

Inquiry and the *National Science Education Standards*

A Guide for Teaching and Learning

Committee on
Development of an
Addendum to the
National Science
Education Standards
on Scientific Inquiry

Center for Science,
Mathematics, and
Engineering Education

National Research Council

4

Classroom Assessment and Inquiry

The *National Science Education Standards* point out that "assessments provide an operational definition of standards, in that they define in measurable terms what teachers should teach and students should learn" (NRC, 1996, pp. 5-6). In the context of inquiry, assessments therefore need to gauge the progress of students in achieving the three major learning outcomes of inquiry-based science teaching: conceptual understandings in science, abilities to perform scientific inquiry, and understandings about inquiry.

Just as these objectives differ from those of other approaches to science education, so assessments of inquiry-based science education differ from more traditional assessments. Conventional multiple-choice or short-answer questions typically ask students to identify facts, concepts, or vocabulary. Such tests have proven too broad in their coverage, too shallow in the depth of reasoning required, and too narrow when it comes to measuring outcomes like "understanding the nature of science and the work of scientists." These tests are more likely to require recognition and recall rather than in-depth reasoning and application of underlying concepts. As such, they can pose a serious obstacle to inquiry-based science teaching. Teachers are less likely to focus on the goals of inquiry if their students' performance is evaluated on district or state-wide tests that assess isolated facts (Neill and Medina, 1989). Furthermore, when large-scale external examinations take these forms, teachers tend to create similar assessments for their classes (Raizen and Kaser, 1989; Baron, 1990).

Assessment in inquiry-based classrooms takes a broader perspective on the rich learning called for by the *Standards*. It asks what each student knows and understands, what is fuzzy or missing, and what students can do with what they know. Assessment determines whether students can generate or clarify questions, develop possible explanations, design and conduct investigations, and use data as evidence to support or reject their own explanations. At the broadest level, it measures the capacity of students to evaluate the kinds of questions that scientists investigate, understand the purposes of investigations, and assess the qualities of data, explanations, and arguments.

Assessment can take many forms in inquiry-based classrooms, and it serves many

purposes. Assessments can range from the questions teachers ask during a lesson to end-of-unit tests and statewide and national examinations. Assessment data can be used to plan a lesson, guide a student's learning, calculate grades, determine access to special programs, inform policy, allocate resources, or evaluate the quality of a curriculum or instruction. In the breadth of its application, assessment merges seamlessly into considerations of the curriculum and teaching.

An important distinction needs to be made between formative assessment and summative assessment. Formative assessments can occur at any time and are used to influence a teacher's plans to meet specific student learning experiences and needs. Summative assessments typically occur at the end of a learning activity to determine its impact on student learning.

The vignettes in the previous chapter included many examples of formative assessments. For example, Ms. Flores asked her students where they might find worms and how they could build homes for their worms. Mr. Gilbert listened as his students constructed their models of the earth-moon-sun system and asked questions to assess and further their understanding. Similarly, Mr. Hull observed his students' drawings of forces to gauge their understanding. In general, teachers in inquiry-based classrooms are continually assessing to know what to do next, what abilities are developing, which are still underdeveloped, and whether the objectives of a particular lesson or unit are being achieved.

Formative assessments are important for general planning and guidance, but they generally are too informal and insufficiently documented to answer many of the hard questions posed by parents, principals, and teachers: What have students actually learned? What evidence demonstrates that they are learning? How well are they learning it, and at what level of competence?

Formative assessments also are not sufficient to support high-stakes decisions about an individual or changes in policy or professional development designs. Such decisions require summative assessments that provide evidence to parents, teachers, and policy-makers that a student or class is progressing toward meeting the standards for inquiry or falling behind. Such assessments require more standardized instruments and a way of recording student responses, whether a test, interview protocol, or observation guide for a performance assessment. Stable, quantifiable ways of converting student responses to numbers and averages can better support accountability decisions.

The results of summative assessments of student learning can take many forms, from descriptions of individual achievement to formal comparisons across time or with other students. For example, Mr. Gilbert assigned a take-home exam at the end of his session on phases of the moon in which he asked students to summarize all of their evidence that supported or refuted their understanding of the phases of the moon. Ms. Idoni assigned as a final assessment a report describing how each student would investigate an unexpected phenomenon in the lake they had studied. In general, the results of such assessments need to be presented in such a way that they can be summarized and compared with other evidence so that judgments can be made.

This chapter describes features of classroom assessments that support inquiry and the *National Science Education Standards*. It first discusses the "what" -- what are students supposed to know, understand, and be able to do as a result of their education in science. It then discusses "who" should be responsible for various aspects of assessment activities, with a particular focus on students. Finally, it looks at "how"-- the formats and procedures of assessment.

