DOCUMENT RESUME

ED 340 601 SE 052 474

AUTHOR	Blank, Rolf; Selden, Ramsay
TITLE	Alternative Methods for Assessing Science: Report to the States. Science & Mathematics Indicators Project. Conference Report of State Assessment Directors and State Science Supervisors (Tampa, Florida, January 13, 1989).
INSTITUTION	Council of Chief State School Officers, Washington, DC. State Education Assessment Center.
SPONS AGENCY	National Science Foundation, Washington, D.C. Directorate for Science and Engineering Education.
PUB DATE	Feb 89
NOTE	26p.
PUB TYPE	Collected Works - Conference Proceedings (021)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Academic Achievement; *Educational Assessment; Elementary Secondary Education; *Evaluation Methods; *Evaluation Problems; Informal Assessment; *Science Education; Testing; Validity
IDENTIFIERS	*Science Achievement

ABSTRACT

State assessment directors and State science supervisors discussed alternative methods for assessing student learning in science at a conference. The conference had two objectives: (1) to increase the knowledge of state science supervisors and assessment directors of recent experience at international, national, and state levels with alternative methods of student assessment in science, such as hands-on exercises; and (2) to inform and assist states in planning alternative methods for state science assessment programs. Presentations and discussions of alternative methods by Senta Raizen, Rodney Doran, Walter McDonald, Joan Baron, Douglas Reynolds, Edward Roeber, and Zack Taylor are provided. A discussion on the issues of validity, role of assessment to "drive" versus "reflect" the curriculum, time requirements, cost, and use for trend data concerning the development and use of alternative methods of assessment in science is included. (KR)

Council of Chief State School Officers
State Education Assessment Center

SCIENCE & MATHEMATICS INDICATORS PROJECT

National Science Foundation

February 1989

ALTERNATIVE METHODS FOR ASSESSING SCIENCE: REPORT TO THE STATES

CCSSO STATE EDUCATION ASSESSMENT CENTER
400 North Capitol Street, N.W. o Washington, D.C. 20001 o (202) 624-7700

Council of Chief State School Officers
State Education Assessment Center

ALTERNATIVE METHODS FOR ASSESSING SCIENCE: REPORT TO THE STATES

February 1989

This report summarizes the results of a conference of State Assessment Directors and State Science Supervisors which was held on January 13, 1989 in Tampa, Florida. The conference was organized by the CCSSO Science/Math Indicators Project and sponsored by the National Science Foundation.

Report prepared by:
Rolf Blank and Ramsay Selden
Council of Chief State School Officers
400 N. Capitol Street, Suite 379
Washington, D.C. 20001
(202) 624-7700

ALTERNATIVE METHODS FOR ASSESSING SCIENCE: REPORT TO THE STATES

The Council of Chief State School Officers (CCSSO) convened state assessment directors and state science supervisors to discuss alternative methods for assessing student learning in science. The conference was planned and organized by the CCSSO Science/Math Indicators Project. Funding was provided by the National Science Foundation (NSF), Office of Studies and Program Assessment, Science and Engineering Education.

OBJECTIVES

The conference had two objectives: a) to increase the knowledge of state science supervisors and assessment directors of recent experience at international, national, and state levels with alternative methods of student assessment in science, such as hands-on exercises; and b) to inform and assist states in planning alternative methods for state science assessment programs. The longer term goal is to increase the number of states using alternative methods in assessing science and to increase the consideration of alternative methods with national assessment programs.

RECENT DEVELOPMENTS IN SCIENCE ASSESSMENT

The National Science Foundation has recently supported two projects that developed new methods of assessing student knowledge and skills in science. In 1986, an experimental study was conducted by the Educational Testing Service which tested new hands-on science exercises and analyzed their potential application to large-scale assessment programs. In November 1988, NSF sponsored a conference of researchers and educators to review the existing knowledge concerning the use of alternative methods of science assessment in national and state-level assessments. NSF is also supporting several major projects to

develop new curriculum and materials for elementary science which will demand new techniques in how student learning is assessed. The U.S. Department of Education is currently supporting a National Center for Improving Science Education which is focusing on how student assessment affects science education and how assessment might be improved. At the January conference, information and findings from these efforts were disseminated, so state representatives could consider alternative methods of science assessment in light of these developments.

