DOCUMENT RESUME

ED 269 253 SE 046 571

AUTHOR Raizen, Senta A.
TITLE Report on the Work of the Committee on Indicators of Precollege Education in Science and Mathematics.
PUB DATE [86]
NOTE 25p.; Paper presented at the Annual Meeting of the American Educational Research Association (70th, San Francisco, CA, April 16-20, 1986).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Educational Improvement; *Educational Quality; Elementary Secondary Education; Evaluation Criteria; *Mathematics Education; *Performance Factors; *Science Education: *Scientific Literacy
IDENTIFIERS *Indicators

ABSTRACT

This commitee of the National Research Council was established to develop improved indicators of mathematics and science education. In its first report, the committee used a model of education consisting of inputs, processes, and outcomes. Indicators recommended in these areas include teacher quantity and quality, quality of the curriculum content, instructional time and course enrollment in science and mathematics, and student achievement. Data and information currently available on these indicators are summarized and their adequacy assessed. Conclusions on how to improve the database are provided. The committee's current work is also summarized. Areas currently under consideration for development of indicators include resource investments in mathematics and science education, assessing teaching effectiveness, further consideraton of curriculum quality indicators, improving the assessment of learning, assessing attitudes and motivation, and indicators of science literacy in the general population. (Author)

ABSTRACT

REPORT ON THE WORK OF THE COMMITTEE ON INDICATORS OF PRECOLLEGE EDUCATION IN SCIENCE AND MATHEMATICS

Presentation for the AERA 1986 Annual Meeting

Senta A. Raizen

This committee of the National Research Council was established to develop improved indicators of mathematics and science education. In its first report, the committee used a model of education consisting of inputs, processes, and outcomes. Indicators recommended in these areas include teacher quantity and quality, quality of the curriculum content, instructional time and course enrollment in science and mathematics, and student achievement. Data and information currently available on these indicators are summarized and their adequacy assessed. Conclusions on how to improve the data base are provided. The committee's current work is also summarized. Areas currently under consideration for development of indicators include resource investments in mathematics and science education, assessing teaching effectiveness, further consideration of curriculum quality indicators, improving the assessment of learning, assessing attitudes and motivation, and indicators of science literacy in the general population.

REPORT ON THE WORK OF THE
COMMITTEE ON INDICATORS OF PRECOLLEGE
EDUCATION IN SCIENCE AND MATHEMATICS

Presentation for the AERA 1986 Annual Meeting
Senta A. Raizen

This presentation summarizes the work of a committee of the National Research Council which has been established to develop improved indicators of the condition of science and mathematics education in the nation's schools. The impetus for the work came from the convocation held by the National Academy of Sciences in spring 1982 on mathematics and science education (National Academy of Sciences, 1982) and the various reports on the condition of education that appeared in the fall of 1982 and spring 1983 (see, for example, National Science Foundation, 1983; National Commission on Excellence in Education, 1983; Twentieth Century Fund, 1983; Boyer, 1983). These reports all found serious inadequacies in precollege education; a number of them suggested that most U.S. students leave high school without adequate preparation in science and mathematics, whether for the job market or for continuing their education. The reports identified such specific school deficiencies as teacher shortages, inadequate curricula, and low standards of student performance. Perusal of these reports led to widespread concerns about the state of schooling. However, there were also questions raised about the quality of the information used to formulate the conclusions and policy recommendations in the reports (see for example, Peterson, 1983; Stedman and Smith, 1983). This concern led to the creation of the NRC committee. The committee is charged with laying the foundation for the development of an adequate monitoring system for use at the national, state, and local levels, so that the condition of mathemathics and science education could be tracked, particularly the effects of current improvement efforts.

The committee was confronted with defining what an indicator of science and mathematics education might be, selecting--at least on a preliminary basis--particular schooling variables as a basis for constructing indicators, and reviewing the available data pertinent to the selected variables. A major problem, in fact, is the large amount of statistical data and research information available on education. These data derive from diverse sources, address similar questions differently, and are collected and analyzed with various degrees of rigor. Hence, they cannot readily be cumulated nor adapted for use as indicators. Moreover, despite the wealth of information, some pertinent issues are not addressed. The potential for confusion and misuse of data has been vividly illustrated in recent attempts to report on the nation's schools. Therefore, the committee early in its work decided to select a quite limited set of indicators and concentrate on assessing the quality of pertinent information. Selecting a limited number of indicators, at least initially, is also important if one wishes to develop a monitoring system that is feasible and affordable at the federal, state, and local levels. (It is instructive to remember that economists use nearly a hundred indicators to monitor the state of the nation's economy, even though only a few of these--gross national product, unemployment rate, cost of living, interest rates--are generally reported by the mass media.)

