

Assessment of Science Inquiry **by George E. Hein and Sabra Lee**

Different types of classroom assessment can give teachers different kinds of evaluation information. This chapter surveys the assessment methods available to teachers, and talks about the challenges inherent in evaluating learning in the inquiry classroom.

All teachers assess what their students know, where they need help, and what they should do next. Teachers do this informally countless times each day, and more formally after completing a topic, or at a fixed time, such as at the end of a marking period or semester, or the end of a unit.

On a larger scale, administrators and policymakers use assessments to determine how well their schools are educating the next generation. Assessment is a more modern and more inclusive term than the traditional “testing.” It provides the connection between teaching and learning; it lets us know the result of any educational activity. Until recent years, assessment of science education was not a major concern in K–12 education because very little science was taught, especially in grades K–8. With increased attention to science, and recognition that science instruction is important in preparing students for the modern world, science inquiry and the assessment of science inquiry are now seen as crucial in schools.

Assessing Science Inquiry

It is generally agreed that inquiry science includes some hands-on interaction with the natural world; that is, “problem solving,” “investigations,” or “inquiries” must involve actively doing as well as thinking

and reasoning. But this still leaves room for considerable variation in definitions of inquiry science. In some classrooms, children are given carefully prescribed materials and asked to use them in specific ways—they carry out activities that illustrate known scientific principles. For example, they may all be asked to measure a pendulum's period (the time it takes for one complete swing) as the length of the pendulum is changed. In other inquiry classrooms, children carry out independent investigations, exploring questions for which no one knows the answers. They may be asked to find the acidity of water in a local pond, for instance, and then figure out how that affects nearby plant and animal growth.

In each of these classrooms, the records children keep of their work, as well as other assessments developed by the teacher, can form the basis for determining what children have learned. In the first classroom, the teacher can tell whether the children's data conform to the expected Newtonian results for pendulums. In the second classroom, since the acidity of the local pond may, indeed, be unknown, any result may be correct—or incorrect—and the teacher has to look at assessments that demonstrate the methods children used, rather than the results they obtain. In most science inquiry classrooms, some combination of activities and assessments is appropriate.

In order to develop any assessment, the most important issue to resolve is determining what is going to be assessed. In addition, any discussion of assessment of inquiry must start with a clear statement of how inquiry is defined. As the previous sections of this book have demonstrated, definitions of inquiry vary widely.

Assessing “Doing” Science

If we accept the notion that inquiry science involves investigations of the natural world, then such inquiry requires both physical and mental activity. To assess both aspects of inquiry requires “performance assessments.” Such assessments are likely to include a number of components. First, they should address how well students are able to carry out physical processes, such as measurement, observation, experimental design, problem solving, etc. The level of students' thinking and reasoning skills should also be addressed—that is, whether students draw valid conclusions, choose appropriate methods, recognize regularities in nature, and so on. In addition, it's also important to look at students' knowledge of science concepts, and science content.

Uses of Assessment

Assessment can be used for a variety of purposes. Each presents its own opportunities and challenges. The six most common are as follows:

- Diagnostic assessment (pretests) to help determine what students know when they begin any educational task.
- Formative assessment to help guide day-to-day classroom activities.
- Student outcome or summative assessment to find out what students have learned and mastered in their individual programs.
- Comparative assessment for determining how an individual's or a group's outcome compares to some other group's outcome.
- Assessment to support professional development by using analysis of student work to improve the teacher's performance.
- Student assessment to help determine the effect of a program, curriculum innovation, pedagogic strategy, professional development, or policy initiative.

Let's take a closer look at each form of assessment.

Diagnostic Assessment

Diagnostic assessment is used to determine what knowledge and understanding a student brings to a subject. If teachers were content to have all students doing the same thing—listening to a lecture, for instance, solving problems on a worksheet, or making identical measurements—then diagnostic assessment would be relatively easy. But if teachers want to find out what individual students can do, and how each deals with inquiry, then teachers have to engage their students in inquiry processes. Experienced teachers can use classroom discussions, informal observations of children, examination of children's work products, and short interviews to decide what students can do and what they might be ready for next. Most important for diagnostic assessment is that teachers be clear about what they expect to do in their science teaching and know what qualities they hope to bring out in their students.

