education Evaluation and Policy Analysis*, Vol. 9, Number 3.

Shaughnessy, C.A., Nelson, J.E., & Norris, N.A. (1997). *NAEP 1996 Mathematics Cross-State Data Compendium for the Grade 4 and Grade 8 Assessment. Findings from the State Assessment in Mathematics of the National Assessment of Educational Progress*. Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Silver, E. A., Alacaci, C. & Stylianou, D. (1998, in press). Students' Performance on Extended Constructed-Response Taskzs. In Edward A. Silver and Patricia Ann Kenney, Eds. *Results from the Seventh Mathematics Assessment of the National Assessment of Educational Progress*. Reston, VA: National Council of Teachers of Mathematics.

Shavelson, R., McDonnell, L., Oakes, J., & Carey, N. (1989). *Indicatzt*, C.C. (1985). Mathematics Achievement in U.S. Schools: Preliminary Findings from the Second IEA Mathematics Study, *Phi Delta Kappan*, pp. 407-413.

Walberg, H.J. (1984). Improving the Productivity of American Schools. *Educational Leadership*, 41(8).

Webb, N. (1997). *Criteria for Alignment of Curricular Expectations and Assessments in Mathematics and Science Education*. Washington, DC: CCSSO; National Institute for Science Education, University of Wisconsin-Madison.

Weiss, I.R. (1994). *A Profile of Science and Mathematics Education in the United States: 1993*. Chapel Hill, NC: Horizon Research, Inc.

Zucker, A., Shields, P., Humphrey, D., & Adelman (1998). *SSI Strategies for Reform: Findings from the Evaluation of NSF's SSI Program*. Palo Alto: SRI International.

References for Sample Tasks

"ODOMETER"

Mitchell, J., Hawkins, E., Jakwerth, P., Stancavage, F., Dossey, J. (draft March 20, 1998). *Student Work and Classroom Practices in Mathematics,* p. 76. NCES.

"SAM'S LUNCH"

Mitchell, J., Hawkins, E., Jakwerth, P., Stancavage, F., Dossey, J. (draft March 20, 1998). *Student Work and Classroom Practices in Mathematics,* p. 57. NCES.

"TWO GEOMETRIC SHAPES"

Released item from the NAEP *1996 Mathematics Report Card for the Nation and the States: Findings from the National Assessment of Educational Progress (4th grade).* Washington, D.C.: Office of Educational Research and Improvement, U.S. Department of Education.

"BICYCLE ACCIDENTS"

Business Coalition for Educational Reform (May, 1998). *Formula for Success: A Business Leader's Guide to Supporting Math and Science Achievement,* p. 9. Washington, D.C.: U.S. Department of Education.

"GEOMETRIC PROOF"

Takahira, S., Gonzales, P., Frase, M., & Salganik, L. (1998). *Pursuing Excellence: A Study of U.S. Twelfth-grade Mathematics and Science Achievement in International Context, p. 43.* Washington, D.C.: U.S. Department of Education.

"AMOUNT PAID FOR PORTION OF ITEMS"

Beaton, A., Mullis, I., Martin, M., Gonzalez, E., Kelly, D., & Smith, T. (November, 1996). *Mathematics achievement in the middle school years: IEA's Third International Mathematics and Science Study (TIMSS),* p. 98. Boston: Center for the Study of Testing, Evaluation, and Educational Policy, Boston College.

"CHERRY SYRUP"

Reese, C., Miller, K., Mazzeo, J., & Dossey, J. (February, 1997). NAEP *1996 Mathematics Report Card for the Nation and the States: Findings from the National Assessment of Educational Progress,* p.22. Washington, D.C.: Office of Educational Research and Improvement, U.S. Department of Education.

"MEANING OF EQUATION"

Mullis, I. (Draft, February 27, 1998). *TIMSS as a Starting Point to Examine Mathematics Assessments: An in-depth Look at Geometry and Algebra,* p. 95. Washington, D.C.: Office of Educational Research and Improvement, U.S. Department of Education.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC®

TM029827

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: Improving Mathematics Education Using Results from NAEP and TIMSS	
Author(s): Linda Dager Wilson and Rolf K. Blank	
Corporate Source: Council of Chief State School Officers	Publication Date: 1999

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 1	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2A	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2B
Level 1	Level 2A	Level 2B
☑	☐	☐
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other then ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, → please

Signature: Rolf K Blank	Printed Name/Position/Title: Rolf K Blank, Dir of Ed Indicators	
Organization/Address: CCSSO 1 Mass Ave NW Washington DC 20001	Telephone: 202 336 2044	FAX:
	E-Mail Address: rolfb@ccsso.org	Date: 5/11/99

(over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, *or*, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

THE UNIVERSITY OF MARYLAND
ERIC CLEARINGHOUSE ON ASSESSMENT AND EVALUATION
1129 SHRIVER LAB, CAMPUS DRIVE
COLLEGE PARK, MD 20742-5701
Attn: Acquisitions

Send

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
1100 West Street, 2nd Floor
Laurel, Maryland 20707-3598

Telephone: 301-497-4080
Toll Free: 800-799-3742
FAX: 301-953-0263
e-mail: ericfac@inet.ed.gov
WWW: http://ericfac.piccard.csc.com