Selecting Indicators

The committee chose a simple model of the education system for identifying areas of science and mathematics education to be monitored, consisting of educational inputs, schooling processes, and student outcomes. Since the primary goal of instruction in science and mathematics is student learning, the most explicit student outcome is student achievement in these fields. Other outcomes such as choice of college majors or careers and later career paths are also important to society, but it is more difficult to tie them directly to schooling variables. Another student outcome, student motivation and attitudes toward science and mathematics, was considered but not treated in the committee's preliminary work. Choosing student achievement as the outcome variable of greatest interest determines to a considerable extent what schooling input and process variables need to be selected--those that have some causal relationship to student achievement. Based on research evidence as well as on educational practice and experience, time spent on subject matter was selected as a proxy for schooling processes that should be monitored. The most obvious schooling inputs linked to student achievement in science and mathematics are the numbers (and quality) of teachers responsible for these areas of instruction and the content of the curriculum.

For each of these four areas--teacher quantity and quality, curriculum content, instructional time (and course enrollment in secondary school), and student achievement--the committee reviewed the data and information currently available, provided some findings on temporal trends and comparisons with other countries, judged the adequacy of the information available on the selected indicators, and made recommendations for improvement. (For details on the data sources and research reviewed, see Raizen and Jones, 1985.) This summary of the committee's work will concentrate on the adequacy of the information available rather than on the state of science and mathematics education, although I would be glad to comment on the latter in the discussion period.

Before summarizing the committee's conclusions and recommendations on the adequacy of information, however, some general comments are in order:

- Most of the pertinent data have been collected through surveys and student tests, although some research is available that describes classroom processes in greater detail (e.g., Stake and Easley, 1978). But there is no common method nor even objective. Some surveys and tests use whole populations, others are based on national or state samples, still others are characterized by self-selection of participants. Surveys may be cross-sectional, documenting a single point in time, they may be repeated at irregular intervals, they may be designed as longitudinal studies or repeated annually. Obviously, periodic replication of studies is necessary if temporal trends are to be identified. Yet there are constraints: careful thought must be given to reducing the response burden in surveys, the disruption that sometimes accompanies classroom observation and other qualitative research, and the unforeseen negative consequences of increased student testing. One way of limiting both the expense and the interference of periodic surveys, testing, and qualitative research may be to set up a carefully selected panel of schools, with systematic rotation of schools into and out of the panel, to provide a consistent data base.

- Much of the data used in the reports critical of education come from national surveys or nationally administered tests. However, education in the United States is decentralized and, despite some tendencies toward conformity, quite diverse in inputs, processes, and outcomes. The richness and sometimes even the meaning of information is obscured by reporting only national averages. Moreover, nationally aggregated statistics are of limited use in formulating policy at the state and local levels--and, after all, it is the states and localities that largely determine what happens in schools in this country, not a national body. Therefore, if the condition of science and mathematics education is to be portrayed so as to inform key people and policy makers involved in education, indicators must be selected to be useful at the state and local levels as well as at the national level.

- There is one important national goal not well met in science and mathematics education, the goal of equal opportunity. Information pertinent to this goal ought to be available. Therefore, it is important to collect certain data by gender and minority status, such as different enrollment rates in mathematics and science courses. Other demographic descriptors may also be important, for example, density of population or economic characteristics of different communities.

- Some indicators may have to be represented by different measures at different levels of education. For example, the teaching of science and mathematics in elementary school is not generally provided by specialist teachers and enrollment is not recorded by specific courses, as is the case in high school. A special problem in this regard is the middle or junior high school, which may be organized either like elementary or like secondary school.

- Supposing that indicators of mathematics and science education were available, how should they be interpreted? Most commonly, indicators are used to make comparisons over time: Have test scores risen or fallen in the last year, the last decade? Are students taking fewer or more science courses than last year, five years ago? Comparisons can also be made among groups or geographic entities, for example in examining distributional issues. Temporal trends can, of course, be established within comparisons of population groups, e.g., over the last five years, has mathematics course enrollment of females increased or decreased more rapidly (slowly) than that of males? A third basis for comparison is to establish an ideal value for an indicator and compare the recorded value to the ideal, as is attempted in teacher demand-and-supply studies. The problem with this method is that establishing ideal values is usually difficult, but all three methods of interpreting indicator values are appropriate, given proper caution.

Findings on the Information Pertinent to Selected Indicators

The committee published its first report in April 1985 (Raizen and Jones, 1985). Below is a summary of the findings in the report on the adequacy of the currently available data and information. Recommendations for improvements are also given.

The Quantity and Quality of Teachers

Findings

- Aggregate estimates of teacher supply and demand mask great differences among regions of the nation, states, and local school districts within states.

- All estimates of teacher supply and demand are accompanied by large uncertainties.

With respect to supply, there are three major gaps in knowledge:

(1) The data on the actual numbers of teachers assigned to mathematics and science classes are inadequate, especially as aggregated at the national level.

(2) The number of inactive teachers who return each year to fill vacancies is unknown. Since the number of trained teachers who do not enter teaching or who leave teaching is sizable, this represents a considerable resource. The number of teachers drawn from the inactive pool may increase as desirable job opportunities arise.

