DOCUMENT RESUME

ED 279 696 TM 870 083

AUTHOR Romberg, Thomas A.
TITLE National Assessment of Mathematical Performance.
PUB DATE 18 Aug 86
NOTE 46p.; One of 46 papers commissioned by the Study Group on the National Assessment of Student Achievement and cited in Appendix B to their final report "The Nation's Report Card" (TM 870 049). For other papers in this group, see TM 870 050-094.
PUB TYPE Viewpoints (120)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Comparative Testing; *Educational Assessment; Elementary Secondary Education; Epistemology; *Mathematics Achievement; *Mathematics Tests; *Measurement Objectives; *National Surveys; Profiles; Test Format; Testing Problems; *Testing Programs; Test Results
IDENTIFIERS *National Assessment of Educational Progress

ABSTRACT

Four topics related to mathematics testing for the National Assessment of Educational Progress (NAEP) are discussed: (1) mathematics achievement; (2) past NAEP approaches; (3) needed changes; and (4) a new conceptual basis for profiling mathematical performance, including recommendations for strengthening current practice. The basic strategy for gathering profile information for students at several age (grade) levels used in the past assessments has been reasonable and has yielded useful information, affecting school mathematics. Current procedures could be improved by discarding content-by-behavior matrices and by adopting a network model, such as conceptual fields. The types of exercises included in the batteries should be expanded to reflect the network model. New contexts should be included so that the construction of knowledge can be assessed. The sampling base can be changed so that data can be gathered for state comparisons. From the rich data such an improved assessment would yield, it should be possible to construct reasonable indicators for use by policymakers. The final four recommendations propose that: (1) work be initiated to identify major conceptual fields in mathematics; (2) future assessments encourage the development of a variety of alternate items and testing formats; (3) the data base be increased so that it reflects current expectations about how students construct mathematical knowledge to build a theoretical model of mathematical performance; and (4) reasonable indicators be constructed from that model for policy purposes. (An extensive list of references is appended.) (GDC)

NATIONAL ASSESSMENT OF MATHEMATICAL PERFORMANCE

by

Thomas A. Romberg
Professor of Curriculum & Instruction
(Mathematics Education)
Faculty Associate, Wisconsin Center for Education Research
University of Wisconsin-Madison

Paper prepared for: Study Group on National Assessment
Honorable Lamar Alexander
Governor of Tennessee, Chair
United States Department of Education
August 1986
(Revised April 1987)

Paper commissioned by

THE STUDY GROUP ON THE NATIONAL ASSESSMENT OF STUDENT ACHIEVEMENT
1986

INTRODUCTION

On May 14, 1986, Secretary of Education William J. Bennett and Tennessee Governor Lamar Alexander announced the formation of a Study Group for National Assessment. The group's purpose was to study and propose ways to strengthen national assessment of student achievement. The purpose of this paper is to examine national assessment of mathematical performance for the Study Group. The paper includes four parts to review past national assessment strategies for mathematics and suggest ways of strengthening them. First, a description of mathematical achievement is given. Second, there is a brief examination of past approaches to national assessment of mathematics. Third, the rationale for a changed or refocused intent in light of current needs is presented. Finally, a new conceptual basis for profiling mathematical performance has been outlined. Included are recommendations for strengthening current practice.

MATHEMATICS ACHIEVEMENT

The purpose of this section is to describe both what is meant by achievement and what methods of assessing mathematical performance are appropriate for national policy purposes.

Achievement

Achievement can be considered as the reasonable pupil outcomes following a set of instructional experiences in school courses. Detailing what those outcomes are is of necessity quite complex. However, at least acquisition of concepts and skills, maintenance of those concepts and skills, preparation for new concepts and skills, acquisition of a positive attitude toward mathematics, and use of concepts and skills to solve problems should be included.

Academic achievement is a subset of achievement associated with academic courses. Such courses are in contrast to, for example, vocational, technical, and physical education courses. The concepts and skills of academic courses are associated with subject-matter disciplines (language arts, mathematics, physics, . . .). The goals of such courses not only emphasize acquisition and maintenance of concepts and skills but, in particular, stress preparation for later study in the subject area in higher grades and then even later use of that knowledge, in various occupations.

For national assessment both the level and variability for a diverse set of academic outcomes for students at certain age levels should be assessed, as should the students' readiness to use what they have learned.

Methods of assessment

Not only is the question of what outcomes should be examined quite complex, but also we must ask the difficult question of how to elicit the information needed. The "units" about which the decision is to be made for national assessment are groups such as classes and schools, not individuals. Thus, the measurement procedures and decision rules to be used must involve specifying, to best estimate a group's performance on a diverse set of outcomes, the sources, the scaling procedure, the reliability, and the validity of the measurement process.

The most common method of gathering information about mathematics achievement is administering paper-and-pencil tests to groups of students. Although other procedures (interviews, observations, judgements about work samples) could be used, the ease of development, the convenience, and low cost of such group testing has made paper-and-pencil tests common in American schools. In addition, to validly span the scope of reasonable outcomes at any age level, multiple matrix sampling is commonly used to estimate a group's performance. Matrix sampling yields a profile of scores for each group assessed. Furthermore, if the assessments are repeated over two or more time intervals, growth curves can be plotted and compared.

Summary

It is clear that the purpose of National Assessment should be to provide educators and policymakers with profiles of mathematics achievement for groups of students over several time periods. Such profiles are of necessity complex, because achievement involves a variety of different outcomes. In addition, measures of performance should be related to what has actually been taught or what is expected to be taught in classrooms and whether what has been learned can be used by students. Finally, repeated assessment is important so that the effects of change in policy and practice can be determined.

However, the combination of assessing what is taught and conducting repeated measurements has created a major problem. During the past decade there has been not only a shift in the mathematical concepts and skills that are important, but a shift in emphasis from acquiring a large number of concepts and calculation routines toward estimating, conjecturing, and problem solving strategies. Such a shift in what is expected to be taught suggests that the next assessments should reflect these intended changes. In fact, the tests must change or they will be inhibitors to accomplishing such change. At the same time elements of past tests must be retained so that growth can be observed. Thus, national assessment must reflect both the attainment of what is new and changes in the attainment of what is still important.

THE PAST NATIONAL ASSESSMENT ACTIVITIES

When the United States Office of Education was founded in 1867, one charge set before its commissioner was to determine the nation's progress in education. That century-old charge was not answered systematically in the United States until 1972-73 when the first National Assessment of Education Progress (NAEP) in mathematics was carried out. Nationally based mathematics testing had been previously done, but not by the Office of Education. Both standardized tests and profile tests had been given before NAEP was first administered. To summarize past activities, both standardized achievement tests and other profile tests are discussed prior to discussion of the activities of NAEP. This section closes with a brief outline of the Assessment of Performance Unit, the national assessment project in the United Kingdom.