WHAT SHOULD BE ASSESSED?

The three learning outcomes of inquiry-based education involve both knowledge and understanding. The *Standards* define these two terms as follows:

Scientific knowledge refers to facts, concepts, principles, laws, theories, and models. . . . Understanding science requires that an individual integrate a complex structure of many types of knowledge, including the ideas of science, relationships between ideas, reasons for these relationships, ways to use the ideas to explain and predict other natural phenomena, and ways to apply them to many events. Understanding encompasses the ability to use knowledge, and it entails the ability to distinguish between what is and what is not a scientific idea (NRC, 1996, p. 23).



Although understanding has a higher status in science education than knowledge, it is a mistake to think that all instruction or assessments should aim for the higher level of outcome. Indeed, when students fail at complex tasks, one never knows whether they are lacking specific skills or the knowledge needed for success unless one also has examined these requisites. For example, at the beginning of their units on the phases of the moon and static forces, Mr. Gilbert and Mr. Hull probed their students' knowledge of the phenomena being investigated to establish a foundation on which to build more complex ideas.

Some of the abilities of inquiry can be assessed in a relatively straightforward way. For example, teachers can observe and listen to students to determine whether they can "use data to construct a reasonable explanation" (as specified in the K-4 standard), "develop descriptions, explanations, and models using evidence" (5-8), and "formulate and revise scientific explanations and models using logic and evidence" (9-12).

Other inquiry abilities, such as designing and conducting a scientific investigation, are more complex assessment challenges. Champagne, Kouba, and Hurley (in press) have proposed that teachers assess student inquiry by examining four phases of student investigations: precursor, planning, implementation, and closure/extension. For each phase, the teacher should delineate the expected products, abilities, and information. For example, in the planning phase the products include the plan, its rationale, and critiques of peers' plans; abilities include developing a plan, explaining it, and revising it; and the information includes descriptions of characteristics of investigations whose methods are well matched to the question under investigation.

DeJong and Van Joolingen (1998) have summarized a parallel body of research done on inquiry abilities and understandings. Students often are unfamiliar with what a hypothesis should look like (i.e., variables and the relationships between them), are not able to state or adapt hypotheses on the basis of data gathered, and avoid hypotheses that have a high chance of being rejected. In designing experiments, they tend to seek information that confirms a hypothesis, change too many variables at one time, or manipulate variables irrelevant to the hypothesis. Frequent problems in the interpretation of data include confirming the hypothesis

regardless of what the data indicate and difficulty in interpreting graphs (Roberts et al., 1997). Teachers benefit from assessing their students' initial ideas about what it means to conduct an investigation and think scientifically and how these ideas and their skills change over time.

It is easy to say that students should not simply learn isolated facts or definitions without understanding. It is harder to say what the understanding of a concept looks like or how students should produce evidence of their understanding. In the New Standards Project, in which several states and urban districts are working together to develop an assessment system based on the *Standards*, conceptual understanding is described as follows:

The student demonstrates conceptual understanding by using a concept accurately to explain observations and make predictions and by representing a concept in multiple ways (through words, diagrams, graphs, or charts, as appropriate). Both aspects of understanding -- explaining and representing -- are required to meet this standard (New Standards, 1997, p. 133).

Similarly, the AAAS Assessment Blueprint (AAAS, 1998) suggests posing questions that stress reflective thinking, requiring the integration of information, rather than reflexive thinking, where a memorized response is called for. As the Blueprint puts it, "Students should be asked to address questions such as, 'How do we know this?' and 'What difference does it make?' rather than being asked to reproduce memorized vocabulary items or the like."

Again, many of these strategies were apparent in the vignettes in [Chapter 3](#). For example, when Mrs. Flores wanted to assess her students' understanding of the idea of a fair test, she had them evaluate whether a design they had not previously encountered was fair. Ms. Flores also gave her students rich and open tasks such as designing soda bottle homes for their worms based on their observations of the places where they found worms naturally.

However, many of the assessments in [Chapter 3](#) guided the actual day-to-day evolution of lessons, making those assessments susceptible to general judgments and off-the-top evaluations of competence. For assessments that carry stakes, whether of passing courses or assigning grades, "standardized" ways of evaluating knowledge and abilities are needed, preferably ways that can be systematically and reliably reduced to quantitative form. Knowledge and understanding also need to be probed in multiple ways, thus ensuring that a memorized definition does not mask misinformation or misunderstanding.

WHO SHOULD DO THE ASSESSING?

Assessments originate from different parts of the educational system, including administrators and teachers. But a particularly important form of assessment is students' self-assessment. Engaging students in assessment of their own thinking and performance allows them to be more self-directive in planning, pursuing, monitoring, and correcting the course of their own learning. Self-assessment nurtures discovery, teamwork, communication, and conceptual connections.

