

Science Testing: Creative Alternatives

by Edward Chittenden
and Deborah Meier

Conventional achievement testing may not meet the assessment needs of science education and may in fact undercut its major goals.

Last year the New York City school system initiated a city-wide testing program in science, administering tests to all fifth and eighth graders, based on the content of the city's science education curriculum guides. The decision to undertake testing in science is part of a broader national trend, as educators, politicians, and the public turn attention to this neglected area of curriculum.

This trend is welcome, to the extent that it represents a commitment to assessment and improvement of instruction. However, to the extent that it simply indicates the spread of conventional testing practices that have dominated the assessment of reading and mathematics, science educators should view it with considerable skepticism. The conventional model of achievement testing—multiple-choice instruments yielding single score estimates—may not meet the assessment needs of science education programs, and may in fact undercut the major goals of these programs.

While traditional tests yield reliable estimates of students' performance compared to a norm population, they have little diagnostic power and provide meager information about the substance of learning; they were not designed to do so. Moreover, the multiple-

choice format introduces a significant bias in learning assessment, as it is difficult if not impossible for such an instrument to measure productive thinking and the ability to deal with loosely structured problems. Since tests inevitably tend to drive the curriculum, educators should carefully consider the effects of a testing program upon both teachers and pupils before launching a quest for scores.

The city's decision to initiate a testing program prompted us to study alternatives to conventional methods, with the hope of identifying more sensitive measures of pupils' knowledge, and of addressing the full range of instructional objectives.

We examined issues related to the purposes and design of testing, as well as the more specific problem of test format and the need for alternatives to multiple-choice. In considering these questions, we reviewed contemporary assessment literature, interviewed teachers and science coordinators, and field tested open-ended, "constructed-response" items with upper elementary students. The following four recommendations summarize our conclusions.

- Testing in science should yield information about pupils' prior knowledge, interests, and

preconceptions, in order to meet the needs of instructional planning and curriculum evaluation.

Contemporary research has amply demonstrated that children's prior knowledge and experience profoundly shape their understanding of science, and hence do much to determine the effectiveness of science programs. Students' everyday notions about the world can constitute either resources or barriers to their understanding of scientific concepts.

Such research points up the need for testing methods that are sensitive to such expectations and that capitalize upon the diagnostic value of pupils' errors and misconceptions. Multiple-choice formats, requiring recognition rather than construction of answers, inevitably screen out evidence of children's own thinking.

- An assessment program in science should involve teachers in all major steps of the testing process, from content specification to data interpretation.

Conventional testing methods place teachers on the periphery of assessment. While teachers administer tests and receive printouts, they have no role in the more interesting questions of assessment such as test content or interpretation of students' responses. Pre-

In the case of science skills tests, the individual testing model is inappropriate and inefficient.

dictably, as numerous studies have shown, practitioners misinterpret test information when they attempt to make use of it, because they have had no stake in understanding the instrument.

Practitioners should not only contribute to decisions of test content, but also participate in scoring and interpreting complex response data that may be elicited by open-ended testing methods. There is an interesting precedent for this in teacher participation in scoring of writing samples. Such involvement of teachers can promote their knowledge of pupil learning. It also entails a closer examination of the science curriculum and hence contributes to teachers' knowledge of subject matter.

• A testing program in science should be built around sampling strategies of group measurement; scores for individuals (percentile ranks, etc.) are neither desirable nor necessary.

Basic skills tests yield scores for individual pupils. To meet requirements of reliability and internal consistency, such tests are necessarily lengthy and their item content is homogeneous.

In the case of science skills tests, the individual testing model is inappropriate and inefficient. On a practical level there is typically little need for individual data beyond the teacher's own observations and tests. On a theoretical level the idea of individual levels of science attainment (analogous to grade level estimates of reading) makes no sense. The subject matter of science is enormous; the interests and experiences of children are rich and varied. The measurement of individual status relative to a norm group is not useful, even if state-of-the-art testing could meet the challenge.

For such reasons, science testing does not need to adopt the individ-

ual testing model of skills assessment. It can look to item sampling and other techniques of group measurement, as developed for national assessment projects. Studies demonstrate that, relieved of requirement for individual scores, science assessment can provide a far richer array of information.

While all children may participate in testing, they do not need to respond to the same set of items. The scope of inquiry can be expanded; the pool of items or exercises can be large and varied while the number of questions an individual answers can be modest.

• To carry out the recommendations above, assessment in science should capitalize upon computer technology for creating more open and flexible testing methods.

The dominance of the multiple-choice format is directly linked to machine scoring and to technology of testing that developed in mid-century. Despite the obvious limitations of this format for assessing children's knowledge, the practical constraints surrounding test administration have meant that multiple-choice is still the most expedient method.

Computer technology, however,

introduces the possibility that large-scale testing can handle much more complex forms of responses. The data manipulation power of computers makes it possible, for example, to organize short-answer, free-response data in ways that make it amenable to scoring or coding. The flexibility of computers should also support a testing system that provides teachers with ready and continuing access to such information. Computers can organize and reorganize data in ways that are relevant to particular school concerns.

In summary, it is clear that science assessment will require testing practices that are fundamentally different from those associated with basic skills. The development of new methods is timely, given advances in cognitive research and in measurement strategies. With a technology that supports open and flexible systems, it is also eminently feasible.

Edward Chittenden is a research psychologist at Educational Testing Service in Princeton, New Jersey. Deborah Meier is a principal at the New York public elementary school which piloted the alternative testing measures in this article.

Science by the Month

The new 1986 Smithsonian Family Learning Project Science is now available. The Calendar is a giant 16x32 hanging, illustrated throughout with both amusing and instructive drawings. According to the NSTA publication, *Science and Children*, "It offers children the chance to learn that science is not only around them, but it is also inviting, accessible, and fun."

The 12 unusually innovative science activities—one for each month—have clear, easy-to-follow instructions and an "Afterwards" which gives background to the science exercise. However, the most important asset offered in the Calendar is parent involvement—the magic ingredient for successful performance in school.

Order from G.M.G. Publishing, 45 West 43rd St., New York, NY 10036. Include prepayment: \$9.95 plus \$2.50 shipping and handling. For quantity discounts contact Lisa MacDonald at 212-354-8840.