Standardized Achievement Tests

Ever since standardized tests have been given,[1] normative data have been gathered which each test publisher claims is representative of the national population. However, there are several reasons for arguing

[1]The first standardized test (on arithmetic reasoning) was developed by Stone (a student of Thorndike's) in 1908 (Ayres, 1918).

that such tests yield poor measures of mathematical performance for national assessment.

First, the purpose of norm-referenced standardized tests is to order respondents with respect to a particular type of mental ability or achievement, to indicate a respondent's position in a population. To do this a standardized test is created from a set of independent questions. The same items are then administered to every student, and the number of correct answers tallied. Each test is accompanied by an appropriate table for transforming the resulting scores into meaningful characterizations of pupil mental ability or achievement with grade-equivalent scores, percentiles, stanines, and so on. For example, millions of students each year take one of the major college admissions tests, the Scholastic Aptitude Test (SAT) or the American College Test (ACT). Both are standardized tests. Scores derived from these tests are used to make selection and placement decisions. Unlike standardized tests, national assessment should not order students on a single scale; we should assess group achievement on a set of variables over time.

Second, although each standardized test is designed to order individuals on a single trait, such as quantitative aptitude, the derived score is not a direct measure of that trait. It is as if one were measuring the Houston Rockets' basketball star Ralph Sampson's height and reporting not that he is 7' 4" but that he is at the 99th percentile for American men. For mathematics achievement there is no theoretical single trait (like height) that is being assessed. National assessment should provide profile data on several aspects of mathematics for groups, not single scores on individuals.

Third, because individual scores on standardized are compared with those of a norm population there will always be some high and some low scores. This is true even if the range of scores is small. Thus, high and low scores can not be judged as "good" or "bad" with respect to the underlying trait. For national assessment we should be primarily interested in levels of performance on what has been taught, not just the relation of individual performance to the performance of a norm population.

Fourth, the items on standardized tests are assumed to be both independent and equivalent to each other. They are selected on the basis of general level of difficulty ($\underline{p}$ value) and some index of discrimination (e.g., nonspurious biserial correlation). National assessment should be interested in interdependent items that reflect specific domains.

Fifth, there are two specific problems with the norm referencing of standardized tests. The representativeness of any norm group (a national sample of students tested at a given time) is questionable. Also, because of the expense involved, norms are updated infrequently. Thus, comparisons of scores with a norm group may be both unrepresentative and out of date. National assessment should be based on a timely representative sample.

Finally, a primary weakness of standardized tests is that they are ofen used for decisions they were not designed to address. For example, aggregating standardized scores for students in a class, school, or district to get a mean of achievement is very inefficient; it provides too little information for the cost involved. Unfortunately, the common use of test scores appears to be more strongly related to political rather than educational uses. For example, it is claimed that elected officials and educational administrators increasingly use the scores from such tests in comparative ways--to indicate which schools, school districts, and even individual teachers give the appearance of achieving better results (National Coalition of Advocates for Students, 1985). Such comparisons are misleading.

One can only conclude that standardized tests are unwisely overused and that their derived scores are of little value as indicators of achievement for national assessment of mathematical performance.

Profile Achievement Tests

These tests, in contrast to standardized tests, are designed to yield a variety of scores for groups of students. As early as 1931 Ralph Tyler outlined a procedure for test construction and validation that clearly pointed out the essential dependence of a program of achievement testing on the objectives of instruction and the recognition of forms of pupil behavior indicating attainment of the desired instructional outcomes. Since then, profile tests have become very popular alternatives to standardized tests. They have been developed for several major studies of mathematical performance such as the National Longitudinal Study of Mathematical Abilities (NLSMA) (Wilson, Cahen, & Begle, 1968-72), the First International Mathematics Study (FIMS) (Husén, 1967), the Second International Mathematics Study (SIMS) (Crosswhite, Dossey, Swafford, McKnight, & Cooney, 1985), and several different state assessments. However, in these studies either the sampled population is not nationally representative (e.g., NLSMA and state assessments) or the content assessed does not reflect the American mathematics curricula (e.g., FIMS and SIMS).

There are five features of profile assessments that make them quite different from standardized tests. First, there is no assumption of an underlying single trait. Instead instruction at any grade in mathematics is assumed to be on several topics. The tests are designed to reflect the multidimensional nature of mathematical outcomes. It must be noted that the temptation to aggregate and derive a single total score would yield a very misleading score.

Second, the approach to identifying what is to be assessed in profile testing is to specify a content by behavior matrix. For example, the matrix used for profiling eighth-grade performance in the Second NAEP is shown in Figure 1 (Carpenter, Corbitt, Kepner, Lindquist, & Reys, 1981). Content topics are crossed with hypothesized cognitive levels. The content topics are judged to be appropriate for a grade

Insert Figure 1 Here

level, and the cognitive levels are usually based on some adaptation of those in Bloom's Taxonomy (1956). Items, similar to those in standardized tests, are prepared for each cell in the matrix. Item data then can be reported in several ways. They can be reported in terms of item means; cell means can be calculated; or item scores can be aggregated, either by columns to yield cognitive level scores or by rows to yield topic scores.

CONTENT

PROCESS		A Numbers and Numeration	B Variables and Relationships	C Shape, Size and Position	D Measurement	E Other Topics
	I. Mathematical knowledge					
	II. Mathematical skill					
	III. Mathematical understanding					
	IV. Mathematical application					

Figure 1. Scheme for developing objectives and exercises for the 1977-78 NAEP (Carpenter et al., 1981).

Third, the unit of investigation is a group not an individual. Matrix sampling is often used so that a wider variety of items can be given.

Fourth, comparisons between groups are done graphically on actual scores. No transformations are needed.

Finally, validity is determined in terms of content and/or curricula validity. Mathematicians and teachers are asked to judge whether individual items reflect a content/behavior cell in the matrix and sometimes to judge whether or not the item represents something that was included and taught in the curriculum.

The strength of profile achievement tests is that they can provide useful information about groups. They are particularly useful for general evaluations of changed educational policy that directly affects classroom instruction. Hence, they are ideal for national assessment. However, there are several weaknesses of these tests. First, because they are designed to reflect group performance, they are not useful for individual ranking and diagnosis. An individual student takes only a sample of items. Second, they are somewhat more costly to develop than standardized tests and harder to administer and score, and their results are more difficult to organize for interpretation. In particular, because they yield a set of scores, comparisons between groups are via differential profiles that do not yield simple distinctions.

However, their primary weakness is in the outdated assumptions underlying the two dimensions of content by behavior matricies. The content dimension (for example see Figure 1) involves a classification of mathematical topics into "informational" categories which lack conceptual validity.

