

EDUCATION UPDATE



[Buy this issue](#)

August 1995 | Volume 37 | Number 6
Designing Performance Assessment Tasks

Designing Performance Assessment Tasks

Philip Cohen

As Beth Larkins remembers it, developing new performance assessment tasks has not always been smooth. "I told my kids the first year I did this: 'You are the guinea pigs,'" says Larkins, a teacher at Rochambeau Middle School in Southbury, Conn., who writes and conducts training on performance assessment. But the effort has been worth it. "I've never seen them learn so much," she says of her 6th graders. "In 20 years of teaching, this is the first buzzword that really works for kids."

Teachers like Larkins are among advocates who emphasize the positive role that performance assessment can play in learning. The efforts of these advocates to popularize performance assessment have led to increased demand for carefully crafted assessment tasks. And despite the availability of off-the-shelf assessment tools, many districts and schools—including Larkins'—are creating their own performance assessment tasks.

Creating effective assessment tasks requires thinking through curriculum content to establish learning outcomes, then designing performance activities that will allow students to demonstrate their achievement of those outcomes, and specifying criteria by which they will be evaluated, experts say.

Grant Wiggins, director of programs at the Center on Learning, Assessment, and School Structure, recommends designing curriculum "backward from the assessment tasks"—deciding what students should be able to demonstrate they know and can do before deciding what to teach them. Such an approach lends coherence to the entire curriculum, he wrote in the 1995 ASCD Yearbook, *Toward a Coherent Curriculum*. "With clarity about the intended performances and results, teachers will have a set of criteria for ordering content, reducing aimless 'coverage,' and adjusting instruction en route; and students will be able to grasp *their* priorities from day one."

To develop meaningful performance assessment tasks that will reveal the learning that teachers hope to see, educators need to take an assessment perspective from the beginning, Wiggins believes. "If you think like an assessor, you're thinking, 'Given what I want them to learn, what counts as evidence that they understand that?'" he says. "That's a very different question than, 'What is a good activity?'"

Getting at Criteria

Performance assessment tasks should include carefully defined criteria. These are "the basis on which we judge," says Judy Arter, manager of the Evaluation and Assessment Program at the Northwest Regional Educational Laboratory (NWREL). Performance criteria specify what tasks are required of the student and how each element will be assessed.

Developing good criteria requires a change in approach for most teachers, Arter says. "We tend to be better at designing meaty, rich, 'authentic' tasks than at developing the criteria we use to judge quality performance on the task," she says. "We are used to having assessment mean task, activity, project, or problem, because that's the way it is on multiple-choice tests—you just score them right or wrong."

Joan Herman, associate director of the Center for Research on Evaluation, Standards, and Student Testing (and coauthor of ASCD's book *A Practical Guide to Alternative Assessment*), illustrates the need for clear criteria with a common example. Frequently, a writing assessment will offer students a narrative prompt and ask them to describe a personal experience. But then the teacher will evaluate the writing according to fixed criteria (such as plot, narrative coherence, and character development)

that were never explicitly told to the students. Vagueness in communicating criteria often stems from a misplaced fear of stifling student creativity, Herman adds.

At Gainfield Elementary School in Southbury, Conn., 4th-grade teacher Kelly Pelletier teaches a unit on the plant life cycle. In the last two years, she has developed a performance assessment that asks students to create a children's book explaining the subject to 3rd graders. With the new assessment, "the children's learning is tremendously increased. Now they know that what they're doing has a valid purpose," she says. "They are very proud of their results," and "they are constantly striving to improve."

With conventional paper-and-pencil tests a common problem is "teaching toward the test," or worrying more about how students will score on a test than about how well they actually learn, Jay McTighe of the Maryland Assessment Consortium explains. But the "paradox of performance assessment," he believes, is that "if the outcomes are worth spending time on, if the tasks really are demonstrations of understanding," and if good criteria are clearly explained, "then that's what we *ought* to be teaching to." Teaching toward the assessment becomes the goal instead of a problem, and the learning process is more coherent as a result, he says.

