KNOWING *When You* DON'T KNOW

Supporting teaching and learning using a new generation of tests

William E.J. Doane, Rebekah R. Rice, and Paul Zachos

Picture this: *It's the first week of September. You're giving your students a test to find out what they know about density before you begin teaching. Every student is fully engaged for 30 minutes in describing and interpreting the behavior of floating and sinking objects. You watch them struggle with the task but can see the excitement in their faces. They want to understand what they're seeing. You know that their desire to understand means they will be ready to learn the next time they walk into your classroom. That night, you score their tests. The next day, you're confident that your plans for instruction precisely meet their learning needs.* Is this what happens in your classroom? If not, it can be.

This article describes an alternative way of testing—a performance assessment—that is different both in process and goals from typical tests, but closer to what teachers believe assessment should be. Teachers who have helped to create, refine, and use this new generation of tests find them valuable because they measure the skills and concepts that they believe are most important for their students to develop.

A new generation of tests

With very few exceptions, current tests are constructed so that the number of correct items are added up to obtain a total score. Summed scores do not tell us what students can and cannot do, and so they cannot be effectively used to plan instruction. Since tests are typically administered when instruction is already over, they cannot help to plan instruction. Testing does not have to be this way.

How are these new assessments different? The assessments motivate students and focus their attention through the use of stories and discrepant events. These assessments:

- allow students to solve practical problems based on natural phenomena,
- target higher-level thinking abilities,
- produce information that supports teaching and learning, and
- can be used both for formative and summative purposes.

These alternative assessments also share some desirable properties with current tests: The learning goals are aligned with state and national standards, the assessments are reliable, and they are efficient. The assessments, like current tests, are standardized in their administration and scoring so that teachers can have confidence in the objectivity of the results that are obtained. Many believe that the problem with current tests is standardization. But standardization is necessary to ensure validity and reliability, both desirable features of tests. Conventional tests are problematic not because they are standardized, but because they are noneducational (Zachos 2004). Noneducational tests do not serve to inform instruction or improve learning.

An example

Cubes & Liquids is an assessment that grew out of a lab created by Cindy Sargent, a New York State Earth science teacher, who used the lab to teach density of solids and liquids, careful observation skills, and critical thinking. In Cubes & Liquids, students are presented with a discrepant experience. Several cubes of varying size and density are placed in similar looking liquids. The cubes and liquids have been chosen so that "ob-

Correlating standards and benchmarks.

New York State Standards and Task Learning Goals:

An alignment of Cubes & Liquids learning goals with New York State learning standards is available online at *http://acase.org/tasks/cubes-liquids/alignment.pdf.*

AAAS Benchmarks and Task Learning Goals:

Distinguishes observation from inference

"What people expect to observe often affects what they actually do observe" (AAAS 1993, p. 12).

Technical description

"Accurate communication with a science discipline results in part from the use of technical language" (AAAS 1993, p. 295).

Density of solid objects

"Equal volumes of different substances usually have different weights" (AAAS 1993, p. 78).

Density of liquids

"Equal volumes of different substances usually have different weights" (AAAS 1993, p. 78).

Uses a 2x2 classification scheme to organize relevant factors

"Organize information in simple tables and graphs and identify relationships they reveal" (AAAS, p. 297).

Proportional reasoning

"Use ratios and proportions... in appropriate problems" (AAAS 1993, p. 291).

Students, when asked to make predictions about whether a particular cube will float or sink in a particular liquid, need to consider that they may not have sufficient information to make a good prediction. In other words, they must discover, as Socrates did, the value of knowing when they do not know.

vious" explanations quickly become insufficient to explain the observed phenomena. For example, some cubes float in some liquids while sinking in other liquids. Also, some large cubes float while small cubes sink. Confounding factors have been incorporated into the assessment in order to encourage higher-order thinking by students. [**Editor's note:** A complete version of the Cubes & Liquids activity can be downloaded from *http://acase.org/tasks/cubes-liquids*. The site also contains teacher instructions and a demonstration of how student progress on learning goals over time can be maintained for planning and evaluation purposes.]

The Cubes & Liquids assessment begins with a story of how Socrates became a philosopher. (Other stories connected to the task's learning goals could also be used.) Students, when asked to make predictions about whether a particular cube will float or sink in a particular liquid, need to consider that they may not have sufficient information to make a good prediction. In other words, they must discover, as Socrates did, the value of knowing when they do not know.

Building on the story, students are asked to predict which of several objects of different sizes will float (or sink) in beakers of indistinguishable liquids. Custom cubes of varying densities but identical outward appearance are ideal for this task, but teachers have also successfully used different-sized plastic eggs filled with insulating foam and weights to adjust density. Once students have made their predictions, the teacher places the cube in each of the liquids and students are asked to record their observations.

(Cubes & Liquids can be a challenging task for graduating high school seniors and yet equally engages and provides meaningful entry points for sixth and seventh graders and learners of diverse abilities.)

At the end of the assessment, students bubble with curiosity about what happened. Were the liquids (visually indistinguishable) actually different? What were they? What were the cubes made of? Why did such and such happen? Students are primed for learning. The Cubes & Liquids activity is efficient because it does not take time away from instruction. Learning begins with the first administration of the assessment. This is the heart of formative assessment—assessment cannot be formative if it is not followed by instruction.