The behavior dimension of matricies has always posed problems. All agree Bloom's Taxonomy (1956) has proven to be useful for low level behavior (knowledge, comprehension, and application) but difficult for the higher levels (analysis, synthesis, and evaluation). Single answer multiple-choice items are not reasonable for those levels. In fact, the Taxonomy fails to reflect current psychological thinking. It is based on the naive psychological principle that simple individual behaviors become integrated to form a more complex behavior. In the past thirty years our knowledge about learning and information processing has changed and expanded. We should discard Bloom's Taxonomy and use a contemporary alternative for profiling.

National Assessment of Educational Progress (NAEP)

As stated earlier, in 1972-73 the first National Assessment of Educational Progress for mathematics was carried out. Its intent was to provide to educational policymakers and practitioners information that could be used to identify educational problem areas, to establish educational priorities, and to determine national growth in education. The eight specific goals established for NAEP are:

Goal I: To measure change in the educational attainments of young Americans.

Goal II: To make available on a continuing basis comprehensive data on the educational attainments of young Americans.

Goal III: To utilize the capabilities of National Assessment to conduct special interest "probes" into selected areas of educational attainment.

Goal IV: To provide data, analyses, and reports understandable to, interpretable by, and responsive to the needs of a variety of audiences.

Goal V: To encourage and facilitate interpretive studies of NAEP data, thereby generating implications useful to educational practitioners and decision-makers.

Goal VI: To facilitate the use of NAEP technology at state and local levels when appropriate.

Goal VII: To continue to develop, test, and refine the technologies necessary for gathering and analyzing NAEP achievement data.

Goal VIII: To conduct an ongoing program of research and operational studies necessary for the resolution of problems and refinement of the NAEP model. (Implicit in this goal is the conduct of research to support previously mentioned goals.)

(Carpenter, Coburn, Reys, & Wilson, 1978, pp. 4-5)

Since 1972-73 three more mathematics assessments have been conducted: 1977-78, 1982, and 1986. The contract for conducting the first three assessments in mathematics for the Department of Education was held by the Education Commission of the States. The last assessment is being conducted by Educational Testing Services.

Each of the assessments has involved administering profile achievement tests. The tests are comprised of a set of questions or tasks called exercises. Subsets of exercises were administered to a scientifically determined national sample of students at three age levels representing educational milestones attained by most students: age 9, when most students have been exposed to a basic primary education; age 13, when most students have finished their elementary school education; and age 17, when most students are near completion of their secondary education. It should also be noted that only in the first assessment were both 17-year-olds who were not in school and adults (ages 26-35) tested; these groups were not represented in later assessments.

The exercises in each of the tests have predominantly been given in a multiple-choice format, although in the first NAEP many open-ended exercises were given. These were not given in later assessments because of the time and cost involved in scoring them. Also, the exercises have been split into two categories: secure items, to be readministered in later assessments to show change; and published items, to be used to report results.

Over the four assessments there have been three important changes. First, ETS has taken over the administration of NAEP from ECS. The consequences of this shift are not as yet clear, although it is argued that ETS is more capable of developing and administering an efficient national assessment. Second, the testing and sampling was simplified after the first assessment, primarily to reduce costs: open-ended exercises are no longer used, and both 17-year-olds and adults are no longer tested. Third, the major change has been in the reconceptualization of the content-by-process matrix on which each assessment was based. In 1972-73 a three dimensional matrix was used. The content dimension had 17 areas:

1. Number and Numeration Concepts
2. Properties of Numbers and Operations
3. Arithmetic Computations
4. Sets
5. Estimation and Measurement
6. Exponents and Logarithms
7. Algebraic Expressions
8. Equations and Inequalities
9. Functions
10. Probability and Statistics
11. Geometry

12. Trigonometry

13. Mathematical Proof

14. Logic

15. Miscellaneous Topics

16. Business and Consumer Mathematics

17. Attitude and Interest

(Carpenter et al., 1978, pp. 9-10)

The process dimension had six categories:

1. To recall and/or recognize definitions, facts and symbols

2. To perform mathematical manipulations

3. To understand mathematical concepts and processes

4. To solve mathematical problems--social, technical, and academic

5. To use mathematics and mathematical reasoning to analyze problem situations, define problems, formulate hypotheses, make decisions, and verify results

6. To appreciate and use mathematics

(Carpenter et al., 1978, p. 9)

A third dimension--uses of mathematics--had three categories:

1. social mathematics (the mathematics needed for personal living and effective citizenship in our society),

2. technical mathematics (the mathematics necessary for various skilled jobs and professions), and

3. academic mathematics (the formally structured mathematics that provides the basis for an understanding of various mathematical processes).

(Carpenter et al., 1978, p. 9)

This ambitious matrix was considerably simplified for the second assessment. Simplification was carried out in part because it was impossible to adequately assess each of the 306 cells of the matrix and in part for economic reasons. The framework adopted for the second assessment was shown in Figure 1. It contained only two dimensions; the uses dimension was dropped. The content dimension had 5 categories rather than 17:

1. numbers and numeration
2. variables and relationships
3. size, shape, and position
4. measurement
5. other topics

And the process dimension was comprised of four categories rather than six:

1. mathematical knowledge
2. mathematical skill
3. mathematical understanding
4. mathematical application

(Carpenter et al., 1981, p. 4)

Obviously this simplification from 306 to 20 cells in the matrix made exercise writing and summarization of results much easier than in the first assessment. However, this simplification may have been too drastic, particularly because of the elimination of open-ended items that could better measure higher order thinking skills.

The third assessment given in 1982 used the same basic framework as the second. The process dimension was unchanged; while in the content dimension "size, shape, and position" was relabeled "geometry" and "other mathematics" was split into "probability and statistics" and "graphs and tables" (Education Commission of the States, 1983).

For the fourth assessment there was a radical rethinking of the matrix. First, both the content and the process domains were restructured. Seven content areas were specified:

1. fundamental methods of mathematics

2. discrete mathematics
3. data organization and interpretation
4. measurement
5. geometry
6. relations, functions, and algebraic expressions
7. numbers and operations

(Educational Testing Service, 1985, pp. 4-13)

The new labels and their order indicate a significant shift in emphasis from the previous assessments. In particular, the identification of such fundamental methods as modeling, induction, deduction, algorithms, logic, and proof--when combined with the new categories of "discrete mathematics" and "data organization and interpretation"--indicates more of an emphasis on "knowing how" than on "knowing what."

For the process dimension five categories were stated:

1. problem solving
2. routine application
3. understanding/comprehension
4. skill
5. knowledge

(Educational Testing Service, 1985, pp. 1-2)

The category of problem solving has been added and given prominence that reflects the intent of ETS to shift the emphasis of the assessment from knowledge toward higher order thinking skills. The results of this assessment promise to be different from past assessments. However, it will be a year or more before summaries will be available.