(3) The most recent data on the annual supply of newly certified entrants to teaching are four years old. Hence, the effects of current incentives to draw people into the field are unknown.

With respect to demand, there are four unknowns:

(1) While enrollments are dropping, vacancies tend to be filled with teachers from other fields who have tenure in a district, rather than with new entrants certified in the field with vacancies. This practice, the extent of which is unknown, reduces the demand for additional teachers, even though it may be detrimental to the quality of science and mathematics teaching.

(2) The extent to which school systems will seek to replace out-of-field teachers or will choose instead to provide in-service training is unknown. Such choices will in part be influenced by state and federal support policies for teacher education and in part by local board policies and teacher contracts.

(3) To the degree that increased high school graduation requirements will entail having to offer more courses in mathematics and science, teacher shortages will be aggravated, but how much is unknown.

(4) Demand forecasts are generally based on extrapolation of current conditions, taking account of likely changes in enrollment, class size, and curriculum. They do not take into account possible structural changes in the education system.

Since the committee completed its review, several studies have been initiated to address some of these gaps.

With respect to the quality of teachers, the following problems were identified:

- Adequate information is lacking on the qualifications of the teachers who are responsible for teaching mathematics and science in high school, middle/junior high school, or elementary school.

- Information on certification, the only proxy available for qualification, is lacking for all but new entrants, although data on a national sample of the teaching force are now being collected.

- Even when available, information on certification is of questionable use as a measure of qualification because state certification requirements and preservice college curricula reflect a wide range of views on what constitutes a qualified or competent teacher in mathematics or science. Moreover, teachers currently certified obtained their certification at different times that may have required different types of preparation; therefore, certification even within the same state does not connote equivalent preparation.

Conclusions and Recommendations

- A suitable indicator to assess the sufficiency of secondary school science and mathematics teachers would be either the ratio of or the difference between projected demand and anticipated supply of qualified teachers. The ratio would indicate how close to balance demand and supply are; the difference would indicate the number of teachers that need to be added or that exceed the demand. The construction of such an indicator on teacher demand and supply is at present not feasible at the national level because of the lack of a meaningful common measure of qualification.

- Individual states and localities might construct this type of indicator by using certification as an approximation for qualification or developing alternative criteria for teacher competence. In each case, an adequate determination would entail estimates of both demand and supply under alternative sets of assumptions about anticipated enrollments in mathematics and science classes and new entrants into the teaching of these fields. Aggregation of the state data might provide a useful national picture, especially if, in addition, information was reported concerning differences among states.

- The disparate views on teacher qualification and the variation in certification standards indicate the need to rethink the initial preparation and continuing training appropriate for teachers with instructional responsibilities in science and mathematics. Guidelines that have been prepared by professional societies need to be considered by the wider educational community, including bodies responsible for the certification of teachers and accreditation of teacher education programs. Requirements should be detailed separately for teachers in elementary school (grades 1 to 5 or 6), middle or junior high school (grades 6 or 7 to 8 or 9), and high school (grades 9 or 10 to 12), with particular attention to requirements that can be translated into effective college curricula and in-service education for teachers.

- The development of guidelines for the preparation and continuing education of teachers would be advanced if the attributes of successful teaching in science or mathematics were better understood. Further research is necessary on the relationships between teacher training and student outcomes; for example, the effects on student achievement of different types of preservice and in-service training and of teaching experience. Current initiatives to augment the pool of science and mathematics teachers should be monitored to assess their effectiveness.

Curriculum Content

Findings

- Although commonly used textbooks and tests introduce a modicum of similarity in the range of topics generally treated within a year's course of instruction, emphasis varies from text to text, class to class, and test to test. Hence, for the nationally normed achievement tests often used at the elementary and middle school levels, there may be a discrepancy between a student's opportunity to learn and the subject matter covered on the test, while at the same time the student may have learned considerably more than the test indicates.

- To a large extent, the content of instruction is based on the textbook used in a class, yet there is no continuing mechanism to encourage periodic and systematic analysis of the use and content of science and mathematics texts.

- At the secondary school level, and particularly in mathematics, course titles are a questionable indicator of content studied. The current practice of accepting similar course titles as representing exposure to similar material is likely to produce data of questionable quality.

- There are no established standards for content derived either from past practice, practice elsewhere, anticipated need, or from theoretical constructs developed, say, from the nature of the discipline being taught or from learning theory. Until some consensus can be reached on instructional content that represents desirable alternatives for given learning goals, it is premature to suggest a specific indicator for this area.

- Information on what is currently taught should be collected and analyzed, and reviews of the curriculum should be done by scientists, mathematicians, and other experts in the disciplines as well as teachers and educators. The reviews should evaluate material covered at each grade level or by courses, such as first-year algebra or introductory biology; consider relationships among grade levels or courses; and identify the knowledge and skills expected of students at the completion of each grade or course. Such reviews are needed in conjunction with addressing the critical matter of what content should be taught in mathematics and science.