Formative Assessment

Assessment used to support day-to-day instruction, called formative assessment, makes use of all the normal activities of a classroom. What turns any instructional activity into an assessment is the explicit intention

of a teacher to use it for that purpose, the systematic recording of student results, and the application of some criteria for judging the quality of a child's performance. Many recent NSF-supported science curricula include "embedded assessments," specific activities that can be used to assess students' progress. Thus, students may be asked several times during a unit to draw pictures of a complete circuit, place pictures of plant growth and development in chronological order, draw graphs, or provide a complete description of a scientific term such as "biosystem." Such student products can inform teachers of what ideas have been understood by individual children and what needs to be done next.

Summative Assessment

Traditionally, summative assessment consists of tests at the end of a period of instruction. The term needs to be expanded to include any judgment based on all available evidence of what a student has learned after working on a particular topic.

The most powerful evidence of student growth is provided when teachers combine data from pretests (student work done *before* the topic is studied), embedded assessments (classroom activities recorded *while* a topic is being studied), and post tests (drawings, descriptions, or answers to questions done *after* a topic has been studied). Together, this information provides a summative assessment. For example, if a student does a drawing of a plant, diagrams a functioning motor, gives a specific description of an environment, or carefully draws and correctly labels a graph at the end of a unit, that information can provide powerful evidence of growth in learning, especially when compared to work done just before studying the unit. This form of evidence is particularly valuable in classrooms where traditional paper-and-pencil activities are minimal and time is spent in doing and talking. It often furnishes compelling evidence of student achievement for parents, as well.

Comparative Assessment

Much of the discussion above has stressed individual growth. When assessment is used to compare students with others in a larger arena, however, problems associated with assessing inquiry become more complex. In order to compare students to each other, standards need to be established about what would serve as an appropriate measure of achievement. What is an acceptable experiment for a second grader? How detailed

should a fourth grader's plant drawing be? How many variables can a sixth grader be expected to consider in designing an inquiry?

At this level, problems of sampling also come to the fore. Since any one test can ask only a limited number of questions, the results may not accurately reflect what a particular student knows or can do. But a teacher has available a more complete, if informal, knowledge of the student's abilities and skills. Assessment results that are strikingly different from what a student usually does can be modified by including additional information, reassessing, clarifying what is expected, or providing specific instruction.

When tests are used to compare students against district performance or national standards, the tests may not match what actually was taught in individual classrooms. Since the range of what is learned in inquiry science is so large, it is particularly difficult to develop assessments that cover what individual teachers may be doing in their classrooms. In addition, questions about equity—the background children bring to science and the role of inquiry science in various cultures, inside and outside of school—need to be taken into account (Goodwin, 1997).

Teachers who have participated in study groups that look carefully at children's work, or who are engaged in developing performance assessments, frequently comment about how much they have learned from the process and that it has dramatically and immediately influenced their practice as teachers.

Assessment for Professional Development

Engaging teachers in the process of developing performance assessments or interpreting students' responses to them is a powerful form of professional development. Teachers who have participated in study groups that look carefully at children's work, or who are engaged in developing performance assessments, frequently comment about how much they have learned from the process and that it has dramatically and immediately influenced their practice as teachers.

Student Assessment as a Measure of Program Effectiveness

Higher student achievement should be the central goal of all science education activity. Using student assessment for teacher or program evaluation can be problematic.

When teacher professional development is related to student assessment, it assumes that there is a direct relationship between teacher education and student success (Hein, 1996). However, even when professional development is excellent, there may be many other factors affecting student performance. Changes in local administration, for example, may be a primary influence on student test results. Better teaching may not outweigh other factors, such as increased poverty, administrative turnover, shifts in curriculum priorities, or natural disasters that close schools, any one of which can negatively influence assessment results.