Results from the assessments have been reported with the full cooperation of the mathematics education community, and they have had considerable impact. For the first two assessments, in addition to reports prepared by ECS, the National Council of Teachers of Mathematics appointed a committee to study the results and prepare summaries (Carpenter et al., 1978, 1981). For the third assessment, ECS organized a committee of mathematics educators (most of whom had worked on the previous reports for NCTM) to prepare their basic report that focused on change over the three assessments (Education Commission of the States, 1983). ETS is currently working with a group of mathematics educators to prepare a report on the fourth assessment.

The impact of the NAEP findings is hard to document. It is clear that the reports and articles based on them have been widely read and cited. For example, the finding that students have learned to add, subtract, multiply, and divide simple whole numbers has been used to allay the fears of the "back to basics" advocates. At the same time, the finding that a large percentage of students could not use those skills to solve word problems has provided needed ammunition for the "problem solving" advocates. Other examples could be given with respect to rational numbers, geometry, probability, and so on.

There are two difficulties with these reports. First, as noted earlier, profile tests yield results which are not easy to interpret. The NCTM committee took three years to produce each of the first two reports after the data were collected. Even then the resulting pictures were complex with information related to performance on items within cells, rows, or columns of the matrix. There is no simple set of indices that policymakers can easily use to make judgements about the health of mathematics instruction in their schools, districts, states, or even the nation. Unfortunately, this lack of simple indices undoubtedly contributes to the continued reference to SAT scores and the use of standardized tests, even though they are invalid.

The Assessment of Performance Unit (APU)

The Assessment of Performance Unit in Britain has much the same commission as the National Assessment of Educational Progress in the United States: to prepare a national profile on the educational achievement of children. The work of the APU is geared toward causing educational change by having assessment procedures precipitate curricular change (Clegg, 1985). The direction of change is essentially that outlined as desirable by the Cockroft Commission (Committee on Inquiry into the Teaching of Mathematics, 1982). This commission advocated, among other things, links with other curricular areas, practical work, the importance of language, a diagnostic approach to testing, mathematics for the majority, a graduated assessment, and records of progress. In the process, they gave several batteries of tests to a large number of students.

The tests were developed based on a typical content-by-behavior matrix to which a third dimension had been added to address understanding, practical application, problem solving, and attitudes. The third dimension, involving their more innovative ideas, was assessed separately. The basic battery included a large set of open-ended items (not multiple choice) given via matrix sampling to a large sample of 11- and 16-year-old students. This administration was followed by the practical and problem solving tests and an attitude inventory given individually to small samples of students.

The assessment methods for the practical and problem solving parts (Foxman & Mitchell, 1983) are a combination of pencil-and-paper answers to complex and realistic situations and practical assessment with manipulatives. Both also involve a diagnostic assessment interview (Denvir & Brown, 1985). The situational questions are largely analogous to the super-item (Collis, Romberg, & Jurdak, 1986) approach, in that there is a problem situation with considerable information followed by a series of increasingly complex questions. The diagnostic interviewing was conducted according to a script, but with some flexibility for clarification, limited prompting, or amended answers. Responses were checked against a precoded list. However, unanticipated answers were recorded in detail. The result yields valuable insight into students' mathematical thinking (Burstall, 1986). The APU approach to national assessment is obviously different from that of NAEP, but one which should be examined as changes are being proposed.

Summary

On a national, regional, or state basis information about the mathematical performance of groups of students is best obtained via profile tests. While standardized, norm-referenced tests are often used for this purpose, they are inadequate and yield too little information. Profile tests, on the other hand, yield rich data sets. Experience from the National Longitudinal Study of Mathematical Abilities, the First and Second International Mathematics Studies, the Assessment of Performance Unit, and the four National Assessments of Educational Progress in mathematics have provided the mathematics education community with lots of valuable information.

In particular, the NAEP assessments have been well documented and the information has been used. This is true although the assessments have been hampered by inadequate resources and inhibiting testing traditions such as the reliance on paper-and-pencil multiple-choice items and the use of out-of-date content-by-behavior matricies.

NEED FOR CHANGED NATIONAL ASSESSMENT

The information reported from the first three national assessments has, as noted above, proven to be of considerable value to mathematics educators, and the information from the recent fourth assessment promises to be of even more value given its shift in emphasis. Nevertheless, a change or modification in what mathematics is being assessed and how the assessments are carried out and reported is warranted. To build the argument for change, four aspects of national assessment are examined:

1. The need to challenge testing traditions,
2. the need to understand that we are in a new economic era,
3. the recognition that mathematics is a growing, dynamic discipline in which there have been significant changes over the last decade in what is deemed fundamental, and
4. the policy need for valid indicators of mathematical performance.

Challenging Testing Traditions

Sometimes educational reform is directed toward making schooling more efficient. Under those conditions expected outcomes have not changed, and assessment procedures may remain the same if they reflect those expectations. However, when expectations have changed, new assessment procedures should be developed. It is necessary to compare and contrast the "old" and "new" expectations, use the assessment tools designed for both, discard tools no longer appropriate, and develop new procedures when needed. Today schools should be planning to change the emphasis from drill on basic mathematical concepts and skills to explorations that teach students to think critically, to reason, to solve problems, to interpret, to refine their ideas, and to apply ideas in creative ways.

The current approach to gathering information about pupils' mathematical performance by administering a set of individual multiple-choice paper-and-pencil questions to students and then tallying the number of correct answers is out of date. The procedure is an outgrowth of the "scientific testing movement" which began at the turn of the century.

The testing movement was a product of its times. It grew out of the machine-age thinking of the industrial revolution of the past century. The intellectual contents of the machine age rested on three fundamental ideas. The first was reductionism. The machine age was preoccupied with taking things apart. The idea was that in order to deal with anything you had to take it apart until you reached ultimate parts. The second fundamental idea was that the most powerful mode in thinking was a process called analysis. Analysis is based in reductionism. It argues that, if you have something that you want to explain or a problem that you want to solve, you start by taking it apart. You break it into its components; you get down to simple components; then you build up again. The third basic idea of the machine age has been called "mechanism." Mechanism is based on the theory that all phenomena in the world can be explained by stating cause and effect relationships. The primary effort of science was to break the world up into parts that could be studied to determine cause and effect relationships. The world was conceived of as a machine operating in accordance with unchanging laws.

These ideas gave rise to what we now call the first Industrial Revolution. In this world, work was conceived of in physical terms, and mechanization was about the use of machines to perform physical work. Man was supplemented by machines as a source of energy. Man-machine systems were developed for doing physical work to facilitate mechanization.

This whole process is clearly reflected in what has happened in school mathematics during the last half century. Mathematics was segmented into subjects and topics, eventually down to its smallest parts--behavioral objectives. At this point, a network diagram, a hierarchy, was created to show how these components were related to produce eventually a finished product. Next, the steps by which one travelled that hierarchy were mechanized via textbooks, worksheets, and tests. In particular, tests that could be efficiently administered and reliably scored were a central feature of this conceptualization. Furthermore, teaching was dehumanized to the point that the teacher had little to do but manage the production line. Businesses, industry, and, in particular, schools have been conceived and modified based on this mechanical view of the world since before the turn of the century and continue to operate in a mechanical tradition.