- At a minimum, periodic surveys should be conducted to determine the relative frequency of use of various mathematics and science textbooks at each grade level in elementary school and for science and mathematics courses in secondary school. Timing of surveys should take into account the common cycles of textbook revision.

- Surveys of textbook use should be followed by content analyses of the more commonly used texts. Analyses should proceed along several different lines: balance between the learning of recorded knowledge (concepts, facts) and its application (process), emphasis given to specific topics, adherence to the logic of a discipline, opportunity and guidance for student discovery of knowledge, incorporation of learning theory.

- Intensive studies should collect information from teachers and students on topics actually studied within a given grade or course. Observation of samples of individual classrooms can help to document the content of instruction. Such studies could help to inform curriculum decisions by local districts, even though the results may not lend themselves to generalization over a state, let alone over the United States as a whole.

- Improved definitions of secondary school courses, based on their content, should be developed. As a first step, use of a standardized course title list, such as the Classification of Secondary School Courses (Evaluation Technologies, Inc., 1982), should be considered.

- Critical analysis of standardized tests should continue so as to establish their degree of correspondence to the instructional content of the class subjects for which they are used. Consideration should be given to inviting the judgment of teachers (and older students) concerning the students' opportunity to learn the material that is covered on each test.

Instructional Time and Course Enrollment

Findings

- For elementary schools, not enough data are available to discern clear trends over the last 20 years with respect to amount of instructional time spent on mathematics and science. Existing information, however, points to great variability from class to class in the amount of time given to instruction in general and to each academic area specifically.

- A number of problems attend enrollment data currently available: uncertainties generated by using self-reports, differences in questions and method from survey to survey, and ambiguities created by similar course titles in mathematics that refer to different content or different levels of instruction.

Conclusions and Recommendations

- The average amount of time per week spent on mathematics instruction and on science instruction should be measured periodically for samples of elementary schools. This measure would serve as an indicator of length of exposure to pertinent subject matter; values can be compared for different years. Care must be taken, however, to ensure common understandings in collecting measures of time as to what constitutes science or mathematics instruction. Time given to mathematics or science, expressed as a percent of all instructional time, would indicate the priority given to these fields.

- Efficiency of instruction should be assessed by comparing allocated time with instructional time and with time that is actually spent on learning tasks that appear to engage students, as established by observation.

- Experimentation and research should be carried out to develop a proxy measure for time spent on instruction that would permit collecting the pertinent information at reasonable costs.

- For grades 7 to 12, enrollments in mathematics and science courses at each grade level and cumulatively for the six years of secondary school or for the three or four years of senior high school should be systematically collected and recorded. Alternatively, the mean number of years of mathematics or science taken or percentages of students taking one, two, or three or more years of such courses can be used as a measure.

- The disparities in mathematics and science enrollment among various population groups warrant continued monitoring, so that distributional inequities can be addressed. National data on student enrollments collected in connection with the periodic surveys recommended above may be insufficient for this purpose. States should consider biennial or triannual collection of enrollment data by gender, by ethnicity, and by density of the school population.

Student Outcomes

Findings

- It has proved difficult with current test methodology to construct tests that can be used for large numbers of students and yet are adequate for assessing an individual's cognitive processes, for example, the ability to generalize knowledge and apply it to a variety of unfamiliar problems. However, existing tests of mathematics and science of the kind employed by NAEP, HSB, and IEA are sufficiently valid for the purpose of indicating group achievement levels.

Conclusions and Recommendations

- Systematic cross-sectional assessments of general student achievement in science and mathematics, such as the ones carried out through NAEP, should be carried out no less than every four years to allow comparisons over relatively short periods of time. The samples on these assessments should continue to be sufficiently large to allow comparisons by ethnic group, gender, region of the country, and type of community (urban, suburban, rural, central city).

- Longitudinal studies such as High School and Beyond are important for following the progress of students through school and later and should be maintained.

- International assessments in mathematics and science education such as those sponsored by IEA need to be carried out at least every 10 years.

- Developmental work on tests is needed to ensure that they assess student learning considered useful and important. Instruments used for achievement testing should be reviewed from time to time by scientific and professional groups to ensure that they reflect contemporary knowledge deemed to be important for students to learn. Such reviews may lead to periodic changes in test content--an objective that must be reconciled with the goal of being able to compare student achievement over time.

- Work is needed on curriculum-referenced tests that can be used on a wider than local basis, especially for upper-level courses. This work will require careful research on the content of instruction, tests constructed with a common core of items, and alternative sections of tests to match curricular alternatives.

- Assessments should include an evaluation of the depth of a student's understanding of concepts, the ability to address nonroutine problems, and skills in the process of doing mathematics and science. Especially for science, it is desirable that a test involve some hands-on tasks.