FIGURE 1

Operational learning goals for Cubes & Liquids.

Distinguishes observation from inference

1. Records observations; makes no unnecessary inferences
0. Makes inferences where only observations are called for

Technical description

2. Correctly indicates all critical actions and objects for every event
1. Indicates all critical actions and objects for at least one event
0. Does NOT indicate the critical actions and objects even once

Density of solid objects—Coordinates mass and volume of solid objects

2. Correctly coordinates mass and volume of solid objects
1. Attempts to coordinate mass and volume of solid objects
0. Does NOT coordinate mass and volume of solid objects

Density of liquids—Coordinates mass and volume of liquid

2. Correctly coordinates mass and volume of liquid
1. Attempts to coordinate mass and volume of liquid
0. Does NOT coordinate mass and volume of liquid

Uses a 2 × 2 classification scheme to organize relevant factors

2. Forms a COMPLETE classification scheme including all levels of both factors
1. Forms an INCOMPLETE classification scheme including all levels of one factor
0. Does NOT form a scheme to classify objects

Proportional reasoning—Coordinating solid and liquid densities

2. Correctly coordinates two ratios
1. Attempts to coordinate ratios
0. Does NOT attempt to coordinate ratios

Learning goals

The concepts and process skills called for by this assessment are aligned with state and national standards. However, state and national standards are sometimes too broad, and consequently too vague for practical use in the classroom. They must be operationalized or translated to a level of specificity appropriate for teaching a lesson or unit (Figure 1).

The six learning goals in Figure 1 are representative of valued capabilities identified in science standards and benchmarks (see "Correlating standards and benchmarks," p. 47). The learning goals are valuable across the

By implicitly encouraging only correct answers, grades undermine our ability to observe what is perhaps the most powerful information available for instructional planning: genuine examples of students' misconceptions.

curriculum, are foundational for more advanced learning, and are applicable to authentic problems. Yet they remain some of the most difficult learning goals for which to demonstrate student attainment!

This assessment requires that the learning goals be applied in the context of solving a practical problem and assures that the goals are higher–level thinking skills, that is, at the level of Application or higher in the cognitive domain of Bloom's taxonomy of educational objectives (Bloom 1956). Teachers can assess student performance on all six of these learning objectives in about two minutes or less for each student.

Grading versus rating

But what about grading students? The learning goal scales in Figure 1 give us a way to rate student attainment without assigning a summary grade. In practice, an assessment is conducted before instruction on density (or observation versus inference, or proportional reasoning, and so on) begins. This establishes the prior level of student attainment, and can highlight areas of concern for instructional planning. Assigning a grade based on this pretest would truly be punitive.

Indeed, using the results of this type of assessment to generate grades is detrimental. Doing so prevents students from fully engaging with the task—from recording possibly incorrect notions, for example. By implicitly encouraging only correct answers, grades undermine our ability to observe what is perhaps the most powerful information available for instructional planning: genuine examples of students' misconceptions. So, we encourage teachers working with our assessments to explicitly distinguish grading from assessment.

If a grade must be given, we suggest grading based on full engagement with the activity, not on attainment. Making this distinction clear to students is vital, otherwise students will be trying to feed us what they think we want to hear about the phenomena instead of what they truly believe. The latter is the true foundation for planning instruction. If the student is fully engaged in learning activities, follows all of our instructions, and still does not attain the learning goal, whose fault is it?

Another benefit of ratings versus grades is that ratings on learning goals are meaningful to parents and students (and principals, and superintendents, and policy makers). Information on the extent to which learning goals are being achieved over time suggests areas where more attention is needed and where resources may be allocated effectively, rather than encouraging comparisons of students' grades (e.g., "I need to get better at applying ratios and proportions," rather than, "I did better than you did").

After instruction is complete, summative assessments of these learning goals (individually or as a whole) can be conducted to determine the change in students' level of attainment. It is possible to design additional assessments targeting these learning goals and to use them during instruction, as well.

We hope that we have been able to demonstrate that a new approach to assessment is not only possible, but desirable. We begin by assuming that the primary purpose of these assessments is to support teaching and learning. We concentrate on the skills, concepts, and dispositions that educators and the public value most and circumvent the superficialities that characterize much current testing. These tests can be as valid and reliable as standardized tests, yet they are nonthreatening. In fact, they can be fun to administer and experience. The tests are efficient classroom activities that require minimal time for scoring and can be a learning experience in themselves. They can actually motivate students to learn more about the subject matter. We encourage science teachers to explore and try out these possibilities. ■

William E. J. Doane (wdoane@skidmore.edu) *is an adjunct instructor at Skidmore College, University Without Walls, 819 Broadway, Saratoga Springs, NY 12866; Rebekah R. Rice* (rrrice@acase.org) *is an associate researcher, and Paul Zachos* (paz@acase.org) *is director, both with the Association for the Cooperative Advancement of Science and Education (ACASE), 110 Spring Street, Saratoga Springs, NY 12866.*

References

American Association for the Advancement of Science (AAAS). 1993. *Benchmarks for science literacy*. New York: Oxford University Press.

Bloom, B. 1956. *Taxonomy of educational objectives*. New York: David McKay Company.

Zachos, P. 2004. Discovering the true nature of educational assessment. *Research Bulletin* 9(2): 7–12.