Objective test items administered under standardized conditions undoubtedly will continue to be used, but they are products of an earlier era in educational thought. Like the Model T Ford assembly line, objective tests were considered an example of the application of modern scientific techniques in the 1920s. Today we ought to be able to develop better indices of achievement.

A New Economic Age

We are now in a new economic age--the Information Age--which will significantly alter the character of American schooling. Labeling the new age as the Information Age gives it a rather lofty, intellectual, cerebral sound, especially in comparison to the muscular, grinding, "dark, satanic mill" connotations of the Industrial Age. Early designations, such as the Post-Industrial Age (Bell, 1973) or the Super-Industrial Age (Toffler, 1985), simply recognized that our industrial economy has changed so drastically that a new description was needed. Caused by a revolution in communications which started with the telegraph, it could equally have been described as "the Communications Age." However, the integration of telephone, television, and computer permits instant transfer of information between people anywhere. This, with the geometric growth of knowledge, has combined to make Information Age a more apt label.

Information is the new capital and the new raw material. The ability to communciate is the new means of production, the communications network providing the relations of production. Industrial raw materials only have value if they can be put together to form a desirable product; the same is true of information.

The works of several authors (Naisbitt, 1982; Shane & Tabler, 1981; Toffler, 1985; Yevennes, 1985) point toward some of the attributes of the shift from an industrial society to an information society. First, it is an economic reality, not merely an intellectual abstraction. Second, the pace of change will be accelerated by continued innovation in communications and computer technology. Third, new technologies will be applied to old industrial tasks first but will then generate new processes and products. Fourth, basic communication skills are more important than ever before, necessitating a literacy-intensive society.

Information only has value if it can be controlled and organized for a purpose. To tap the power of computers, it is obligatory, first, to be able to communicate efficiently and effectively; that means being both literate and numerate. In addition, in an environment of accelerating change, the old approach of training for a lifetime occupation will have to be replaced by developing learning power, which also depends on the abilities to understand and to communicate. Finally, concurrent with the move from an industrial society to one based on information is awareness of the change from a national economy to a global economy. The change is important for the simple reason that the United States and the advanced societies of the West are losing their industrial supremacy. Mass production is more cheaply accomplished in the less-developed parts of the world.

In particular, Zarinnia and Romberg (1987) have recently argued that

> the most important single attribute of the Information Age economy is that it represents a profound switch from physical energy to brain power as the driving force, and from concrete products to abstractions as the primary products. Instead of training all but a few children to function smoothly in the mechanical systems of factories, adults who can think are needed. . . . This is significantly different from the concept of an intellectual elite having the responsibility for innovation while workers take care of production. (p. 12)

Thus, thinking skills must be the focus of instruction in mathematics in the near future, and assessment procedures need to be developed to portray not only the number of correct answers students can produce but the thinking that produced those answers.

Unfortunately, as Lauren Resnick (in press) has pointed out,

> American schools, like public schools in other industrialized countries, are the inheritors of two quite distinct educational traditions--one aimed at the education of an elite, the other concerned with mass education. These traditions conceived of schooling in different terms, had different clienteles, and held different goals for their students. Only in the last sixty years or so have the two traditions merged, so much so that in American schools it is now difficult to detect the separate threads. Yet a case can be made that it is a continuing and as yet unresolved tension between the goals and methods of elite and mass education that is producing our current concern for the teaching of [thinking] skills. (pp. 4-5)

Furthermore, she argued that

> clearly one of the most important challenges facing the movement for increasing higher order skills learning in the schools is development of appropriate evaluation strategies. Part of the problem is our penchant for testing. American pressures for standardized testing, especially at the elementary and secondary school levels, makes it difficult for curriculum reforms that do not produce test score gains to survive.

> But most current tests favor students who have acquired lots of factual knowledge and do little to assess either the coherence and utility of that knowledge or students' ability to use it to reason, solve problems and the like. (pp. 40-41)

Future national assessments must be in tune with this emerging world view.

Mathematics: A Dynamic Discipline

This is not the place for a detailed discussion about the changes that have occurred and are occuring in mathematics and the mathematical sciences. However, three issues must be mentioned. First, the mathematical expectations or goals for our students have changed in light of the current social revolution. Procedural skills such as computational algorithms are no longer as important; the calculator and computer have not only freed man from the necessity of performing such tedious calculations, they have made other extremely complex models and calculations possible. Quantitative reasoning, mathematical modeling, statistics, and problem-solving are now more important than ever before. It is premature to detail the new expectations at this time, since several groups--including the Mathematical Sciences Education Board, the National Council of Teachers of Mathematics, the American Association for the Advancement of Science, and the Council of Chief State School Officers--are preparing frameworks, criteria, and standards for the new fundamentals of mathematics. Nevertheless, new expectations imply that new methods of assessment will be needed.

Second, the new expectations reflect a shift in emphasis about mathematics. As Romberg (1983) put it,

> When nonmathematicians, such as sociologists, psychologists, and even curriculum developers

> look at mathematics, what they often see is a static and bounded discipline. This is perhaps a reflection of the mathematics they studied in school or college rather than a sure insight into the discipline itself. John Dewey's distinction between "knowledge" and "the record of knowledge" may clarify this point. For many, "to know" means to identify the artifacts of a discipline (its record). For me and many others, "to know" mathematics is "to do" mathematics. (p. 121-122)

Third, the new emphasis on process implies that the content of mathematics is its own epistemology (Romberg & Zarinnia, 1987); several things follow. First, context, content, and process are inextricably related. Second, interdisciplinary activity is a natural corollary, once mathematics is seen as a process in search of content and context. It makes more sense for children, trying to understand entirely abstract processes, to root their understandings in concrete contexts from the real world, whether cake-baking or stream flow. Third, a clear understanding of the significance of an epistemological emphasis is essential to the creation of a framework for assessing the mathematical progress of children.

Epistemology is concerned with the origin, nature, methods, and limits of knowledge. Therefore, emphasis on the creation of knowledge virtually requires an epistemological perspective. Knowing involves making cognitive structures match the reality that they are supposed to represent. However, because experience is the way to knowing, knowledge is necessarily subjective and constructive and cannot be separate from the knower. In this context, public knowledge structures ensue from communal agreement about private cognitive structures.