Current Work

The committee's first report represented a preliminary selection of indicators and review of relevant data and information. The report was criticized by scientists and mathematicians for concentrating on conventional indicators and current data bases rather than developing imaginative new indicators that would provide more penetrating insights on the condition of science and mathematics education. With support from the National Science Foundation, the committee is continuing its work. Some of the problems discussed above are receiving further attention, for example, defining teaching effectiveness, developing indicators of the quality of curriculum content, and improving assessment of student performance. Also, some potential indicators identified but not selected for discussion in the first report are being reexamined.

In the fall of 1985, the committee held a workshop involving some 50 outside experts to develop improved approaches to indicators. Three areas of schooling input and three areas of schooling outcomes were addressed: financial investment at the federal and state and local levels in mathematics and science education, teaching effectiveness, curriculum quality, learning of mathematics and science, motivation and attitudes, and scientific literacy. At this stage, it is not possible to report the committee's conclusions and recommendations, but I will try to highlight progress to date.

Financial Indicators

At the federal level, current budget information does not allow a separate accounting for federal resources invested in science, mathematics, and technology education. There are several reasons: with few exceptions, the budgets of agencies that support relevant activities do not include separate line items for science and mathematics education; programs designed as supporting science and mathematics education vary from agency to agency as to what subjects, educational levels, and populations are included; and agencies have different priorities regarding salient areas of policy interest and therefore maintain different budgeting categories. Unless considerably more resources were to be spent on creating compatible budgeting systems across agencies that would make possible identification of investments in mathematics and science education, a financial indicator at the federal level will be difficult to create.

In any case, federal financial indicators will be useful only if they can be considered in conjunction with state and local financial data. Federal resources are only a small part of the total educational effort, especially in precollege science and mathematics education. Hence, the federal contribution can best be appreciated in relation to the larger context, particularly if it could be shown how the federal funds are leveraging state and local funds. However, at these levels as well, a major constraint is the current variability in recording financial and accounting data. Moreover, school budgets generally do not reflect investments in instructional programs, and the budget categories that are in common use do not make it possible to generate the requisite figures.

Any attempt to profile financial information indicative of investments in science and mathematics education will require a significant R&D program. A useful initiative would be to experiment with a pilot management information system on science and mathematics education in a few selected districts, perhaps some 20 sites. Such a project would entail adoption of a standard cost accounting arrangement by the pilot districts and developing the needed reporting capacity at the local level to track funding of science and mathematics activities. Criteria for selecting pilot districts might include demographic characteristics (e.g., urban, suburban, rural) and character of the mathematics/science programs (e.g., programs meet minimum standards, programs are recognized as high-quality). If the feasibility of such an approach can be demonstrated, further issues that will need to be resolved include corrections for cost-of-living and other market conditions that vary among localities, the need to modify the data base over time as schooling practices change, and changing inputs to education, such as the changes in the demand for and supply of teachers.

Assessing Teaching Effectiveness

As pointed out in the committee's first report, a key problem in creating an indicator on teacher supply is defining teacher quality. Attributes that characterize effective teachers of mathematics and science include intellectual curriosity, subject matter knowledge, and ability to convey subject matter and intellectual curiosity to students. All three attributes, not just one or two, are important to effective mathematics and science teaching. The reason for stressing the combined importance of these attributes is to forestall the adoption of policies raising the average level of just one attribute--say, subject matter expertise--at the expense of lowering the level of the other two.

Identifying attributes important to effective science and mathematics teaching is only the first step; there remains the thorny problem of measuring these attributes. To assess intellectual curiosity, one may wish to consider teachers' use of time both inside and outside the classroom. For example, information might be collected on the extent to which teachers use their own time to keep up with their subject area, whether they take courses on their own, and whether they read such periodicals as Science and Scientific American. However, there are many ways of displaying intellectual curiosity, and it would be misleading to restrict measures to participation in a small, finite set of activities. Using professional time within school to pursue intellectual interests requires that time is set aside in the normal school day for such purposes. Appropriate space that allows teachers to think, read, and talk about ideas with colleagues also is required. Measures of such preconditions for maintaining an active and productive concern with one's teaching responsibilities might lead to desirable changes in the school environment provided for teachers.

One would assume that knowledge of subject matter, the second attribute, ought to be easiest to define and assess. Most recent reports on improving the quality of teachers have stressed the need for increased subject matter preparation. There is, however, a surprising lack of evidence in the literature on the relationship between increased knowledge of subject matter and teaching effectiveness. Obviously, teachers must understand the content they are expected to teach, but it is not clear what level of subject-matter knowledge beyond this minimum will improve their performance. They also need to have a sense of the structure of the discipline they teach and of the relevant importance of different facts associated with the discipline. This is not necessarily engendered by taking more undergraduate courses. Nevertheless, there appears to be consensus that high-school mathematics and science teachers, as a minimum, should possess the skills of an undergraduate major in the subject area. Thus, a gross general indicator would be to collect information on the proportion of teachers who have undergraduate majors in the subjects they teach. Because undergraduate majors in a particular science or mathematics vary enormously in substance and difficulty, this gross indicator should be augmented by accreditation models based on peer review, that is, judgments by subject-matter experts on what constitutes a reasonable academic program. Since many high school science teachers teach more than one science, data also need to be collected on the percentage of students taught by teachers having adequate subject matter preparation, as defined above.

