

The ABCs of Assessment

*Aligning assessment
with instruction*

Ann W. Wright

A contradiction exists between teachers' instructional goals and their assessment strategies. In a study of high school biology teachers, it was found that, "overall teachers wanted their students to develop a general interest in and understanding of biology as well as its real-world applications" (Bol and Strage, 1996, 145). However, the same study found that more than half of the teachers' assessments evaluated students' basic knowledge and only 4 to 5 percent evaluated students' application knowledge (Bol and Strage, 1996). Therefore, teachers need to improve and find alternative types of assessment.

Unfortunately, assessment often carries a negative connotation. But, it is as important as curriculum and instruction and needs to be viewed as a positive enhancement of learning. To improve their assessment practices, teachers need to familiarize themselves with the many types of assessments and how they are used. For instance, assessment can be summative, diagnostic, or formative, depending on when it takes place during an instructional program. Summative assessments, which teachers use most often, are given at the end of instructional units and determine students' grades. Diagnostic assessment, given before learning takes place, determines student understanding of topics before learning takes place. Diagnostic assessment can be used if students need remedial help. Formative assessments are given during the instructional unit, and the outcomes are used to adjust teaching and learning.

Assessments also can measure the cognitive domain, which is students' intellectual ability; the affective domain, which is students' attitudes, values, interests, and awareness; and the psychomotor domain, which governs students' motor or manipulative skills. The assessments used during a course should be composed of various elements to effectively assess all domains. Depending on one type of assessment, such as a multiple-choice test, to



SCILINKS
THE WORLD IS A CLICK AWAY

Explore assessment strategies at www.scilinks.org.
Enter code TST1060.



PHOTO BY MIKE OLLIVER

determine the learning outcome is not a comprehensive assessment of student learning. To be comprehensive, assessment must be authentic, meaning it resembles the classroom experience or “real life experience.” Simply “having the students perform or do things in a testing situation does not make a test item authentic” (Reynolds et al, 1996, 13). Students must be given the opportunity to learn the content.

Types of assessment

Many types of assessment exist; the wheel has already been invented. Traditional assessments include paper-and-pencil formats such as multiple-choice, short-answer, and essay. Summaries of practical skill assessment by the International Association for the Evaluation of Education Achievement indicate that there is a qualitative difference between performance on specific process skills in practical tests and the same skills on paper-and-pencil tests (Tamir et al, 1992). Some people blame the low scores on process skills on international tests on the predominant use of multiple-choice assessments.

Alternative assessments are often more effective. They can take non-traditional forms, such as performance (hands-on) assessment, portfolio assessment, time-series design (Lin and Lawrence, 1999), and predict- and-explain assessment. In any new endeavor,

teachers should begin small, get feedback from colleagues, and ask their schools or school districts to purchase reference books and supplies.

Performance assessment

Performance assessment should measure content outcomes as well as inquiry skills. Inquiry skills include process skills and the understanding of scientific ideas. Observing, explaining (hypothesizing), predicting, raising questions, planning and conducting investigations, interpreting data, and verbalizing conclusions are process skills, which can be assessed in any order.

Performance assessment can test the affective domain. The affective domain that measures attitudes can include “willingness to collect and use the evidence, to change ideas in the light of evidence, and review procedures critically” (National Research Council, 1999, 89).

As on any assessment, when students have their performance assessed, they need to know the objective and how the outcomes will be used. Performance assessment can be diagnostic, summative, or formative. Performance assessments must be at the appropriate level, have a clearly defined purpose, and have scoring rubrics developed with them.

Figure 1 (page 59) is a sample performance task that could be used in a high school biology class. Even

To be comprehensive, assessment must be authentic, meaning it resembles the classroom experience or “real life experience”

though the example is biological in nature, it can be used as a framework to develop a performance task in physics, chemistry, or mathematics.

To complete the task in Figure 1, students need a basic understanding of enzymatic activity and should know that enzyme activity is dependent on pH and temperature. Digestive enzymes, which can be bought from biological supply companies, can be used to illustrate enzymatic properties. Amylase, pepsin, trypsin, and lipase are digestive enzymes that work best at body temperature—37°C. If the temperature is increased above or decreased below body temperature, enzymatic activity will decrease. Amylase, trypsin, and lipase are most effective at a pH between 7.35 and 7.45. A pH outside this range will decrease enzyme activity.

Amylase breaks down starch. Salivary amylase, found in saliva, can be used in a laboratory activity prior to the assessment. Salivary amylase added to a 10 percent starch solution at pH 7.4 and 37°C will break down the starch into glucose. Benedict’s solution can be used to qualitatively determine the concentration of glucose. Salivary amylase can be added to 10 percent starch solutions with a pH above 7.45 and below 7.35 to determine the effect of pH on salivary amylase activity. Salivary amylase can be added to 10 percent starch solutions at a pH of 7.4 and temperatures above and below 37°C to determine the effect of temperature on salivary amylase activity.