The number of states with science assessment programs has more than doubled in the past four years (now 29 states), and many states are considering how they can design science assessments which are more consistent with state objectives for curriculum and instruction in science. States have also expanded their role in setting goals and objectives for elementary and secondary science. Based on a 1987 CCSSO survey, 38 states have a curriculum framework or standards for science education (Blank & Espenshade, 1988). The frameworks are used to select or recommend textbooks, design student assessments, guide school curricula, and improve teacher training and in-service. Now, one of the concerns of states is how to improve the "alignment" of state curriculum goals with what is tested in science assessments.

Many state curriculum frameworks emphasize teaching science as process, as opposed to science as facts. One way of ensuring that state assessment programs reflect science as process is to develop and implement direct methods of testing student knowledge of these processes, such as hands-on science exercises.

ORGANIZATION OF THE CONFERENCE

With these interests in mind, the planning conference was designed as a forum for state representatives to review the recent models for and developments in science assessment and to learn from the experience of states, such as Connecticut, New York, California, and Michigan, that have begun to incorporate hands-on exercises and other alternative methods into their assessment programs. The conference also provided an excellent opportunity for states to share knowledge, ideas, and strategies for improving their assessment programs.

The conference was scheduled to coincide with the mid-winter meeting of the state assessment directors. CCSSO requested that chief state school officers send their state science supervisor, as well as their assessment director, to the conference. A total of 60 state education staff participated in the conference representing 39 states.

CCSSO invited presenters from states as well as national experts on science assessment (as shown on the attached agenda). The national experts included: Senta Raizen (National Center for Improving Science Education), Rodney Doran (Second IEA Science Study), and Walter McDonald (National Assessment of Educational Progress' science assessment). The state presenters incluued: Joan Baron (Connecticut), Douglas Reynolds (New York), Ed Roeber (Michigan), and Zack Taylor (California).

Wayne Welch, head of the Office of Studies and Program Assessment at NSF provided an overview of current studies and activities of the Office. Ramsay Selden, Director of the CCSSO State Education Assessment Center, explained how innovation in science assessment is needed to correspond to desirable state and national goals in science curriculum, and Rolf Blank, CCSSO Science/Math Indicators Project Director, explained the development and role of the conference in the Project's efforts to improve state-level indicators.

PRESENTATIONS AND DISCUSSION OF ALTERNATIVE METHODS

SENTA RAIZEN presented findings from a new report of the National Center for Improving Science Education, entitled, "Assessment in Elementary Science Education" (1989). The Center's mission is "to promote changes in state and local policies and practices in the science curriculum, science teaching, and the assessment of student learning in science." Towards this goal, the report synthesizes findings concerning assessment in elementary science based on recent studies and experiments and recommendations of an advisory panel of scientists, educators, and assessment experts. The report was written to serve as a practical resource for policy-makers and educators.

Raizen's presentation focused on how assessments can be designed and used to improve instruction in science. The report takes the position that assessment in elementary science must be viewed from the perspective of the elementary classroom teacher. It recommends improving the alignment of curriculum content, classroom assessments of instruction, and district and state assessment programs. At the same time, there should be greater national, state, and local correspondence in the content, methods, and uses of science assessments.

Raizen outlined what should be assessed in elementary science:

1. Science Knowledge
   - o factual
   - o theoretical
   - o about the scientific enterprise

2. Science Skills
   - o laboratory skills
   - o science thinking skills
   - o generic thinking

3. Disposition
   - o applying science knowledge and skills

4. Learning Over Time

Among the various methods of assessment, the report states that the dominant paper-and-pencil method found in national, state, and local assessments for monitoring is mainly useful for testing science knowledge. Reliance on this method is not consistent with efforts to teach science through inquiry. The report strongly urges new forms of assessing science that would be appropriate to measure science skills and students' disposition toward science. However, the report cautions that the process of developing valid skill assessment exercises is more complex than that for knowledge assessment.

