

Putting Performance Assessment to the Test

Can performance assessment yield dividends for teachers and policymakers alike? Yes, several states have concluded.

JOHN O'NEIL

In Vermont this month, hundreds of 4th and 8th grade teachers will gather at regional sites, carrying portfolios of their students' work. In meetings expected to last three days, the teachers will score their students' mathematics portfolios — which will include five to seven “best pieces” of work — according to common criteria and scales. At another half-day meeting, a sampling of students' writing portfolios (already scored by the classroom teacher) will be cross-

scored by other teachers, again according to common criteria for quality. In both subjects, students are expected to show in their portfolios a range of skills and the ability to do complex thinking.

For Vermont, which had no state assessment program before inaugurating its portfolio system this year, the rationale for its new plan is clear. “We wanted kids to be engaged in doing interesting work,” says Ross Brewer, director of planning and

policy development for the Vermont Department of Education. “And we wanted to give teachers a test that was worth teaching to.”

Vermont's new assessment plan reflects a move away from the dominance of standardized paper-and-pencil testing in state testing programs, experts say. The inclusion of “performance” assessments — which include performance tasks as well as portfolios — may represent a new emphasis in state testing programs.

The new push for performance tests results, in part, from intense criticism of the impact of state testing programs. (State tests have doubled in number over the past two decades with 47 states now having such programs.)

The minimum competency testing programs that swept the nation during the 1970s and early '80s caused state policymakers to increase their use of standardized exams in an attempt to ensure that all students mastered at least some basic skills. “Their intent was to provide some rigor for the high school diploma, so that students could read their diploma when they received it,” quips Ed Roeber, an assessment expert with the Council of Chief State School Officers. In many ways, experts say, such state testing programs achieved a measure of success: increasing numbers of students mastered a modicum of basic skills, and state policymakers, some of whom had tied education funding increases to testing programs, were mollified.

But testing programs focusing on mastery of basic skills had other effects as well, critics say. By using



Photo: Paulson

multiple-choice standardized tests to assess skills that had been broken down into discrete parts, state testing programs rarely succeeded in assessing students' thinking skills or their ability to synthesize content or solve problems. Moreover, educators felt pressured to "teach to the test" because high stakes were attached to test results: schools were ranked by test results, for example, or students who failed a basic skills test were denied a diploma.

"Teachers get messages from tests," says Cheryl Tibbals of the state education department in Kentucky, which is overhauling its state assessment system. "What we've been doing for years is telling teachers to break down skills into 'factoids,' and we've been testing for those factoids." By slighting more complex student outcomes, critics say, basic skills testing essentially cemented into place low standards for student achievement.

A New Vision

The desire to ensure that students graduate with more than basic skills — with the ability, for example, to use skills to solve novel problems, work cooperatively in groups, or synthesize knowledge across disciplines — has fueled interest in performance assessment. Students taking part in performance assessments might be called upon to write an essay, perform a group science experiment, defend in writing how they answered a math problem, or keep a portfolio of their best work. In contrast, standardized paper-and-pencil tests, which typically require students to work individually and select answers from multiple choices, seem less appropriate for such outcomes.

"The achievement targets we have are more complex than ever before," says Richard Stiggins, who directs the

Center for Classroom Assessment at the Northwest Regional Educational Laboratory. "We now realize that without performance assessment methodology, it's not possible to create a complete portrait of student achievement. You can't evaluate writing without asking students to write, and you can't evaluate whether

The inclusion of "performance" assessments — tasks as well as portfolios — may represent a new emphasis in state testing programs.

a student has learned a foreign language without asking that student to speak."

Increasingly, states are looking to performance assessment to get a more complete picture of students' abilities. A 1990 survey by the Center for Research on Evaluation, Standards, and Student Testing at the University of California-Los Angeles found that nearly half of state testing programs either had performance assessments in place, were planning to implement them, or were actively exploring the idea.

Under the new state assessment plan being launched in Vermont, for example, all 4th and 8th grade students are keeping portfolios documenting their work in mathematics and writing. All 4th and 8th graders

also take a uniform test in writing and mathematics that will include a writing sample (in writing) and open-ended and multiple-choice questions (in math). Various pieces of the testing program are used to assess the progress of individual students, their districts, and the state as a whole. The state is considering a portfolio requirement at the high school level, says Brewer.

In Kentucky, students in grades 4, 8, and 12 this year took part in several performance assessments. They kept writing portfolios and participated in performance "events" in science, social studies, and math. Students also completed a more traditional assessment consisting of multiple-choice and open-ended questions. The results of these tests, combined with other data, will establish a baseline this year for district improvement. Teachers in schools showing a certain amount of improvement will receive financial rewards — up to a third of their annual salaries — and schools scoring very far below expectations face sanctions. Under a legislative assessment mandate that is among the most ambitious in the nation, the state assessment program must be "primarily performance-based" by 1996. "If you're going to have a results-oriented system, you have to have assessments that are worth teaching to," explains Thomas Boysen, the state's education commissioner.

