Scoring Active Assessments

Setting clear criteria and adapting them to your students are the keys to scoring classroom performance.

By Sabra Price and George E. Hein

Increasingly, educators are examining collections of children’s drawings, writing, recorded conversations, and other work to find out what their students have learned in science class. Yet, classroom teachers may hesitate to base their assessment on this material because they don’t know how to score it.

What’s the Score?
Scoring, the placing of relative value on a piece of work, must be part of any assessment system. In most schools, teachers assessing children’s work must use letter grades. In schools where grades are not an important issue, teachers still may need to assess students if they are to report to students’ parents, justify the science program’s existence, or meet accountability standards. Scoring is usually essential if you must answer such questions as “What is this child learning?” “What evidence is there of progress?” and, “Where is this child in comparison to others who are the same age?”

Teachers should develop expected levels of achievement for science by observing what their students actually do in the classroom.

Unfortunately, most current framework statements and new curriculum materials are not specific enough to be used as guides to scoring children’s work. For example, in Benchmarks for Science Literacy, the American Association for the Advancement of Science states that children in grades three through five “should be encouraged to observe more and more carefully, measure things with increasing accuracy, . . . record data clearly in logs and journals, and communicate their results in charts and simple graphs as well as in prose” (1993, p. 10).
To apply these recommendations for assessment, you need to develop specific descriptions of what children are expected to do. To establish scor-
Before establishing a scoring system, decide what behaviors, skills, and products you plan to observe.

Defining criteria

You must first decide what kinds of learning you value, and what behaviors, skills, and products you plan to observe.

Define Your Criteria

You can apply the same criteria you would use for interpreting children’s work to scoring it for assessment. For example, in looking at a child’s transplanted seedling, the criteria might include depth of planting or whether the seedling is still alive; in looking at a drawing of a mealworm, the criteria might be inclusion of particular body parts; and in judging a student’s graph of an experiment, the criteria might be whether the axes have been labeled correctly.

Define your criteria based on what you believe students should get out of the lesson or unit. The more specific your criteria are, the easier it will be for you to score your students’ work. Criteria can include science content, processes, skills, and attitudes. Your students can benefit both from participating in setting criteria and from scoring themselves.

Ranking Achievement

For each of your criteria, define, describe, and rank the range of behaviors you can expect from your students. Remember, there is no universally accepted, absolute standard for judging the science activities of children. We simply don’t have enough information or research evidence to determine what is an “appropriate” or “outstanding” observation for a six-year-old, or what level of experimental design can be expected from a 10-year-old.

With this in mind, develop your levels based on what you observe students actually doing in the classroom. Decide what would demonstrate a student’s full knowledge of a concept or complete mastery of a skill. When setting a standard, you must take into account the classroom opportunities your students have had to learn the assessed skills, concepts, and attitudes. Your definition of levels for each criterion may be unique to your situation. Figure 1 (see page 29) shows a generic scheme that can be applied broadly to all scoring systems. Remember, too many levels is confusing and makes the task of scoring unnecessarily complex. Most schemes consist of three or four levels.

Refine Your Scoring

Any scoring system needs to be tried out and refined in the classroom. You may find that you need to revise your levels of achievement, or that you need to add a drawing or an activity to your active assessment in order to document and score student learning.
more accurately.

Discuss your scoring criteria with colleagues, showing them examples of your students' work. Many teachers have discovered that such conferences give them insight into how they teach and how children learn.

Look to Your Classroom

Most assessment sections in new science curriculum materials do not provide scoring systems; publishers and developers have left this task to the individual teachers or districts. Even when scoring systems are provided, teachers may need to modify them. For example, in the Full Option Science System (FOSS) publication Variables, for fifth and sixth grades, one question in the final assessment provides students with a picture of a stone-throwing catapult and asks them to list all the things that might change the stone's trajectory. The teacher's manual suggests that, in addition to the position of the catapult, the angle of release, and the mass of the stone, "any reasonable answer" should be accepted. This requires the teacher to interpret the children's responses based on their experience with their own catapults. Teachers might consider, for instance, how much time was spent on the activity, what the class discussion covered, and whether some answers should count more than others.