Once students have finished the task, teachers should use the scoring rubric to assess students’ performance. Teachers should take the time to check the scoring rubric for interrater reliability by having a colleague score a number of the students’ papers and then checking how closely the scores match or correlate.

Portfolio assessment

A portfolio is a collection of items produced by a student during a course or over a period of time. Portfolio assessments can be used to evaluate students’ learning during a certain time interval; they are formative assessments that encourage students and teachers to reflect on learning. Portfolio assessments can also be used as self-evaluations and show progress over time.

In the laboratory portion of a class, students should keep journals and write several laboratory reports. A portfolio could contain a laboratory report from the

beginning of the instructional unit and one from the end of the instructional unit. A student should design and plan an experiment around an idea chosen either by a student or the teacher. The experiment could be included in the portfolio. Finally, the portfolio could contain a literature review the student wrote on a topic of interest. With each component of the portfolio, the student needs explicit instruction in how to accomplish each assignment. Often, an example of a grade A assignment is useful.

Time-series design assessment

Although they may be time consuming, time-series design assessments are effective for monitoring student learning and assessing teaching effectiveness (Lin and Lawrence, 1999). A single item is given daily or in frequent time intervals during the instructional unit. An example of a single item might be a question asked during a class on the respiratory system, such as, “Does depth of breathing increase when the rate of breathing increases?” Students can answer the items individually or in groups, and the answers may be oral or written. The results can be used formatively, giving students the opportunity to express how much they have learned. For instance, if a student responds to the previous question by saying that breathing depth does not change or decrease, then the teacher should immediately indicate that this is incorrect so the student knows that this is a false understanding of respiration.

As an example, a time-series design assessment might be used during a unit on genetics, when students learn that a single allele can cause a single characteristic (one allele is dominant and one is recessive). After the lesson, PTC paper can be used to test students’ understanding of this concept. Students should taste the paper and dispose of it. The teacher should ask how many students could taste the paper and then ask the class if the single PTC taste allele is dominant or recessive. The taste allele is dominant. The teacher should pick one student who tasted the paper and ask the class if that student could have only one genotype—a yes-or-no answer. Students should realize that the answer is no because a dominant phenotype has the possibility of two genotypes. A person who can taste the PTC paper could be homozygous or heterozygous. A homozygous person would have two PTC taste alleles, and the heterozygous would have one taste allele and one no-taste allele. Even

FIGURE 1**Example of a performance task.**

Level: High school biology class

Skills assessed: Hypothesizing, investigation, data collection, and interpretation

Content assessed: Enzyme activity

Student information

Purpose: You are working in a lab that tests enzyme activity. Someone has sent in a sample of an unknown enzyme, which you are told is similar to amylase. Your job is to design an experiment to test the enzyme's activity on starch.

1. Write out an experimental design or plan. Be as thorough as possible. Make illustrations if you want. Be sure to label the illustrations.
2. Write a prediction before you begin.
3. Run the experiment.
4. Record your results. Draw a table, chart, or graph of your results.
5. Describe your conclusion.
6. Describe another experiment that should be done to confirm your conclusion about the factors that affect the enzyme's activity.

Scoring rubric

The development of a scoring rubric is as important as the task development.

Design and planning (2 points)

Detailed procedure _____

Logical sequence of steps _____

Prediction (3 points)

Dependent variable _____

Independent variable _____

The relationship of variables _____

Experimental manipulation (points optional)

Observations can be made of students' equipment manipulation skills, such as using a volumetric cylinder appropriately.

Results (3 points)

Chart, table, or graph _____

Axis labeled correctly _____

Data recorded accurately _____

Conclusion (2 points)

Results supported with data _____

Results backed by scientific principles _____

(Enzyme activity depends on pH and temperature)

Extension (1 point)

One other extension proposed _____

Total (11 points)

though the answer is yes or no, students have to extend their knowledge from the lesson to answer the question. This type of formative assessment allows students to evaluate their own learning. The teacher can use this to measure students' understanding of the concept of dominant and recessive alleles.

The next topic presented is gene mutation. Students study sickle cell anemia and learn how an alteration in one nucleotide can alter a protein structure. Because of the structural

change, the function of the hemoglobin is compromised. A yes-or-no question could be, if nuclear radiation changes a single nucleotide in a gene, could the protein the gene codes for have functional changes? The answer is yes. Can this be beneficial for the survival of the species, yes or no? Again, the answer is yes, it could be. Questions can be developed prior to the instruction of the unit. Questions over a period of time provide a measure of students' understanding of a topic.