Raizen highlighted three concerns with skill assessment:

a) Laboratory equipment reveals the difference between knowing how to do something and being able to do it;

b) Assessing intellectual skills of science, such as being able to design an experiment, introduces the additional distinction between generic thinking skills and thinking skills specific to the scientific area of the experiment;

c) Interpreting and scoring performance requires agreement by observers on standards and consistent application of standards; and

d) Administration of alternative techniques, observing them, and scoring them are more labor intensive than paper-and-pencil tests and require trained test administrators and scorers.

To improve elementary science in the classroom, Raizen outlined findings concerning methods of assessment that should be available to teachers Multiple methods must be employed, from short-term formative methods, such as written quizzes and tests, to longer-term summative methods, such as records of student work and documentation of systematic observation of students. One way to judge the appropriateness of current methods of assessment in science is to ask a series of questions about tests:

1. Are there problems that require students to think about and analyze situations?

2. Are there some problems that call for more than one step to arrive at a solution?

3. Are there problems with more than one correct solution?

4. Are students encouraged to use a variety of approaches to solve a problem?

5. Is there opportunity for assessing laboratory and science thinking skills through hands-on-exercises?

6. Are there opportunities for students to make up their own questions, problems, or designs?

A general emphasis of the report is that science assessment should be used as an entry point for improving instruction. To accomplish this objective, the report strongly recommends that teachers, curriculum supervisors, and principals be "brought on board" with what states are trying to accomplish through assessment programs.

RODNEY DORAN, professor of science education at the State University of New York at Buffalo and associate national research coordinator for the Second International Science Study (SISS), explained and demonstrated the "science process laboratory skills test" which was a part of SISS. The presentation was based on the 1988 report, "Science Achievement in the United States and Sixteen Countries."

The SISS was conducted in 17 countries in 1983, and in the U.S. in 1986, with assessments of national representative samples of fifth and ninth grade students. The first international science assessment was conducted in 1969-70. In the two ensuing decades science education had emphasized teaching science process skills. Methods of evaluating or assessing these skills had not kept pace with innovation in the curriculum. Thus, the second international study included an innovative, optional science process skills test, and six countries participated in the skills test. In the U.S., the skills test was administered to a nationally representative sample of 2500 fifth grade students and 2300 ninth grade students.

At each grade level, the process (lab) test had six "live" exercises:

FIFTH GRADE

- Describe and explain color change of bromthymol blue solution after blowing through a straw. (Chemistry)
- Cite at least three similarities and differences of two plastic animal specimens. (Biology)
- Determine if four objects are electrical conductors by testing in a battery-bulb circuit. (Physics)
- Predict and measure the temperature of the mixture of equal amounts of hot and cold water. (Physics)
- Observe and explain the dissolving of coffee crystals in water. (Chemistry)
- Determine which seeds contain oil by rubbing them on paper. (Biology)

NINTH GRADE

- By testing with battery-bulb apparatus, determine the circuit within a "black box." (Physics)
- Using phenophthalein and litmus paper, prepare and execute a plan to identify three solutions as to being acid, base, or neutral. (Chemistry)
- Using iodine solution, prepare and execute a plan to determine the starch content of three unknown solutions. (Biology)
- Using a spring scale and graduated cylinder, determine the density of a metal sinker. (Physics)
- Explain movement rates and separation of water soluble dots in paper chromatography activity. (Biology)
- Using a sugar test tape and iodine solution, identify three unknown solutions as to presence of starch and/or sugar. (Chemistry)

Doran demonstrated the tasks for the ninth grade test so that conference participants could visualize the degree of difficulty, the materials required, and how each task was administered to students.

The tasks were designed to be consistent with the expected level of learning at the grade levels as well as to employ materials that would be typically used in schools. For the test, a research contractor was responsible for sending the test materials to each of the 140 schools in the U.S. that were selected for the study. It was economically feasible to administer and score the exercises to 2,300 students at each grade level. Life science tasks were more difficult to design and incorporate due to the problem of shipping organic materials across the country. Life science tasks were included which used materials that could be easily shipped.