The issue of adequate mathematics preparation for elementary-school teachers might be resolved by analyzing the mathematics content of the grades K-6 curriculum and then basing measurements of teacher quality on this analysis, for example, assessing the competence of teachers in solving problems that involve fractions and decimals. With respect to the science knowledge to be required of elementary school teachers, it is not possible at this time to define an indispensable core, since the science curriculum itself is so variously and vaguely defined and--in fact--often absent altogether in any form recognizable as science by expert observers.

Measures of teachers' competence in conveying subject-matter knowledge and intellectual curiosity to their students need to be based on minimum skills required of the teacher: exposing students to important concepts, presenting multiple representations of concepts, trying to understand students' thinking, asking students to explain their answers, giving clear directions. Observational techniques could be used to determine the extent to which teachers of mathematics and science exhibit these behaviors. Development of reliable indicators of such teacher behaviors are likely to be difficult and expensive, requiring extensive training of observers and observation of particular teachers for several hours, preferably on different days. Moreover, such indicators probably will not distinguish inspired teaching from merely adequate teaching, but they may identify whether basic conditions are satisfied for instilling in students both a knowledge of and positive attitudes about science and mathematics.

Quality of the Curriculum Content

Indicators are needed for all four aspects of curriculum content: the planned curriculum (e.g., state and local curriculum guidelines), the intended curriculum (e.g., the content of textbooks, laboratory exercises, and lesson plans), the content actually presented to the student, and the content learned by the student. If indicators of these differently defined aspects of curriculum content were available, comparisons could be made, for example, between what state and local policy makers intend and what goes on in the classroom.

Assessment of the content of instruction in mathematics and science cannot proceed without a framework that represents the structure of the subject matter and desirable learning goals, or alternatives among goals. Therefore, such frameworks need to be developed for curricular units. The frameworks must give attention both to the quantity of subject matter covered and to the depth of coverage. They must also consider over what unit of time curriculum should be defined--a school term, a grade in school, or a longer period of time such as grades K-6. Possible curricular units might be: K-8 mathematics, college-preparatory mathematics (algebra I, geometry, algebra II, analysis, calculus, and statistics), secondary-school mathematics for the non-college bound, K-8 science, college-bound science (biology, chemistry, physics), and secondary science for the non-college bound. The frameworks should meet some general criteria: they should array major processes, emphases, or principles in the curriculum against content topics rather than simply list detailed topics; they should represent the best thinking of a combination of disciplinary specialists and specialists in the design of curricula and in teaching the subject; they should be conceived to lead practice, rather than representing a least common denominator of current practice; and they should be flexible, presenting a commonly agreed-on core and allowing for major options or alternatives in the content presented in states, localities, schools, and classrooms. Over time, these frameworks should be regularly and critically reviewed so as to reflect developments in the disciplines.

Once a framework is developed, it can serve as the basis for a series of analyses. One analysis would sample state and local curriculum objectives, guidelines, and testing programs to determine which of the elements in the framework are covered and to gauge the depth of coverage and the variability in coverage among school systems. A second stage would be to identify the most frequently used textbooks and to use the framework to determine which elements are covered. Since textbooks represent the primary tool for planning curriculum content, this stage would yield indicators of the planned curriculum. Analysis should be done often enough to account for periodic textbook revisions. Another use of the framework would be as a basis for assessing the content actually covered in classrooms. One way to do this would be to observe classes, but this is expensive and time-consuming. Instead, a sample of teachers could be asked whether they covered various topics and to indicate the depth of covereage. Such surveys could be conducted in conjunction with student testing and would yield measures of students' "opportunity-to-learn" similar to those used in the IEA studies. Finally, the curriculum framework could be used to review the content of existing student achievement tests and design new ones. This would tie content analysis of the curriculum to indicators of student achievement.

Assessing Learning in Science and Mathematics

The present system for assessing student progress in acquiring knowledge, skills, and understanding in science and mathematics relies primarily on tests composed of multiple-choice items. Such items are used by NAEP in assessments at the national level and at state, district, and local levels. Multiple-choice tests are worth retaining because they provide an efficient and economical method for assessing the extent to which students have acquired necessary factual information and such elementary procedural skills as the algorithms involved in arithmetical computation.

There are, however, some important ways in which the current system for assessing student achievement is deficient:

1. One should be able to assess those educational objectives that are poorly measured by present methods. Existing multiple-choice tests appear to exclude a variety of critically important educational goals.

2. An improved system should provide diagnostic information that would be useful in helping students who fail to understand or master certain skills.

3. An improved system should help teachers make wise decisions about what to teach and how to teach it. Available assessment materials appear to lead teachers to focus on the low-level objectives that are reflected in multiple-choice tests. Improved tests should provide models of performance reflecting the more complex cognitive skills, models that should be emulated by both teachers and students.