Predict-and-explain assessment

In predict-and-explain assessment, students are given some information, such as a picture of an optical concept, and are asked to make a prediction. Students are then given more information, such as another picture of the concept, and must modify their prediction (Lawrence and Pallrand, 2000). Predict-and-explain assesses inquiry skills and can be integrated into laboratory activities by having a student predict an outcome of the activity and explain the prediction.

For example, a predict-and-explain assessment topic could be gas exchange in the lungs. Carbon dioxide is removed from the body while oxygen diffuses into the capillaries to be used by the tissues. Carbon dioxide is a product of some metabolic processes; for example, carbon dioxide is produced during the formation of ATP, which is a higher energy molecule. As carbon dioxide levels increase, the pH of the blood decreases.

As a demonstration, the teacher uses 50 mL of 1 percent sodium hydroxide, which contains phenolphthalein. Phenolphthalein, which should not be handled directly by students, is a pH indicator that turns pink if it is basic and clear if it is acidic. Teachers can ask

students to predict what will happen when they blow through a straw into the solution. Then, they should blow into the solution until it becomes clear. Then, students should revise their predictions and explain why their predictions were accurate or inaccurate. If one rigorously exercises, it increases carbon dioxide levels. Students can compare the period of time before the solution becomes clear between the two situations—no exercise and exercise.



To effectively use any assessment methods, teachers must constantly evaluate the success of the methods used

Analyzing assessment

To effectively use any assessment methods, teachers must constantly evaluate the success of the methods used. Developing an assessment always takes time, but often little time is spent analyzing assessment results. This is a critical step that teachers must include. There are several ways to analyze assessment results: item analyzing, reflecting on the results, and checking the reliability of scoring rubrics. If 90 percent of the two scores are the same, the scoring rubric can be considered reliable.

One way to improve assessment is by doing an item analysis. Item analyses should be done on all types of assessments. The data from an item analysis should be used to make modifications in instruction. Two useful values of item analysis are the difficulty number and the discrimination factor (Doran et al, 1998). The difficulty number or difficulty index is a percentage of those test takers who correctly answer the questions. The formula for this equation is $P = R/N \times 100$, where R is the number of students choosing the correct answer, N is the number of the students taking the test, and P is the difficulty number. The criterion used for determining whether or not to accept a question is 50 percent or within a range around 50 percent.

The discrimination factor is another way to analyze an assessment item. The discrimination factor differentiates between the high achieving students and the low achieving students. The formula used to determine the discrimination factor is $D = (H-L)/(N/2)$. The total number of students in the sample is N. The number of students in the high group is H, and the number of students in the low group is L. The discrimination factor is D. An index greater than 0.30 is a widely used criterion. The item must have a positive discrimination factor to be considered useful. These two item analyses can be used for question construction and to help ensure that the assessment is valid and as fair as possible.

Teachers should reflect often on assessment and ask questions such as, "Have students learned what they were supposed to learn?" Or, "What concepts do the students have trouble with, and why are they having trouble?" Reflection can improve assessment and will essentially im-

prove learning. The time for evaluating the assessment after its administration should be equal to the time used for assessment development.

Even though a change in assessment is difficult and time consuming, it is worth a try. Colleagues can help each other review assessments. Administrations should also be interested in new ideas of assessment, especially if the assessments are aligned with the *National Science Education Standards* (National Research Council, 1996) and their state standards. ∞

Ann W. Wright is an assistant biology professor at Canisius College, 2001 Main Street, Buffalo, NY 14139; e-mail: wrighta@canisius.edu.

References

- Bol, L., and A. Strage. 1996. The contradiction between teachers' instructional goals and their assessment practices in high school biology course. *Science Education* 80(2): 145–163.
- Doran, R., F. Chan, and P. Tamir. 1998. *Science Educator's Guide to Assessment*. Arlington, Va.: National Science Teachers Association.
- Lawrence, M., and G. Pallrand. 2000. A case study of the effectiveness of teacher experience in the use of explanation-based assessment in high school physics. *School Science and Mathematics* 100(1): 36–47.
- Lin, H.S., and F. Lawrence. 1999. Using time-series design in the assessment of teaching effectiveness. *Science Education* 83: 409–422.
- National Research Council. 1999. *Foundations: Inquiry: Thoughts, Views, and Strategies for the K–5 Classroom*. Washington, D.C.: National Academy Press.
- National Research Council. 1996. *National Science Education Standards*. Washington, D.C.: National Academy Press.
- Reynolds, D.S., R.L. Doran, F.H. Allers, and S.A. Agruso. 1996. *Alternative Assessment in Science: A Teacher's Guide*. Buffalo, N.Y.: New York State Department of Education State University of New York at Buffalo.
- Tamir, P., R.L. Doran, and Y.O. Chye. 1992. Practical skills testing in science. *The Journal Studies in Education Evaluation* 18(3): 263–275.

Online extension

For more examples of authentic assessments, NSTA members can log on to *The Science Teacher* through the NSTA website at www.nsta.org.