Several aspects of the test organization and administration were discussed. The room for test administration was organized so that adjoining tasks were not the same, and 12 students could be tested at a time. Each student was allowed 10 minutes to complete the task. No group tasks were used, only individual tasks,to reduce the complexity of scoring results. To increase the reliability of results, test administrators were specially trained and responses were centrally scored by trained scorers.

Findings were provided in the full report, but one finding of particular interest was that girls scored equally as well as boys on these hands-on exercises. Most paper-and-pencil science tests show boys scoring better than girls. About half of the teachers of the tested students reported that the tested activities were among the instructional experiences of students in their school during the year or a previous year.

WALTER MCDONALD, science assessment coordinator for the National Assessment of Educational Progress (NAEP), presented results from NAEP's experimental study of the feasibility of hands-on assessment in science. The study was conducted by NAEP staff at the Educational Testing Service and supported by the National Science Foundation. The study report, entitled "Learning by Doing: A Manual for Teaching and Assessing Higher-Order Thinking in Mathematics and Science," includes 11 exercises that were field tested in the study (1987).

McDonald reviewed the rationale for developing and using hands-on assessment techniques in science. Like other presenters, he emphasized that the lack of these techniques in current assessment methods constrains the integration of hands-on and laboratory instruction into science teaching. As a result, many students have little opportunity to learn how to apply scientific concepts, or to actually "do science."

The study involved 1000 third-, seventh-, and eleventh-grade students from 12 school districts across the country. Twenty-two administrators conducted the tests in teams during April 1986. Approximately 100 to 300 responses were obtained for each task.

The tasks were designed to test a hierarchy of skills:

First Level-- Classify and sort

Second Level-- Observe, infer, and formulate hypotheses using materials, equipment, or apparatus that represent scientific or mathematics phenomena or relationships

Third Level-- Detect patterns in data and interpret the results

Fourth Level-- Design and conduct complete experiment.

McDonald showed slides of the materials and apparatus that were used to test many of the exercises and explained how the tasks were tested and responses scored. Each student was tested individually at eight stations with eight minutes allowed per station. Students were tested in some tasks as groups, because that is how science often takes place. There was one complete investigation in which the student designs an experiment and carries it out. Overall, the study included 30 different tasks--six group activities, 20 individual stations, and four complete experiments. (Each student took only a subset of tasks.) Many of the tasks were adapted from those designed by the Assessment Performance Unit in Great Britain.

The results provided useful data on the quality and appropriateness of the tasks for each grade level, and they can provide a sound basis for further development of the task designs. Student responses reflected the hierarchy of expected skills as well as the grade levels that were tested. The results demonstrated that "hands-on assessment is feasible and extremely worthwhile" (Learning by Doing, p. 7). The study also showed that managing the equipment and training administrators and scorers requires considerable effort and

preparation. Developing standards for scoring the different possible correct responses to each task was found to be very important.

During the discussion, a question was raised about the validity of scores on a given task in relation to the science objective being tested, since there are different possible correct responses. McDonald indicated that the study established criteria for categorizing the types of responses to the tasks and for scoring the responses. Some open-ended information was obtained by test administrators on how and why students gave their responses.

Validation studies of the relationship of test responses to the intended objective should be done, just as would be done with paper-and-pencil multiple choice tests. A current limitation of hands-on exercises for large-scale assessments is the small number of pre-tested exercises that are available to assess a given science learning objective. More exercises will be needed in order to equate from one assessment period to the next.

JOAN BARON, director of the Connecticut Assessment of Educational Progress, outlined the findings from a "practical test" of science skills that was administered to a subset of the students in the 1984-85 statewide assessment of science in Connecticut. The exercises required students to measure and observe things, manipulate objects, and conduct simple investigations. The results of the test are described in a report, "Connecticut Assessment of Educational Progress 1984-85: Science."