In a review of more than 580 articles on formative assessment, Black and Wiliam (1998a) point out that "students should be trained in self-assessment, so that they can understand the main purposes of their learning and thus what they need to achieve." Black and Wiliam also

found that improved formative assessment -- including self-assessment -- was most effective in raising the performance of students at the low end of the performance scale, although students who perform well also benefit from better formative assessment. This approach to assessment therefore narrows the gap in performance between the highest and lowest achievers.

Involving students in assessment both reduces the burden on teachers and lets students know what's expected of them. Unless students can see the criteria by which they will be judged and examples of successful performance, assessment becomes a game of guessing what's in the teacher's head. Students frequently fail to make explicit the connection between what they have just done and the question or problem posed. In this respect, it is not surprising that lower-achieving students benefit the most from learning the criteria for success and being shown examples of how to achieve these criteria.

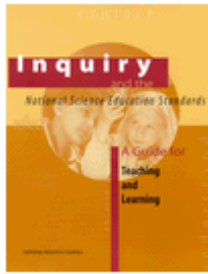
One way of involving students in assessment is to engage them in devising the scoring guide for a task or project. Their first person statements, "I explain my ideas clearly and in detail," and "I used words, numbers, drawings, tables, diagrams, or graphs to show my ideas," are the students' translations of the performance standards for inquiry abilities. Giving students the rubric before they start does not mean giving them the "correct answer" to their investigation. Rather, it is giving them the criteria by which the quality of their conclusions will be judged.

An example of such criteria can be seen in the [Chapter 3](#) vignettes in the journals Ms. Idoni has her students keep as they conduct their field work. By having her students organize their journals according to the inquiry abilities described in the *Standards*, Ms. Idoni provides them with a way of monitoring their own progress in achieving the standards. The conceptual organization of the journals also provides a framework that students can use in their final project at the end of the course.

HOW SHOULD STUDENT LEARNING BE ASSESSED?

Educators long have known that an effective teacher learns a great deal about what students know and do not know, and how they think about scientific ideas, simply by listening to them. A number of years ago, Rowe (1974) identified the very effective instructional strategy of "wait time," where teachers' silence allows students to pose and answer more thoughtful questions than they do when teachers quickly break a silence. She suggested thinking in terms of questions that individual students bring with them -- for example, questions of values (e.g., "Who cares?"), ways of knowing (e.g., "What is the evidence?"), actions (e.g., "What must I do with what I know?"), and consequences (e.g., "Do I know what would happen?"). In writing about assessment, she noted that,





BUY THIS BOOK!

"Learning to have conversations instead of inquisitions is a very powerful way of starting to get data into context" (Rowe, 1991, p. 91).

Gallas (1995) also emphasizes the value of listening to students; she reports gathering her elementary students for open-ended discussions around a particular topic or question that she calls "Science Talks." She allows her students to explore their own ideas, which may or may not be related to the experiences she has planned for them. "Children know when we are 'taking over' their agenda. They can sense when the 'I wonder' in their questions is absorbed into a teacher's 'let's find out and show' agenda" (Gallas, 1995, p. 71). She always asks students to draw, right after the talk, an idea or ideas that they felt answered the question best, which she uses to follow and document the progress of their learning.

Several important dimensions of the familiar formats of multiple choice, constructed response, projects, and portfolios, are displayed in [Table 4-1](#). The challenge for teachers increases from the left side of the table to the right, as the products of assessment go from being right or wrong to having qualities that must be negotiated with other members of the school community. In other words, what are the teacher's, the school's, or the science community's criteria for an excellent response to a particular question?

Table 4-1. Assessment Formats and Procedures

	On demand	----- -->		Over time
Formats	multiple choice, true/false, matching	constructed response, essays	investigations, research reports, projects	portfolios, journals, lab notebooks
Amount of time	typically ~1 min 2-3 min with justifications	1-2 min short answers 5-15 min open-ended responses	days, weeks, or months	months or even years
Whose questions? (audience for the answer)	anonymous or the teacher's	anonymous or the teacher's	the teacher's or the student's	the teacher's or the student's
What kind of questions?	posed narrowly	posed narrowly	posed more openly	varies
Source of answer	anonymous or the teacher's	the student's	the student's	the student's

What kind of answers?	right/wrong	extent of correctness	standards or criteria for quality	standards or criteria for quality
Resources available during assessment	usually none	none or some equipment	equipment, references	equipment, references
Opportunity for feedback, revision	none	usually none	usually some from teachers and peers	usually some from teachers and peers

Discussions among teachers at a school or district level, calibrated with the participation of outsiders, are a component of most effective assessment systems. As Daro (1996, p. 260) puts it:

If standards are to have any real consequence, it will have to be through the engagement of teachers in a professional community holding each other to a mutually accountable standard. They can only hold each other to standards they understand in terms of their own students' work. Thus, deliberating upon their students' work with their colleagues in open but moderated scoring discussions will be needed to make standards a reality for teachers and thereby for students.