4. An improved system should produce tests that are coachable only in the good sense that teaching a student to deal with the test problems is equivalent to teaching the knowledge and skills the test is intended to measure rather than improving test scores without improving skills.

5. An assessment system should be based on some understanding about curricular priorities. Unfortunately, the present fragmentation and lack of consensus with regard to curriculum goals, especially in science, constitutes a potential barrier to the development of instruments for assessment.

Some innovative assessment methods are under development. Continuing advances in cognitive theory promise to provide ideas regarding what skills should be assessed, and computer science promises better tools with which to assess them. For example, research on the misconceptions that many students have regarding physical phenomena provides instances of how better understanding of cognitive processes may lead to more useful assessments. Another example is the research on differences between novice and expert problem solvers. Such studies strongly suggest that methods for discovering how one represents a problem internally and how these representations are altered with training and practice would be of great value in instruction. A third example is research concerned with the extent to which certain aspects of problem solving can be performed automatically, with a minimum of attention, through pattern recognition and other short-cuts.

Computer simulations of "hands-on" performance in doing experiments allow assessment of student skills in scientific thinking. It is also possible, through analyzing protocols of student input, to develop a refined picture of a student's processes for solving problems. Coaching systems developed for computer use could be adapted for assessing performance in very complex domains. Thus, the computer opens up several new possibilities for assessment. Paper-and-pencil simulations represent another option: Descriptions of proposals, research investigations, and experimental results can be presented to students for analyses, criticism, and explanation. Trained coders can then assess the quality of student responses.

The methods of assessment described above obviously cannot compete with multiple-choice tests from the standpoint of economy and efficiency. However, the cost of using the methods should be justifiable, not only because they would provide information for a far more accurate and complete assessment of instruction, but also because they would be directly useful in the educational process. Exercises derived from the methods could be used for practice and to provide information for remediation, and assessments based on the methods should raise educational standards by providing models of performance to be emulated by both students and teachers.

Developing alternatives to multiple-choice tests will require considerable creative effort and funding investment. Serious consideration should be given to the creation of a National Library of Science and Mathematics Assessment Materials. This library would be a compendium of science and mathematics exercises designed both for testing and for classroom use as instructional aids. The materials in the library would be publicly available, and a system for receiving feedback from users could facilitate improvements in content and methodology. The library materials should be created by a consolidated effort of scientists, educators, and testing professionals. Its philosophy should be to teach science and mathematics not as rote knowledge but as active problem-solving and to emphasize basic concepts and methodologies.

Assessment of Attitudes and Motivation

Current research has not been highly successful in establishing causal relationships between student attitudes and motivation and student achievement, although frequent correlations have been noted. The apparent lack of connectedness could be a function of three things:

1. The constructs of attitudes and achievement that have been used are unrelated.

2. The measures of attitudes/motivation are faulty.

3. The measures of achievement are faulty.

Generally, it has been assumed that the problems reside in either 1 or 2, but it is quite possible achievement measures in current use are not sufficiently broad to connect up with attitudinal indices. In any case, motivation and attitudes are worth assessing as educational ends in themselves, regardless of achievement.

The choice of indicators in this domain should be guided by constructs that can explain attitudes of students towards science and mathematics and their motivation for learning and performing in these fields. Three constructs considered important are engagement--choosing to attend to ideas and applications of science and mathematics; competence--knowing what strategies are needed to be successful in these fields and being in command of the strategies; and autonomy--the sense that one chooses to engage in an activity for one's own purpose.

Several principles should govern the development of indicators of attitudes and motivation toward science and mathematics. First, qualitative as well as quantitative methods will be necessary to assess the identified constructs. National and state level assessments should be done through large-scale surveys, but at the same time smaller subsamples should be interviewed intensively so as to connect results from quantitative and qualitative research. Second, for both qualitative and quantitative studies, measures need to be domain-specific as well as assessing more general processes. And third, assessment strategies must be developmentally appropriate. The meaning of science will differ between 3rd grade and 12th grade, so the methods for measuring both attitudes and achievement must be sensitive to the student's stage of comprehension.

Regarding competence and autonomy, survey questionnaires and interviews should establish how students perceive their control over understanding mathematics and science in and out of school, how they perceive their capacities, and the degree to which children experience choice in learning mathematics and science. Questions should also probe the extent to which children view mathematics and science as being connected to other meaningful aspects of their lives. More indirect indicators may have to be developed to assess engagement. One such indicator might be the extent to which students choose to participate in opportunities to learn science; another might be the career interests of students. Various non-school behaviors also may reflect student engagement with science--a predisposition to read science materials and view science television and film programs, voluntary participation in science projects, using deductive (scientific) learning to evaluate information outside of science class, and solving puzzles, word problems, and computer simulations for fun.

Indicators of General Science Literacy

Any useful attempt to determine the status of scientific literacy in the population must reflect an informed view of what constitutes scientific literacy. At least four dimensions need to be addressed:

1. The nature of the scientific world view;

2. The nature of the scientific enterprise;

3. The role of science in human affairs; and

4. Scientific habits of mind.

Each of these dimensions is elaborated below.