For the practical skills assessment, three hundred students were tested at each of three grade levels (4, 8, 11). Ten students per grade were tested in each of 30 schools, with the exercises administered by a trained administrator The exercises were adapted from the Assessment Performance Unit of Great Britain, and thus there were no design costs for the exercises.

Baron outlined in her presentation the ways in which the results of the test were interpreted and used in evaluating science learning and instruction in the state. With many of the exercises, students were asked questions about their findings from the exercises and how they arrived at the findings or what they could conclude. For example, after timing 40 swings of pendulums with different weights and string lengths, fourth grade students were asked what they could conclude about pendulums. Sixty-five percent correctly concluded that shorter pendulums swing faster, eight percent concluded the opposite, and 24 percent gave conclusions not comparing speeds.

Students were asked similar questions on the written portion of the science assessment as on the practical skills portion, and responses from the written test were compared to the performance exercises to determine how students responded after experience with a concept as opposed to just being taught the concept. For example, students were given a battery pack with wire leads, a light bulb with wire leads, and insulated wires with alligator clips for connectors. They were instructed to use these materials to make the bulb light, and 85 percent of fourth graders succeeded. On a multiple choice item requiring the students to identify the picture of a simple circuit in which the bulb would light, only 46 percent of the fourth graders succeeded. Students were also asked how much experience they had had with the equipment used in the exercises, and it was found that those with more experience had higher scores on the skills tests.

Baron also reported some general observations concerning instruction in elementary science from a study she conducted in 31 elementary classrooms during the fall of 1988. The classrooms were selected by principals as those having better science teachers. She found that much of the instruction involved hands-on science techniques, but in many of the classrooms there were missed opportunities for having students actually do experiments. For example, there

were few instances of use of predictions, variables, and controls, and students were seldom asked to generalize about what was learned from the hands-on lesson.

Connecticut is developing a new secondary-level assessment called the Common Core of Learning. This assessment will involve multiple forms of assessment and will be designed to test student knowledge and skills across the core subject areas. Baron distributed copies of the basic curriculum objectives to be assessed.

DOUGLAS REYNOLDS, state science supervisor for New York, described the state's "science manipulatives test" to be administered to all 200,000 fourth grade students in the state for the first time in May 1989. The manipulatives test is one part of a statewide model for elementary science program evaluation. The evaluation will produce school-level results for program analysis, not individual student scores.

The fourth grade test and program evaluation model are part of the overall state plan for improving elementary science education. The four elements of the plan are: a) mandate instructional policies, b) assess programs, c) integrate science into the elementary curriculum, and d) provide a teacher and school support system. The state plan for elementary science has the primary goal of developing students' capacity for problem solving, with three kinds of learning expected: science attitudes, skills, and content.

The assessment approach consists of two required components, a written test and the manipulatives test, and five optional components, a survey of student attitudes toward science and surveys about the science program with students, teachers, administrators, and parents.

The manipulatives test will be administered in all of the 4,000 elementary schools in New York. Each school will have one test administrator who has been

trained in one of a series of regional training workshops. All fourth grade students in a school will rotate through one room in which the exercises will be placed. Reynolds demonstrated an example of each of the five types of exercises:

- Measuring physical properties, such as length, temperature, and volume;
- Predicting an event, such as variation in absorption of a liquid;
- Creating a simple classification system, such as with types of seeds;
- Testing objects to make a generalization, such as with an electrical circuit;
- Making inferences, such as about objects in a sealed box.

Reynolds outlined the structure and development of the state Elementary Science Mentors System for teachers. The goal of the system is to improve elementary science through a network of mentors who have received training in assisting elementary teachers with science instruction. The system has 93 regional mentors, 1000 school district mentors, and 4000 school mentors.

EDWARD ROEBER, Michigan's assessment director, described the state's experience with alternative methods of science assessment. He outlined the methods that were used in the three statewide science assessments in 1974, 1980, and 1985. He also explained steps that Michigan has taken to keep costs of test design and administration low. At the same time, the state has tried to introduce innovations that increase the usefulness of test results for teachers as well as for policy-makers and administrators.