In choosing the appropriate format for an assessment, the nature of the standard needs to be examined. Is it something that can and should be assessed "on demand," with little time for reflection or revision? Multiple choice and short-answer responses are convenient for assessing the things that students should know "at the drop of a hat" or "cold." Many of the things valued in the *Standards*, however, require at least the time for reflection (more than a couple of minutes). Consequently, many assessments require formats that take more time.

The vignettes in [Chapter 3](#) emphasize assessments on the right side of [Table 4-1](#), in part to demonstrate the varied uses of assessments. But the full range of assessment formats and procedures could be used in any of the lessons described in [Chapter 3](#). In particular, a combination of evaluative tools likely would be needed to conduct the summative assessments of how much each student had learned from the lessons.

Sometimes teachers, like commercial publishers and district officials, rely on multiple choice formats because they are easy to score accurately, or because teachers are encouraged to prepare students for state or district tests that are in that format. However, it can be difficult to assess understandings, inquiry abilities, or inquiry understandings using just a multiple choice format. One way to make multiple choice questions more meaningful is to ask students to justify their selections, both by saying why they think their choice is best and why the others are not satisfactory.

An additional consideration involves students with limited proficiency in the language of

the assessment. Students who are still acquiring basic knowledge of English vocabulary, syntax, and semantics can have problems both understanding and responding to language-based assessment items. It therefore is important to distinguish between what students know in a subject area and how well they can interpret and respond to specific questions.

The State Collaborative on Assessment and Student Standards under the Council of Chief State School Officers (1999) has developed procedures and materials designed to produce more appropriate assessment of English language learning students. These materials point out that assessments can be affected by linguistic issues (such as the omission of certain letters or sounds that are unknown in a native language), cultural influences (different ways of interpreting a question), and the degree of familiarity with English (whether at a social or academic level). Certain patterns of difficulties emerge among students who are learning English, and a knowledge of these patterns can help make assessments more accurate.

The most comprehensive assessment systems include a variety of instruments. For example, the system developed by the New Standards Project has three interrelated components: performance standards, examinations, and portfolios (New Standards, 1997). The performance standards translate the *National Science Education Standards* into statements that indicate the kinds of activities through which students could demonstrate competence in a standard. These standards also include examples of student work with commentary that explains what aspects of the work illustrate the standard and why it is appropriate for that grade level. The examinations use a combination of selected and constructed response items, including hands-on performance tasks, to yield scores in (1) conceptual understanding, (2) scientific thinking: design and acquisition of knowledge, (3) scientific thinking: analysis and evidence, and (4) life, earth, and physical sciences. The portfolio system includes exhibits for conceptual understanding, scientific thinking, tools and communication, and investigation. Having different exhibits highlights the different types of evidence that need to be presented for these qualitatively different types of standards.

© 2000
National Academy of Sciences

The expectation for quality in the portfolio is higher than the expectation on the exam, as adequate time, feedback, and opportunity for revision are in place for the former. Some of the performance standards, such as working productively in a group, can best be assessed by teacher observation, so certification forms for such expectations are included in the portfolio. Successful implementation depends on the development of a cadre of teachers who are experienced in scoring against a standards-based rubric and on an abundance of examples of standards-setting work from a diverse range of students.

A similar system of multiple formats has been employed in California for several years by the Golden State Exam in High School Biology, Chemistry, and Integrated Science. The examination includes multiple choice items, constructed response items, and laboratory performance tasks. The portfolio is optional and thus is used only to improve a student's score, not to lower it.

CONCLUSION

This chapter has demonstrated that assessment for inquiry-based science education differs

from more familiar, traditional assessments for a number of reasons: the nature of inquiry, the goals of inquiry-based instruction, the alignment of inquiry with the *Standards*, and the capacity of a particular assessment to measure actual progress toward the *Standards*.

These differences in assessment extend both to formative assessments done to guide learning and to summative assessments designed to measure learning, including large-scale (district-wide, state, national, or international) assessments. Summative assessments also must meet a number of additional criteria: they should be systematic, replicable, reliable, equitable for all students, comparable across classes and schools, and interpretable. By meeting these criteria, summative assessments can provide evidence needed to make fair high-stakes decisions -- whether about an individual student's grades or a system's need to redesign professional development approaches for its teachers.

[Previous](#)[Table of Contents](#)[Next](#)