In another example, from the IN-SIGHTS unit Growing Things (Education Development Center, 1990), students observe pairs of growing plants and answer questions about them. The unit does not provide a scoring system, so we developed one (see Figure 2). Note that, even though our scoring system is detailed, it would still need to be adapted to each teacher's situation.

Choosing a System

If you need to assign grades, make comparisons among students, or compare individual students to a general standard, it may be necessary to assign numerical values to individual rankings.

The simplest scoring approach is to treat each criterion separately, assign rankings to each level, and then to add and average the scores. Unfortunately, this method may produce final scores that hide most of the interesting information. For example, if one of the criteria referred to knowledge and another to skills, then the final score would not tell us anything about the student's relative performance on the two components. Average scores do not differentiate among growth in inquiry skills, attitudes, or conceptual understanding. Bearing this in mind, you may wish to keep the scores of different components separate.

It is also possible to develop a holistic scoring system in which the scorer applies general criteria to whole documents, such as notebooks, laboratory reports, or student products. Holistic scoring is relatively simple and can be carried out quickly. It also leaves considerable room for scorer judgment. If you know the student, then it is possible to incorporate into the grade such factors as a comparison between actual and expected performance. You could also consider such questions as:

- Is the performance typical of the student?
- Was the student rushed in completing the work?
- And, Did the student miss a group activity that would have contributed to the product?

If you choose to use a holistic scoring system, it is important that you have a set of clear criteria and are able to document how you applied these criteria to the students' work. Certain criteria may not be clear until you have begun to look at a set of notebooks, drawings, or graphs. In some cases, you will need to examine a number of student papers before you can determine whether labeling a particular component or drawing a certain phase is a sign that students have met a given criteria.

It's Worth the Effort

Scoring systems can be developed for active assessments. It does take time and effort, and it does require teachers and administrators to discuss what they value in education, but a scoring system for active assessments lets students know what is expected of them, helps teachers in their instruction, and helps the public understand a district's elementary science curriculum goals.

Resources


Figure 1. This general scoring rubric could be applied to assess a variety of science learning situations.

**General Scoring Rubric**

*Poor:* The student did not do the task, did not complete the assignment, or shows no comprehension of the activity.

*Inadequate:* The product does not satisfy a significant number of the criteria, does not accomplish what was asked, contains errors, or is of poor quality.

*Fair:* The product or assessment meets some of the criteria and does not contain gross errors or crucial omissions.

*Good:* The product or assessment completely or substantially meets the criteria.

*Outstanding:* All the criteria are met, and the product or assessment exceeds the assigned task and contains additional, unexpected, or outstanding features.

Figure 2. This sample scoring system for a unit on plants shows a practical application of performance assessment.

**Sample Scoring System**

**Question One.** Students are asked to measure the height of two seedlings and to record their results.

**Scoring Rubric for Question One**

0 = The student either did not record results or reported measurements that were inaccurate by more than a certain percentage determined by the teacher.

1 = The student did not record results, but did report approximate measurements. The teacher needs to determine the meaning of "approximate." This will depend on such things as the marklings on students' rulers and students' classroom experiences.

2 = The student recorded approximate measurements.

3 = The student recorded accurate measurements. The teacher needs to determine the meaning of "accurate."

**Question Two.** Students are asked to explain their recorded measurements to their teacher.

**Scoring Rubric for Question Two**

0 = The student provided either no explanation or one that makes no sense to the teacher or is unrelated to any unit activity.

1 = The student's explanation related to unit activities but did not explain the growth pattern.

2 = The student provided an explanation for the growth pattern.

3 = The student gave more than one reasonable explanation for the growth pattern.

---


SABRA PRICE is a research associate at TERC and a senior research associate at the Program Evaluation and Research Group at Lesley College, both in Cambridge, Massachusetts.

GEORGE E. HEIN is a professor at Lesley College and the director of the Program Evaluation and Research Group. This article is adapted from the chapter entitled "Scoring" in the authors' book *Active Assessment for Active Science.*