Politics and Policy

It is one thing to develop a national assessment procedure that is useful for mathematics educators and other experts in the field and quite another to capture the performance of students in a manner that is interpretable by a state legislator or school official with little mathematics background. The past national assessments have provided invaluable information to mathematics educators but obviously have not been as useful to policymakers. In fact, the primary rationale for forming the Study Group on National Assessment is related to this need. It is reasonable for policymakers to expect that information from national assessments be collected, analyzed, and reported to meet their needs. One would hope that important educational decisions would be made using the most valid information available.

Secretary Bennett's seven principles given to guide the Department of Education in developing plans for the future of national assessment reflect this concern. Facilitating comparisons between groups (states) at the same time and over time as well as making the information easily accessible are examples of this concern.

Summary

Past efforts of NAEP have been very useful but we can not be complacent. The assessments need to be continually improved and modified and new procedures developed. Current methods of gathering and reporting information need to be changed, in part because of our emergence into the Information Age, in part because of the dynamic nature of and changes in the mathematical sciences, and in part because of the obvious needs of educational policymakers.

NEW CONCEPTS AND SUGGESTIONS

To complete this paper on national assessment and suggest changes in the future, this section includes a discussion of three aspects of past practices that need to be changed: the model for mathematics content, the nature of the items, and sampling and reporting procedures.

The Model for Mathematics Content

Traditional monitoring practices have consistently used a content-by-behavior matrix as their theoretical framework. However, the mathematical, psychological, sociological, and pedagogical theories embedded in such matricies are, quite simply, inadequate. Unfortunately, their cohesive power exerts a powerful influence that subliminally impedes change.

The classifications of content on which assessment has been based are largely a means towards the linear ordering of work. Often strands and subjects within strands are specified, but no conceptual or psychological dependence has been apparent or assumed. If a strict partial ordering of the segments can be found, a content hierarchy could be constructed. However, if the structure of instruction and assessment is to have a positive influence, mathematical content needs to be arranged, where appropriate, in true hierarchies based on the interdependence of skills and concepts.

The behavioral dimension also has two major problems: fragmentation of objectives and the hierarchy. The categories of behavior rested on the premise that educational objectives stated in behavioral forms have their counterparts in the behavior of individuals, which can be observed, described and, therefore, classified. Some fear was expressed that Bloom's Taxonomy (1956)

> might lead to fragmentation and atomisation of educational purposes such that the parts and pieces finally placed into the classification might be very different from the more complete objective with which one started. (pp. 5-6)

However, it was felt that the structure of the hierarchy would enable users to clearly understand the place of objectives in relation to each other. Unfortunately, this has not proven to be the case.

The hierarchy suggests that "lower" skills should be taught before the "higher" skills. As Resnick (in press) argues,

> This assumption--that there is a sequence from lower level activities that do not require much independent thinking or judgment to higher level ones that do--colors much educational theory and practice. Implicitly at least, it justifies long years of drill on the "basics" before thinking and problem solving are attended to or demanded. A fundamental challenge to this assumption is provided by cognitive research on the nature of basic skills such as reading and mathematics. (p. 10)

A modern alternative to content-by-behavior matricies is in order.

It is important to replace the matrix model with one more capable of handling the complexity and interdependence of content and psychological processing. The new model must be powerful and have both tight internal coherence and congruence with the trends in mathematics, science, and society. The direction should be in terms of network models that are both widely used and consistent in philosophy with approaches to the creation of knowledge. Such models are also capable of modeling complex processes and, in consequence, likely to exert powerful pressure in stimulating change toward the new world view in mathematical education. One such network model comes from the work of the French mathematical psychologist Gerard Vergnaud.

Vergnaud (1983) has a very distinct view of the interrelationship between meaning and complexity, the meaning of mathematics coming from practical and theoretical problems to be solved. He labels his ideas conceptual fields. Crucial to his perception is that mathematics arises from contexts. He emphasized the theory of didactic situations--conceptualizations depend on the context in which they are formulated and are eventually modified in the face of new situations. In other words, knowledge emerges in situ and there is a tight relationship between the context, the conceptual properties of the context, and the best symbolic representation of both concept and context. Conceptual development is so slow that it is desirable to study the same conceptual field year after year, going deeper, meeting new contexts through different problems to be solved (Vergnaud, 1982). Examples have been given for additive structures, multiplicative structures, directed numbers, and measurement. Such fields are derived in the following manner.

1. The symbolic statements (e.g., $a + b = c$ and $a - b = c$; where a, b, and c are natural numbers) which characterize the domain are identified.

2. The implied task (or tasks) to be carried out is specified. For addition and subtraction this involves describing the situations where two of the three numbers a, b, and c in the statements above are known and other is unknown.

3. One identifies the rules (invariants) that can be followed to represent, transform, and carry out procedures to complete the task (e.g., find the unknown number using one or more of such procedures as counting strategies, basic facts, symbolic transformations such as a +

$[\,] = \underline{c} \iff \underline{c} - \underline{a} = [\,]$, computational algorithms for larger numbers).

It should be noted that in these first three steps one only considers the formal aspects of a mathematical system.

4. One identifies a set of situations that have been used to make the concepts, the relationships between concepts, and the rules meaningful (e.g., join-separate, part-part-whole, compare, equalize, fair trading).

The result of following the above steps yields a map (a tightly connected network) of the domain of knowledge.

The problem of complexity is not simply one of memory overload but of the difficulties inherent in conceptualizing tightly interrelated structures of concept, procedure, and representation. This constitutes a serious problem for the transfer of concepts from one context to another. It is a matter of cognitive dissonance.

> This source of resistance to change lies in the fact that an element is in relationship with a number of other elements. To the extent that the element is consonant with a large number of other elements and to the extent that changing it would replace these consonances by dissonances, the element will be resistant to change. (Festinger, 1957, p. 27)

Good teaching therefore requires that a set of relations be learned in one context and then another so that the relational invariants and common structure can emerge. Gradual increase in complexity relies on controlled changes of structure in a fixed context and deliberate transfers of structure from one context to another (Bell, 1985). In other words, control over increases in complexity depends on a moderated introduction of cognitive dissonance.

The Nature of the Items

A practical problem of testing is that any test attempting to be comprehensive in approach requires a long time for children to complete and a long time to grade. Multiple-choice exercises provide one simple approach that NAEP has used. This approach offers several advantages.

1. It made possible much more extensive and representative sampling of the content topics because it tested more topics less deeply.
2. Scoring multiple-choice items is much faster and less costly than scoring open-ended items.
3. Because the items were classified according to location in the matrix, a more detailed profile of groups of students became possible.
4. Questions could be designed to stand alone.

Because the intent now is to assess the creation of knowledge and the processes involved rather than just measure the extent to which children have acquired a coverage of the field of mathematics, a much wider variety of new measures, many considered qualitative, are needed.