Assessment Power

Officials in these and other states integrating performance assessments are betting that if paper-and-pencil basic skills testing had a powerful influence on classrooms, performance assessments might be just as powerful in pushing curriculum and instruction toward more complex outcomes. "Since we have to have high-stakes

tests, we made the decision to use performance-based tests that better reflect instruction as it should be," says Steven Ferrara, a state testing official in Maryland, which is adding some performance tests to supplement its minimum competency assessments. (See box, page 17.)

Experience with direct assessment of student writing, which has a longer history than some other performance assessments, supports the idea that better tests can drive more appropriate instruction, some experts believe. California used a multiple-choice test to assess writing until 1987, when it switched to direct writing assessment, says Dale Carlson, assistant superintendent for the California Assessment Program. A study done after the new test was implemented found that more than 90 percent of English teachers made changes in their teaching — by assigning more and different kinds of writing tasks — partly as a result of the new test.

In Pikeville, Kentucky, a 1,400-pupil Appalachian district, the new performance tests required by the state have helped to refocus classroom instruction, says Anne Keene, an instructional supervisor for the district. Teachers, even some who are not teaching in grades 4, 8, and 12 (the grades tested in the state's accountability plan) have been using authentic tasks and portfolios that push students to demonstrate their mastery of content and thinking power. "Teachers are starting to incorporate these kinds of assessments into their own classrooms," says Keene.

Another benefit of performance assessment, proponents believe, is its potential to draw teachers back into the heart of the assessment process. With centralized paper-and-pencil assessments, exams are shipped to schools in boxes, administered to



© Richard McCorn, Kentucky Department of Education

Some experts say performance assessments are likely to be at least two or three times more expensive per student than multiple-choice tests were.

students, and scored by outsiders. Critics of such exams say that getting teachers involved in developing assessment tasks, devising criteria for rating student work, and scoring student work themselves can be a powerful form of staff development.

"There's a movement, fledgling though it is, but an important one, for teachers to become an important part of the assessment process," says Joan Baron, coordinator of the Connecticut

Common Core of Learning Assessment Program. As part of the program (which receives funding from the National Science Foundation), teachers are helping to develop and try out in their classrooms performance tasks in mathematics and science for high school students. Earl Carlyon, a physics teacher at Manchester High School in Manchester, Connecticut, who has been involved with the program for three years, says it has helped him and his students feel "empowered." "I'm aware of nothing that's more powerful than performance assessment" to change classroom practice, he says. "You cannot assess performance unless you teach performance."

Cost and Time Concerns

Despite the enthusiasm for performance assessments, a variety of concerns confront their use in state assessment programs.

One is cost. Although estimates vary on the costs of traditional machine-scored, multiple-choice tests versus performance assessments, some

Maryland's Reading-Writing Examination

Maryland is using performance assessment as a lever to improve programs. Assessments were developed by the State Education Department and Maryland teachers with the technical assistance of CTB/McGraw-Hill. Following is an example from the reading-writing examination.

Maryland has taken the natural parallel between reading and writing to the point where they are integrated into one assessment. The entire process models how to teach reading and writing together, directly in opposition to the skills approach, which breaks down reading (and then writing) into tiny steps, tests each through workbook drills, and uses basal readers with controlled vocabulary. To show how it works, I will describe the sample grade 8 reading-writing/language use assessment.

Each student is given a reading book, which contains a map of North America on the first page, with Canada, Alaska, and Yukon Territory marked on it; a short story by Jack London entitled "To Build a Fire"; and an excerpt from *Hypothermia: Causes, Effects, Prevention*, by Robert S. Pozos and David O'Born, published by New Century Publishers in 1982. Students also have response books into which they write answers.

The assessment begins with a prereading activity, which focuses students on the topic — the deadly cold of the Yukon Territory and its dangers — by asking them to think about their own experiences of being cold. They are asked to spend 10 minutes writing a journal entry

describing their experience on the appropriate page of their response books. Then they read London's "To Build a Fire" and respond to a series of questions probing their comprehension of the story. The first question can be answered with a drawing of the scene of the action if the student prefers to draw rather than write. A question later in the sequence asks the students to compare their own experience, described in the journal entry, with that of the man in the story who dies in the extreme cold. The final three questions probe the students' reading abilities by asking them to assess the difficulty of the story and explain why they rated it "very easy," "somewhat easy," "about average," "somewhat hard," or "very hard" and describe their reading strategies, that is, what they do to make sense of the story when they come to a word or a reference they do not understand.

On the second day of the assessment, the students begin by writing a 5-minute letter to the man in London's story giving him some advice that might have saved his life. Before they read the excerpt *Hypothermia: Causes, Effects, Prevention*, there is class discussion about the topic, with the teacher writing on the board a cluster of the students' ideas as they respond to the words (succumb, insidious) that they will find in the excerpt. After they read the piece, they respond to a series of questions, again including the option to draw a picture or a diagram for at least one of them.