The state introduced performance assessment exercises in the 1974 science assessment. University professors in the state assisted with the development of items to avoid the high cost of contracting for test development. Test administrators for the performance exercises were unemployed teachers and graduate students who could be hired on a part-time basis and sent out to

schools. A sample of students at selected grade levels were given the performance exercises.

In 1980, the state had to reduce the funding for assessment and there were no funds for performance assessment in science. As a partial substitute, a series of open-ended, paper-and-pencil items were used to assess problem-solving and reasoning skills of students. To reduce the staff time necessary to score open-ended responses, some of the student responses were in a scale or graph form which are machine-scorable.

Michigan has taken several steps to make the state science assessment more useful and accessible to teachers. To convey to all teachers the learning objectives that are being tested, copies of all of the items are sent to each teacher. The state carries out validation studies of the multiple choice items through field interviews with teachers and curriculum specialists. The studies provide information on why students responded with the correct answer or answered incorrectly. Teachers are also provided assistance with how tests in science can be improved and how informal assessment methods can be used in the classroom.

ZACK TAYLOR, science consultant for California, gave a presentation on the development of science assessments as part of the California Assessment Program. The eighth grade science test which is currently being used has the primary objective of evaluating school science programs. The test is designed with 36 unique forms and 15 items on each form. Thus, the overall program assessment is based on a total of 540 items, but individual students take only a small portion of the items. Taylor disseminated copies of the "Rationale and Content" booklet for the eighth grade science test (1985).

Currently California is conducting field tests for new sixth grade and 12th grade science tests. Two field tests will be conducted with each test prior to

the first statewide testing in 1991. Hands-on exercises will be added to the paper-and-pencil test. A team of staff from the California department of education went to Great Britain last summer to learn about the work of Assessment Performance Unit in developing hands-on exercises in science. This year a volunteer group of California teachers have been asked to design hands-on exercises and try them out in their classrooms. The examples will be used in developing a pool of items for consideration in the state science assessments.

ISSUES WITH ALTERNATIVE METHODS

During the course of the conference at least five issues were raised concerning the development and use of alternative methods of assessment in science, especially hands-on exercises: 1) validity, 2) role of assessment to "drive" vs. "reflect" the curriculum, 3) time requirements, 4) cost, and 5) use for trend data.

VALIDITY. Some hands-on exercises can have several possible correct responses based on different methods of reasoning used by students. This leads to the question of how to determine if an exercise is testing the desired objective, since the "desired objectives" appear to be multiple and indeterminate. Several presenters suggested that validation studies must be carried out for these exercises, just as with paper-and-pencil tests, but aimed at their more sophisticated or multiple goals.

In the judgement of several presenters, the most important kind of test validity that can be attained through use of alternative methods of science assessment is construct validity, or "ecological validity," which most testing specialists now regard as the only real issue in validity. These new testing methods seem to provide measures that are consistent with the goals and objectives of the science curriculum. Since different kinds of knowledge and

skills are desired in science instruction, it seems reasonable that multiple forms of assessment are needed to assess outcomes. A related issue is reliability of scoring due to the method of observation and interpretation of results typically used. The experience with international assessments and the NAEP study was that inter-rater reliability can be attained but careful training and supervision are needed.

DRIVING VS. REFLECTING THE CURRICULUM. The use of hands-on exercises and other alternative forms of assessment may "drive the curriculum" instead of reflecting it, by introducing concepts and giving them an importance that would not be there if they were not tested. From the viewpoint of state science supervisors, alternative methods of assessment are needed in order to reflect, track, and encourage the reforms states have been and are currently implementing in elementary and secondary science. Even though more work must be done on design and validation of hands-on exercises, need for these kinds of exercises should be an issue based on the states' curriculum frameworks and instructional goals.

Related to this issue, some testing specialists expressed reservations about such hands-on exercises "contaminating" student knowledge rather than simply tapping it. If students are not otherwise taught these skills, the test itself may teach them, leading to the erroneous conclusion that the skills are being taught systematically and generally to all students.