The single most severe criticism of objective test questions designed to assess a specific item of content at a specific level of content is that they trivialize learning and knowledge (Berlak, 1985). This is almost inherent to such questions for several reasons. First, they are designed to test a single, specific objective in the matrix. Thus, elements in the multiple-choice format are designed so that the candidate can pick an answer which is sufficiently specific to unequivocally demonstrate the sought behavior. This, by definition, tends to eliminate synthesis between content or behavior. Second, the very nature of objective tests that require choosing among alternatives eliminates creativity in answering. Even the intent militates against creativity in answering because the intent is micro-analytic rather than synthetic or creative.

Less trivializing of mathematical thinking was observed in the efforts of the Assessment of Performance Unit (Cambridge Institute of Education, 1985). Their students benefitted from the opportunity to think, achieving different success with the free-form response, practical problems.

In addition to their direct effects, tests exert powerful indirect effects on both the style of teaching and the style of learning. When one studies for an essay exam, one progressively surveys and synthesizes, putting the parts together and developing a mental model of the structure of the subject. One also develops points of view and arguments to advance and support, for those are the expectations. By contrast, in an objective, multiple-choice test, one learns to cover the parts and make fine distinctions between alternative ways of stating the same thing to distinguish a "right" answer from a "wrong" one, the implication being that there is always a single right answer. In other words, the one reinforces the view of mathematics as ground to be covered; the other requires that students create their own models of mathematics.

Another aspect of most objective tests is that, even though some questions may be designed to test lower level thinking and others designed to evaluate higher thought processes, levels of thinking are usually tested independently of each other, allowing little notion of a student's approach to a given problem. Frederiksen (1984) observed that a multiple-choice format does not measure the same cognitive skills as a free-response form and that

> efficient tests tend to drive out less efficient tests, leaving many important abilities untested--and untaught. (p. 201)

One example of a desirable outcome untested and untaught is the ability to cope with ill-structured problems, which are not found on standardized achievement tests.

There should be a strong congruence between the purpose for assessment, the model of assessment, and the tools for assessment. In past assessments there was a cohesion between the hierarchical purpose of ranking, the content-by-behavior matrix, and standardized, objective group testing. For an equally cohesive approach to be developed, alternative methods of assessment must be designed that are congruent with teaching students to create knowledge. While any number of indirect proxies may be postulated, the only direct indicator is the kind of knowledge created by students in the system. Thus, tools are needed to assess students' progress in creating knowledge.

There is an additional consideration. The standardized objective testing approach lends itself readily to quantification because items are scored right or wrong, 1 or 0. But quality, structure, predictive power, collaborative effort, and so on can not be dichotomously scored; the exclusively quantitative nature of group testing is no longer tenable. The first step in developing new scoring procedures will almost inevitably be qualitative, even though means will likely be devised for subsequent quantification.

Work in artificial intelligence suggests that there are two basic facets to creating knowledge:

1. A database of facts and assertions

2. An inference engine

There are, therefore, several ways of adding to knowledge, whether individually or cooperatively:

1. Increasing the power of the inference engine.

2. Adding to the facts in the database.

3. Adding to the network of assertions in the database.

Significantly, power in knowledge creation is primarily a consequence of the knowledge base and only secondarily a consequence of the power of the inference method (Feigenbaum, 1984). Furthermore, the most important aspect of the knowledge base is the structure of assertions (Robinson, 1984). This reinforces the notion of knowledge creation as a matter of searching for new structures. It is essentially similar to the conclusions reached by Pask (1984) on the importance of analogic reasoning in the creation of new knowledge and to the use of analogy in the mathematical modeling of complex systems (Cross & Moscardini, 1985).

In summary, for policy purposes it is important to have tools that monitor children's strategies, problems, and achievements. Simply stated, there is a need for tools that document the production of knowledge and not merely the proxies that contribute to the process. Because knowledge is derived from experience, it seems logical both to monitor the quality of experience in which students learn how to create knowledge and to assess in a practical and realistic context.

Several approaches offer some promise. One is the use of practical assessments. The notion of practical assessment has been typically restricted to such areas as medical school and flight training. However, the APU gave practical tests in measurement of mass and area and in extended problem solving situations as part of its assessment program in mathematics; even more practical testing was given in science. A second tool for group assessment of intellectual structure in context, which is cost-effective, is the use of superitems (Collis, Romberg & Jurdak, 1986).

Sampling and Reporting Procedures

The basic strategy for gathering national data has been to use multiple matrix sampling for a variety of mathematical exercises given to a national sample of students at important age (grade) levels. Results were aggregated, and profiles for the population were estimated. The basic strategy has proven to be reasonable and has yielded valuable information for the mathematics education community. However, two aspects of the procedure could be developed: state profiles and indicators.

State profiles. Given the pressure to make comparisons between states, data based on state samples needs to be gathered. This can be accomplished in either of two ways. States could elect to administer sets of NAEP exercises as a part of their state assessments. This is currently being done in several states (e.g., Massachusetts and Wisconsin). The alternative would be to change the sampling frame for NAEP so that state profiles could be generated.

Indicators. As stated earlier, one of the serious problems with profile achievement tests is that reporting results and comparing profiles for different groups is difficult because of the complex nature of mathematical outcomes. Yet policymakers need simple but valid indicators to make sensible decisions. The answer is not to simplify

NAEP so that it just yields a small set of scores (like standardized tests). Instead NAEP should be encouraged to gather the most extensive and valid set of information possible. Then from that data set, indicators of the health of school mathematics could be constructed.

Economic and social indicators have been developed and used in various ways by governmental and other institutions concerned with formulating and evaluating public policy for decades. They are constructed by sampling information from a rich data base, guided by an explicit theoretical model. For example, the Dow Jones average is an indicator derived from sales information about a sample of stocks. Similarly, the Cost of Living Index is derived by sampling cost data for a variety of products.

SUMMARY AND CONCLUSIONS

National assessment of mathematical performance is important for teachers, mathematics educators, administrators, and policymakers. The basic strategy for gathering profile information for students at several age (grade) levels used in the past assessments is reasonable. The first three assessments have yielded very useful information which has affected school mathematics. The most recent assessment promises to yield even more illuminative results.

However, the procedures now followed could be improved. Content-by-behavior matricies should be discarded. In their place a network model (such as conceptual fields) needs to be adopted. The types of exercises included in the batteries should be expanded to reflect the network model. These should include new contexts so that the construction of knowledge can be assessed. The sampling base can be changed so that data can be gathered for state comparisons. Finally, given the rich data base such an improved assessment would yield, it should be possible to construct reasonable indicators for use by policymakers.

The first recommendation is that work be initiated to identify major conceptual fields in mathematics, such as additive and multiplicative structures. These fields interrelate rather than separate content and behavioral ideas. Assessment information then could be developed to portray the degree to which a student has a coherent system of concepts, relationships, and symbols to use when faced with differing contextual situations within a particular conceptual field.

The second recommendation is that for future national assessments the Department of Education should encourage the development of a variety of alternate items and testing formats.