Similarly, student assessment used to measure the effectiveness of district programs assumes that the assessments being used are aligned with the programs being implemented. Many current large-scale assessments only require that students respond to prompts that include all of the required information (Madaus et al., 1992). One major change in making assessments more appropriate for inquiry science is to include questions that require that students “supply” information, such as explanations, long answers, drawings, and all performance tests, in contrast to traditional multiple-choice or true-or-false test questions for which students “select” correct answers (Madaus, Raczek, and Clarke, 1997). Forms of assessment that require students to supply information can, at least in principle, assess complex chains of ideas and skills, as well as recall of specific knowledge. Questions that only require the supply of information usually assess specific knowledge in small, discreet units. But although most reform efforts require students to use materials, as well as to think and reason about the natural world, performance assessment is still a minor part of most large-scale testing and is not included in many state efforts.

Assessment Challenges

Because inquiry science places a number of demands on assessment processes, and because there are limited resources available to deal with these demands, there are many challenges to creating satisfactory systems for assessing inquiry science, and especially to modifying existing

practices. Usually, however, a reasonable middle ground can be found between conflicting tensions, as described below.

1. Large-scale assessment requires significant standardization, while individual student inquiry must involve some novelty. The practical problems of creating assessments that are appropriate for all students, and yet allow each to fully demonstrate what he or she knows, can be solved given sufficient resources. This is demonstrated by the acceptance of performance assessment in similar situations, such as in group sports, in the performing arts, or in various practical tasks—for example, when a swimmer demonstrates life-saving skills. In these fields, applicants are usually required to complete a standardized task but are allowed some individual variation within a permissible range.
2. Assessing inquiry science requires that teachers document their students' physical skills, such as the ability to observe, measure, and design experiments. Yet, although many science standards refer to the skills that a student needs to do inquiry, there is little empirical evidence about how these skills develop with age. Careful observation is an important skill in science. But, for example, how should we expect a first grader's observations of animal locomotion to differ from a sixth grader's descriptions of the same phenomena? What is a competent measurement for a 6-year-old, and how does this differ from what can be expected from a 9-year-old? We need more research describing the physical and mental capabilities of children of different ages before valid, age-appropriate assessments of science skills can be implemented.
3. Assessment and administrative monitoring usually involves a single test, given at a specific time, and with some ceremony. This process provides a "snapshot" of what a student can do at that time and under those circumstances. Academic test performance, with its attendant test anxiety, may not be the most appropriate measure of student achievement, since society is usually interested in how pupils will perform under "normal" circumstances instead. Some balance between summative judgments made from the accumulation of continuous records, in contrast to judgments from more stressful testing situations, needs to be reached.

4. “Teaching to the test” has a negative connotation among many educators. But when assessment tasks closely mirror what qualified students should be able to do in a particular domain, then instruction and curriculum are closely aligned and teaching to the test is appropriate. In the assessment of inquiry, it is often considered a good idea for the teacher to share criteria for assessment with students, making the whole process open and transparent. To what extent does instruction encourage students to practice what will be assessed? Teachers who have shared assessment criteria with students, or involved students in developing assessment criteria, often report not only increased interest from students, but also improvement in their work.

Conclusion

Assessing inquiry science at the national level is still in its infancy, but over time, teachers have developed a large body of practical experience that can form the basis for good classroom assessments. While school reform efforts are improving education for all children, continuing attention to assessment will help us better understand what children have or have not mastered during their education. As more schools implement inquiry science, we will build a firmer experience base of what it means to do science in classrooms, contributing to the national effort to develop valid, appropriate tests. A growing body of methods is available to assess inquiry science, primarily based on performance assessments. Classroom teachers can develop ways to understand what their students know and can do, and they can utilize this growing body of materials to document student growth.

References

- Goodwin, A.L. (ed.) (1997). *Assessment for equity and inclusion: Embracing all our children*. London: Routledge.
- Hein, G.E. (1996). The logic of program evaluation: What should we evaluate in teacher enhancement projects? In S.N. Friel, and G.W. Bright, *Reflecting on our work: NSF teacher enhancement in K-6 mathematics*. Lanham, MD: University Press of America, Inc.

Madaus, G.A., Raczek, A.E., and Clarke, M.M. (1997). The historical and policy foundations of the assessment movement. In A.L. Goodwin. ed. *Assessment for equity and inclusion: Embracing all our children*. London: Routledge.

Madaus, G., et al. (1992). *The influence of testing on teaching math and science in grades 4-12*. Chestnut Hill, MA: Boston College Center for the Study of Testing, Evaluation, and Educational Policy.