TIME FACTORS. The use of hands-on exercises does require additional time for science assessment. The assessments described in the conference allowed from 8 to 10 minutes per exercise for individual student exercises. With five to six exercises for the assessment, about an hour would be required for each student. The New York assessment, which is for all fourth grade students, is organized to have all students in a school rotate through one room which is organized with

multiple stations. With this approach a number of students can be tested at one time.

COST FACTORS. Based on the conference presentations, three types of costs can be identified: 1) cost of designing hands-on exercises, 2) administration and equipment costs, and 3) scoring costs. Each of the assessments reviewed in the conference adapted exercises from existing exercises, which kept costs low. A recommendation was made to establish a pool of exercises which could be used by states or for national assessments.

The administration and equipment costs can be low. In New York, each school was asked to provide a test administrator from its teaching staff, and the cost of equipment-- about $100 per building-- is borne by schools. The test administrator is trained to do the scoring. In Connecticut's 1984-85 skills test, the per student cost was $6.60, which included administration, equipment, and scoring. This figure is based on the skills test being conducted as part of an existing state science assessment. Costs for scoring vary with the complexity of the exercises and the kinds of questions that are asked. In the international science assessment and the NAEP study, scoring was done centrally by specially trained teams. With this model, initial costs are high but the costs go down as the scoring process is routinized and applied to more students.

TREND DATA. A question was raised about the usefulness of the alternative methods outlined in the conference for analysis of trends in science learning, since trends analysis requires equating different sets of items from one assessment to the next. There is limited experience with hands-on exercises at national and state levels, and hence no baseline data. The number of designed and tested exercises is relatively small, and studies have not been equated. This issue will not be a problem as more work is pursued with alternative methods of assessment and a pool of exercises is established. States could benefit from

exercises already designed and tested in local science assessments, as is being done in California. When new exercises are administered and baselines established for them at state or national levels, trends can be measured in the science skills. Linking new assessments to earlier ones which lacked these results could be a problem requiring bridge studies.

EVALUATION OF THE CONFERENCE

Each conference participant was asked to respond to a series of evaluation questions. The questions and form were developed by CCSSO to assess the effectiveness of the meeting for the sponsor, the National Science Foundation. The evaluation had two purposes: 1) to assess the effectiveness of the conference in increasing participants' knowledge and information about alternative methods of science assessment, and 2) to determine the status of states' current activities and plans with alternative methods. A copy of the evaluation form is attached.

The first question was whether states use performance assessment in science. From the 34 responses, two states (Connecticut and New York) currently use a performance assessment. Maine uses a paper-and-pencil test with results reported by taxonomies of science content and process. Twenty-two states are planning or considering alternative or innovative methods of science assessment.

The participants were asked to rate the level of their knowledge of alternative methods of science assessment before the conference began. On a scale of 1 to 5 (one being low), twelve respondents reported 1 or 2, eight reported 3, ten reported 4, and four reported 5. Thus, only a small number of state staff felt they had a high level of knowledge about the topic of the conference.

Almost all the state representatives found that the content of the information presented at the conference was useful. On a scale of 1 to 5, thirteen rated the content at 4 and eighteen rated it at 5, while only three rated the content at 3. Comments about the content of information included, "The content was cutting edge," "Provided needed information and resources for our state," and "Valuable approaches, philosophies, and techniques discussed."

When asked to rate the effectiveness of the presentations, the participants' responses were also high. Three participants rated the effectiveness of the presentations 3 or less, while fifteen reported 4, and sixteen reported 5. Comments about the effectiveness of the presentations included, "Best group of presenters I've heard in a long time" and "In the limited time available, a great deal of information was exchanged." Three participants noted that additional time for questions and discussion would have improved the quality of the presentations. One comment was, "Too many presentations with limited time for questions."