The third recommendation for the national assessment of mathematics is to increase the data base so that it reflects current expectations about how students construct mathematical knowledge to build a theoretical model of mathematical performance.

The fourth recommendation would be to construct reasonable indicators from that model for policy purposes.

REFERENCES

Ayres, L. P. (1918). History and present status of educational measurements. In _The measurement of educational products: Seventeenth Yearbook of the National Society for the Study of Education_ (Part II). Bloomington, IL: Public School Publishing Co.

Bell, A. W. (1985). Some implications of research on the teaching of mathematics. In A. Bell, B. Low, & J. Kilpatrick (Eds.), _Theory, research & practice in mathematical education_, (pp. 61-79). Nottingham: Shell Center for Mathematical Education, University of Nottingham.

Bell, D. (1973). _The coming of post-industrial society: A venture in social forcasting_. New York: Basic Books.

Berlak, H. (1985). Testing in a democracy. _Educational Leadership_, _43_(2), 16-17.

Bloom, B. S. (Ed.). (1956). _Taxonomy of educational objectives: The classification of educational goals. Handbook 1. Cognitive Domain_. New York: Longmans, Green and Co.

Burstall, C. (1986). Alternative forms of assessment: A United Kingdom Perspective. _Educational Measurement: Issues and Practice_, _11_, 17-22.

Cambridge Institute of Education. (1985). _New perspectives on the mathematics curriculum: An independent appraisal of the outcomes of APU mathematics testing 1978-1982_. London: Her Majesty's Stationery Office.

Carpenter, T. P., Coburn, T. G., Reys, R. E., & Wilson, J. W. (1978). _Results from the first mathematics assessment of the national_ assessment of educational progress. Reston, VA: National Council of Teachers of Mathematics.

Carpenter, T. P., Corbitt, M. K., Kepner, H. S., Lindquist, M. M., & Reys, R. E. (1981). Results from the second mathematics assessment of the national assessment of educational progress. Reston, VA: The National Council of Teachers of Mathematics.

Clegg, A. (1985). Who interprets and transforms research. In T. A. Romberg (Ed.), Using research in the professional life of mathematics teachers: Papers from the Using Research Group, ICME5 (pp. 71-79). Madison, WI: Wisconsin Center for Education Research.

Collis, K. F., Romberg, T. A., & Jurdak, M. R. (1986). A technique for assessing mathematical problem-solving ability. Journal for Research in Mathematics Education, 17, 206-221.

Committee on Inquiry into the Teaching of Mathematics in Schools. (1982). Mathematics Counts (The Cockroft Report). London: Her Majesty's Stationery Office.

Cross, M., & Moscardini, A. O. (1985). Learning the art of mathematical modelling. Chichester: Ellis Horwood.

Crosswhite, F. J., Dossey, J. A., Swafford, J. O., McKnight, C. C., & Cooney, T. J. (1985). Second international mathematics study: Summary report for the United States. Washington, DC: National Center for Education Statistics.

Denvir, B., & Brown, M. (1985). Understanding of number concepts in low attaining 7-9 year olds: Part I. Development of descriptive framework and diagnostic instrument. Educational Studies in Mathematics, 17(1), 1-22.

Education Commission of the States. (1983). The third national mathematics assessment: Results, trends and issues. Denver: Author.

Educational Testing Service. (1985). NAEP mathematics objectives. Princeton, NJ: Author.

Feigenbaum, E. A. (1984). Knowledge engineering: The applied side. In J. E. Hayes & D. Michie (Eds.), Intelligent systems: The unprecedented opportunity (pp. 37-55). Chichester: Ellis Horwood.

Festinger, L. (1957). A theory of cognitive dissonance. Stanford, CA: Stanford University Press.

Foxman, D., & Mitchell, P. (1983). Assessing mathematics: 1. APU framework and modes of assessment. Mathematics in Schools, 1-5.

Frederiksen, N. (1984). The real test bias. American Psychologist, 39, 193-202.

Husén, T. (Ed.). (1967). International study of achievement in mathematics: A comparison of twelve countries. International project for the evaluation of educational achievement in education (IEA). Phase I (Vol. 1 and 2). New York: Wiley.

National Coalition of Advocates for Students. (1985). Barriers to excellence: Our children at risk. Washington, DC: Author.

Naisbitt, J. (1982). Megatrends: Ten new directions transforming our lives. New York: Warner Books.

Pask, G. (1984). Review of conversation theory and a protologic (or protolanguage), Lp. ECTJ, 32(1), 3-40.

Resnick, L. B. (in press). Education and learning to think. Washington, DC: National Research Council.

Robinson, J. A. (1984). Logical reasoning in machines. In J. E. Hayes & D. Michie (Eds.), Intelligent systems: The unprecedented opportunity (pp. 19-36). Chichester: Ellis Horwood.

Romberg, T. A. (1983). A common curriculum for mathematics. In G. D. Fenstermacher & J. J. Goodlad (Eds.), Individual differences and the common curriculum: Eighty-second Yearbook of the National Society for the Study of Education (pp. 121-157). Chicago: University of Chicago Press.

Romberg, T. A., & Zarinnia, E. A. (1987). Consequences of the new world view to assessment of students' knowledge of mathematics. In T. A. Romberg & D. M. Stewart (Eds.), The monitoring of school mathematics: Background papers. Vol. 2: Implications from psychology and outcomes of instruction. Madison: Wisconsin Center for Education Research.

Shane, H. I., & Tabler, M. B. (1981). Educating for a new millenium: Views of 132 international scholars. Bloomington, IN: Phi Delta Kappa Educational Foundation.

Toffler, A. (1985). The adaptive corporation. New York: McGraw-Hill.

Tyler, R. W. (1931). A generalized technique for constructing achievement tests. Educational Research Bulletin 8, 199-208.

Vergnaud, G. (1982). Cognitive and developmental psychology and research in mathematics education: Some theoretical and methodological issues. For the Learning of Mathematics, 3(2), 31-41.

Vergnaud, G. (1983). Why is an epistemological perspective a necessity for research in mathematics education? In J. C. Bergeron & N. Herscovics (Eds.), Proceedings of the fifth annual meeting of PME-NA, The North American Chapter of the International Group for the Psychology of Mathematics Education (pp. 2-20). Montreal.

Wilson, J. W., Cahen, L. S., & Begle, E. G. (1968-72). NLSMA reports: 1. Stanford, CA: School Mathematics Study Group.

Yevennes, M. (1985). The world political economy and the future of the U.S. labor market. World Futures, 21, 147-157.

Zarinnia, E. A., & Romberg, T. A. (1987). A new world view and its impact on school mathematics. In T. A. Romberg & D. M. Stewart (Eds.), The monitoring of school mathematics: Background papers. Vol. 1: The monitoring project and mathematics curriculum. Madison: Wisconsin Center for Education Research.