The scientific world view is made up of ideas and beliefs at various levels of complexity. Grand conceptual schemes are formulated to bring together and reduce to order large numbers of ideas, theories, and observations that are of lesser generality. Particular discoveries and ideas may lead to the development of new theories and are powerful tools for making sense of natural phenomena. The various sciences tend to operate in the context of firmly held notions about the natural world which is assumed to be understandable and not capricious.

Laymen also need to understand that science is a social activity carried out by scores of individuals who collaborate over time and place. It is important to be aware that scientific findings are always tentative, open to modification and correction as new discoveries are made. The scientific enterprise is conducted by individuals who subscribe to a set of value commitments in principle, but in as much as they are human beings, there are variations in the actual behavior of individual scientists.

To a very large extent, people encounter science as it affects human events. The application of science to such matters as health, agriculture, and the environment engage not only scientists but also a large number of policy makers and individual citizens. In a scientifically literate society, people will need to understand some of these relationships. For example, scientists engaged in public matters may behave differently than when they are acting as reseachers. Tensions will always exist between science and society because science is able to deal with uncertainty and ambiguity to a degree that is not possible by organizations that must make policy decisions or by individuals who must make personal ones. Also, the differences and overlaps between science and technology need to be understood as well as the fact that both are likely to have unpredicted impact on society.

Scientific habits of mind include familiarity with natural phenomena; identifying questions or formulating hypotheses relative to problems posed by these phenomena; identifying and seeking out relevant information; using that information to test the hypotheses, answer the questions, or creating new hypotheses; and offering arguments and counterarguments that can be tested by reference to data or accepted principles.

Any plan to generate indicators of scientific literacy should try to estimate the degree to which a population possesses the kind of knowledge and intellectual skills outlined in the above four categories. In doing this the following attributes ought to prevail.

1. A single measure will not do for these multiple dimensions of a complex set of characteristics. The indicators should be matched to the model of scientific literacy.

2. Indicators should be flexible, changing over time in form and content to match changes in the model itself.

3. Even so, there must be enough continuity in the indicators to make it possible to track populations over time.

4. The indicators should recognize that there is no absolute level of literacy and that various levels of attainment in different components of a community or population group are likely.

5. Any measures used to generate indicators should be supplemented by research.

6. Indicators may be expressed in terms of descriptive patterns of behavior and other non-numerical ways.

7. Indicators can be expressed as group literacy as well as individual literacy. Any conclusions based on the indicators must relate only to the unit of analysis.

Current techniques of conducting polls, interviews, and case studies should all be considered in generating indicators. Traditional methods may work reasonably well to assess knowledge, but they should be further developed to probe also the population's understanding of the nature of science and its role in society. It is particularly important and difficult to obtain reliable estimates of problem-solving skills. Assessment of such skills need to go beyond individual paper-and-pencil tests and should include observation and analysis of individual and group responses to carefully selected phenomena involving real objects and filmed sequences of events.

==#==

At this time, the committee is collecting reactions from various groups to its suggestions. I look forward to the discussion period to elicit yours.

References

National Academy of Sciences and National Academy of Engineering
1982 Science and Mathematics in the Schools: Report of a Convocation. Washington, D.C.: National Academy Press.

National Science Foundation
1983 Educating Americans for the 21st Century. National Science Board Commission on Precollege Education in Mathematics, Science and Technology. Washington, D.C.: National Science Foundation.

National Commission on Excellence in Education
1983 A Nation at Risk: The Imperative for Educational Reform. Superintendent of Documents No. 065-000-00177-2. Available from the U.S. Government Printing Office. Washington, D.C.: U.S. Department of Education.

Twentieth Century Fund Task Force
1983 Report of the Twentieth Century Fund Task Force on Federal Elementary and Secondary Education Policy. New York: The Twentieth Century Fund.

Boyer, Ernest L.
1983 High School: A Report on Secondary Education in America. New York: Harper & Row.

Peterson, Paul E.
1983 Did the education commissions say anything? The Brookings Review (Winter):3-11.

Stedman, Lawrence C., and Smith, Marshall S.
1983 Recent reform proposals for American education. Contemporary Education Review 2(2):85-104.

Raizen, Senta A., and Jones, Lyle V.
1985 Indicators of Precollege Education in Science and Mathematics: A Preliminary Review, Washington, D.C.: National Academy Press.

Stake, Robert E., and Easley, Jack A., Jr.
1978 Case studies in Science Education. NSF SE-78-74. Available from the U.S. Government Printing Office. Washington, D.C.: National Science Foundation.

Evaluation Technologies, Inc.
1982 Classification of Secondary School Courses. A report prepared for NCES. Superintendent of Documents No. NCES 300-81-0132. Available from U.S. Government Printing Office. Washington, D.C.: National Center for Education Statistics.