The participants rated the overall usefulness of the conference in increasing their knowledge of alternative methods of assessing science as high. One participant rated the overall usefulness as 2, five participants reported 3, eight reported 4, and twenty reported 5. Once again, the participants stated that more time would have improved the quality of the conference, but other participants noted that, "The conference provided a necessary framework to begin our program."

Most states responding to the evaluation reported that the information from the conference would assist their states in the development of alternative methods of assessing science. Participants noted that the resources that were provided, both the printed and professional resources, would be valuable in their efforts at moving states toward incorporating alternative methods of science assessment.

REFERENCES

Blank, Rolf K. and Espenshade, Pamela H. (1988). State Education Indicators in Science and Mathematics: Results of a State-by-State Inventory. Washington, D.C.: Council of Chief State School Officers.

California Assessment Program. (1985). Survey of Academic Skills: Grade 8, Science Rationale and Content. Sacramento: California State Department of of Education.

Connecticut State Department of Education. (1986). Connecticut Assessment of Educational Progress 1984-85: Science, Summary and Interpretations. Hartford: Author.

Jacobson, Willard J. and Doran, Rodney L. (1988). Science Achievement in the United States and Sixteen Countries, Second International Science Study. New York: Teachers College, Columbia University.

National Assessment of Educational Progress. (1987). Learning by Doing: A Manual for Teaching and Assessing Higher-Order Thinking in Mathematics and Science. Princeton, N.J.: Educational Testing Service.

Raizen, Senta A., Baron, Joan B., Champagne, Audrey B., Haertel, Edward, Mullis, Ina V.S., and Oakes, Jeannie. (1989). Assessment in Elementary Science Education. Washington, D.C.: National Center for Improving Science Education.

Council of Chief State School Officers
National Science Foundation

Conference on
ALTERNATIVE METHODS OF ASSESSING SCIENCE

Tampa Hilton Hotel
200 Ashley Drive, Tampa, Florida
January 13, 1988

AGENDA

8:30 a.m.	Objectives for the Conference Rolf Blank/Ramsay Selden, CCSSO
8:45	Perspective of National Science Foundation Wayne Welch
9:00	National/international studies
	Senta Raizen, National Center for Improving Science Education
	Rodney Doran, Second IEA Science Study
	Walter McDonald, NAEP Science Assessment
10:30	Discussion with assessment directors and science specialists
11:15	State models and experience
	Connecticut: Joan Baron, State Assessment Specialist
	New York: Douglas Reynolds, State Science Supervisor
12:15	LUNCH
1:00	State models and experience (continued)
	Michigan: Ed Roeber, State Assessment Director
	California: Zack Taylor, State Science Consultant
2:00	Discussion with assessment directors and science specialists
2:30	Next steps with states
3:00	ADJOURN

EVALUATION

Conference on Alternative Methods of Assessing Science
January 13, 1989

We ask that you take a few minutes to provide some feedback to CCSSO and NSF on today's conference. Please turn in your completed form at the end of the day to Rolf Blank or Ramsay Selden.

STATE:_____

1. Does your state assessment program include any "alternative" or innovative methods of assessment of student learning in science?

   If YES: a) What alternative method(s)?

   b) When did (will) the method(s) begin to be used in your state?

2. Is your state planning or considering any alternative or innovative methods of student assessment in science?

   If YES: What alternative method(s)?

3. How would you rate your level of knowledge of alternative methods of assessment in science prior to this conference, on a scale of 1 to 5?

   Low 1 2 3 4 5 High

4. How would you rate the content of the information you received on alternative methods of assessment in science on a scale of 1 to 5?

   Low 1 2 3 4 5 High

   Comments on Content:

5. How would you rate the effectiveness of the presentations that were made on alternative methods of assessment in science, on a scale of 1 to 5?

   Low 1 2 3 4 5 High

   Comments on Presentations:

6. How would you rate the usefulness of the conference for increasing the knowledge of state education professionals about alternative methods of assessment in science, on a scale of 1 to 5?

   Low 1 2 3 4 5 High

   Comments on Usefulness:

7. Will the information from the conference help your state to plan, consider, or improve alternative methods of assessment in science? If so, how?