In Pelletier's classroom, she shows the students models of children's books that fit her criteria for "terrific," "OK," and "needs work." Each level includes specific criteria for the plant life cycle content (such as understanding seeds and flowers) and criteria for the written work (including organization and detail). Instead of taking one-shot tests at the end of the unit, Pelletier's students work on their children's books while she walks around the room, pointing out areas where their work falls short of "terrific," and helping them reach a higher standard.

"If the task that you're asking them to do is really authentic, if it's really meaningful to the students, I find that almost all the students strive to be terrific," Pelletier says.

To develop her task criteria, Pelletier drew a children's book performance assessment from her district's bank of generic assessment tasks—with criteria suited to any children's book project—and tailored it to include her own plant life cycle content criteria. In using this method, she says, "we're not getting away from our content objectives; we're just teaching them in a different way. The only real difference is we're looking at it from the perspective of having the students use what they're learning."

For William Spady, director of the High Success Network, however, such an assessment task is still just standing in for an abstract learning outcome; although the task is a much more meaningful experience for the children than a traditional test, it is still a "proxy" for what Spady calls "life-role performances." Spady believes assessment criteria have to reach beyond classroom uses of course content, tapping into "the ability to apply these skills in *real* situations and contexts. You need to look beyond school to define what kids need to learn," he says.

Developing assessment criteria is complicated, and presents challenges for design and implementation. McTighe warns against assessing complex tasks with simple criteria. If an assessment task includes several elements, such as individual research, writing, and group performance, "each of these needs to have a corresponding set of criteria," he points out.

McTighe also urges educators to distinguish between summative and formative purposes in assessment. For summative assessment—which seeks to communicate a general level of progress to evaluators outside the classroom—a single ability score for complex assessments may be appropriate or necessary. This type of assessment could include a single score for writing ability, for example. But for assessments designed to help teachers and students improve individual learning, criteria need to be more specific and detailed. Teachers need to know "not just how well the student has done, but why"—what aspects of the performance are acceptable and where problems exist.

Effective Tasks

McTighe suggests three elements of an effective performance assessment task:

- **Meaningful context.** "Good performance assessments are more contextualized" than traditional tests, he says, "more like how people use knowledge and skills in the larger world." Unlike many multiple-choice tests, good tasks do not jump from one area of knowledge to another.
- **Thinking process.** "Ask students to actually use knowledge," he says, "to thoughtfully apply knowledge and skills to a new situation. If you really understand something, you can work with it, analyze it, argue against it, and present it." Educators should ask of their assessments, "Could students accomplish this task and still not understand what we want to assess?"
- **Appropriate product or performance.** Avoid "products or performances that don't relate to the content" of what is being assessed, even though they may seem like good activities on their own. "Sometimes students get so caught up in the product that they lose sight of what they're actually intending to show with the product." One common problem is an overemphasis on aesthetic elements of an assessment task.

Developing good evaluation criteria also helps avoid assessing the wrong skill or knowledge. For example, "much of what goes on under the banner of alternative assessment is language dependent," Herman says. In trying to make tasks more authentic—by providing context, for example—teachers inadvertently test students on their reading or English skills. This effect undermines "validity" (measuring what you intend to measure).

In NWREL writing assessments, "we never punish the student for being off topic, not following directions," Arter says. "We are trying to assess ability to write, not ability to follow directions. Complex prompts, developed to be authentic, frequently result in tension between scoring responses for following directions versus scoring them for ability to write."

When assessment is more integrated into instruction, it may also become part of other reforms, such as greater student choice, interdisciplinary learning, and cooperative grouping. These methods all contribute to the instructional benefits of assessment—but they pose problems for measurability and equity:

- Student choice. “Student choice has lots of benefits,” McTighe says, “but you want to make sure that opportunities for choice don't get in the way of what you're trying to assess.” Allowing students to choose subjects, resources, methods, and whether to work alone or in groups has instructional benefits, but complicates assessment. “From a measurement perspective, giving students choices is a terrible dilemma,” Herman agrees. Some options or topics may yield easier projects than others, and “not all children are equally good choosers.” On the other hand, assigning topics runs the risk of giving an advantage to students who are more inclined toward what the teacher selects.
- Interdisciplinary tasks. Herman prefers these tasks because of their instructional value. But interdisciplinary assessment is most effective when a teacher is familiar with students' progress in several areas. A writing assessment on a history subject is hard to evaluate unless the teacher can distinguish the level of performance in writing versus that in history. These distinctions are harder to make when the people who rate the assessments don't work with the students every day.
- Cooperative grouping. “Any kind of group activity confounds the measurement of individual ability,” Herman says, although group work supports learning. Many educators include an individual component of the assessment in cooperative situations, but the performance of other students in the group can affect that component, Herman adds. And if teachers want to assess the ability to work as a team, that ability should be included in the criteria.

In each of these areas, experts agree, the teacher who knows the student best can integrate assessment and instruction most effectively. In the classroom, a teacher can also do ongoing assessments and track progress over time. But the higher the stakes of the assessment—when it will determine questions of funding, placement, or scholarships—the more important the issues of accountability and equity in a particular assessment become, Wiggins says.

The trade-off between good instruction and measurability is “a significant issue at the state level,” Wiggins says. “But it's a nonissue at the local level.” As a teacher, “you can factor out these things with your professional judgment and assessments over time.” Teachers can have more leeway for complex assessments in their classrooms, where they do not need the high level of precision necessary to make high-stakes tests accurate and fair. Wiggins adds that teachers should avoid using high-stakes techniques in the classroom when they're not necessary. “You have the whole life of the kid to collect evidence and decide what it means,” he says, so each task does not have to be invested with great importance.

During a yearlong project called “My Country,” Larkins' middle-school students work on an integrated social studies and language arts task to create a folktale for a country they invent. After studying the role of folktales in different cultures and comparing examples, students create their own, using a checklist of required elements: time and place, a problem, figurative language, a resolution, and a moral, for example. Students guide their own assessment, as they use the list to organize their work, filling it in as they write. “You have to leave room for the kids to create,” Larkins says. If the criteria are too rigid, the students are “stifled.” After she grades the stories, students may take them home and refine them for a final presentation.

Benefits for Teachers

Working from the end result backward—from assessment to instruction—is a challenge for educators, but the approach promotes reflection on “all the larger issues of teaching,” McTighe says. Teachers should have the chance to learn how to design quality assessments, he believes. “This is very difficult work. But it's one of the best experiences for educators,” in part because, having learned to develop effective assessment tasks, teachers will become better consumers of assessment products and services from outside the school. Herman agrees. “All our experience suggests that teachers' involvement in these new kinds of assessments has beneficial impacts on their instruction and their thinking about teaching,” she says.

Most teachers were not originally trained for “teaching kids complex performances,” Spady says. And that training is important for implementing and designing assessments. Wiggins argues that teachers should be asked to “justify their designs,” and he suggests a peer review process in which teachers compare one another's assessments with content and assessment standards. In Pelletier's district (Pomperaug Regional School District 15), educators are moving toward a system of selected “anchor tasks” that teachers in every school will use, in addition to their own assessments. These “tried-and-true” tasks will help ensure consistency across the district, she says.

Although creating effective performance assessment tasks may be an imposing job, McTighe is convinced that education will benefit from the effort. Performance assessment “has helped me organize myself. I know what I'm doing and I know what I'm looking for,” Larkins says. “This has been the best thing that's ever happened to me as a teacher.”

Copyright © 1995 by Association for Supervision and Curriculum Development

Requesting Permission

- For **photocopy, electronic and online access, and republication requests**, go to the [Copyright Clearance Center](#). Enter the periodical title within the “**Get Permission**” search field.
- To **translate** this article, contact permissions@ascd.org