On the third day of the assessment, the students are expected to integrate

the information from the two pieces into a written response to one of three situations: informing a group of friends of what they will need to do to stay safe on a winter weekend trip; writing a poem, story, or short play expressing their feelings about extreme states, not only cold but also heat, hunger, or fatigue; or writing a speech to persuade people to avoid travel in the Yukon. As in the case of Arizona and California, teachers will cover these three kinds of writing because they know that one of them — but they do not know which one — will be used in the assessment.

In each case, the student is asked to go through a process of first brainstorming ideas and either listing them or making a web of words with lines connecting them to major ideas. (These graphic organizers are now a recognized part of teaching the writing process.) Students write a rough draft, pause to consider whether it meets the needs of the situation, and then revise the piece. Finally, they use a proofreading guidesheet supplied in the response book to prepare a final copy. . . . The material is graded according to two rubrics: one for the answers to the questions designed to measure reading comprehension and the other for the persuasive or informative writing. The prereading and prewriting activities and the class discussion are recorded but not scored.

— Excerpted from: R. Mitchell, (1992), *Testing for Learning: How New Approaches to Evaluation Can Improve American Schools*, (New York: The Free Press), pp. 47-49.

Testing for Learning: How New Approaches to Evaluation Can Improve American Schools

Ruth Mitchell
New York:
The Free Press, 1992

Testing for Learning makes an important and timely contribution to the educational reform dialogue. In a highly readable style, Mitchell offers a lucid description of current issues, practices, problems, and potentials related to testing and assessment in this country. The book is intended for a broad audience, both within and beyond the professional education community, as indicated by its refreshing absence of psychometric terminology.

The book examines testing practices in terms of their effects on teaching and learning. This perspective guides Mitchell as she traces the history of testing practices in the United States, discusses the move-

ment from testing to assessment, and vigorously critiques the current reliance on multiple-choice tests. Mitchell goes on to describe a number of innovative state and local assessments that call for more authentic performance by students. Using detailed examples, she presents the multiple merits of developing "tests worth teaching to" while acknowledging the unresolved problems with performance and portfolio assessments.

Testing for Learning is highly recommended for educators, parents, and others interested in an up-to-the-minute account of the major issues related to testing and assessment. It should be required reading for national, state, and local educational officials, as well as for board of education members, state legislators, and others whose voices influence educational policy.

Available from The Free Press,
866 Third Ave., New York, NY
10022, for \$19.95 (hardback).

— Reviewed by Jay McTighe, Maryland
State Department of Education, Baltimore,
Maryland.

experts say performance assessments are likely to be at least two or three times more expensive per student. Performance assessments may require special equipment, for example, or funds to hire substitute teachers for educators participating in assessment training or scoring. "The costs could be very high, and we're at a time when a lot of people are cutting their budgets," notes Pamela Aschbacher, who conducted the UCLA survey.

Another concern is the time that performance assessment requires in a school year already crowded with other tests. Students in some grades in Frederick County, Maryland, spend about a week participating in the state's performance assessments in mathematics and language arts, says

Carolyn Strum, the district's coordinator for elementary language arts. Some students also must take a standardized norm-referenced test, and the district has its own assessment program as well. "If there's a disadvantage to performance assessment, it's the amount of time needed for the state piece," says Strum. Moreover, the time spent on administering performance assessment itself pales in comparison to the time needed for teachers to revamp their instruction to better prepare students for the new tests. Both the class time and the preparation time are higher for teachers who aim to meet the demands of the new assessments, says Clare Forseth, a teacher at Marion Cross School in Norwich, Vermont, who

serves on a state committee for portfolios in mathematics.

Psychometric Cautions

In addition, state assessment programs must address concerns that the psychometric capabilities of performance assessments do not match those of traditional standardized tests. When high stakes are attached to performance tests, experts point out, the importance of psychometric quality grows.

"Performance assessments will be done in a high-stakes environment," Brewer says. "We need to be able to show that the tests can be generalized — from tasks to domains, from one day to the next, and from one rater to another."

Beyond specific concerns about the technical quality of tests, testing officials must decide if state assessment programs can serve *both* the goal of improving classroom instruction and the goal of making teachers, schools, and districts accountable for student learning. "We have no choice," says Carlson. "Anybody can build 12 different tests for 12 different purposes. . . . We have to have a system that does both so that the data for the accountability part comes from the right kind of testing."

But if an assessment is put in place to drive improved classroom instruction, should states implement such tests and begin reporting scores even if students have not had an opportunity to learn in ways measured by the tests? "That's exactly what the quandary is," says Baron. "The art of practice needs time to catch up with the assessments you are using to model best practice."

Baron and others are concluding that until such questions are answered more fully, states should not use some of the newer performance assessments

for high-stakes purposes. "Performance assessments could suffer from some of the same problems of standardized tests if the same high stakes are applied to them," says Aschbacher. "While we are developing expertise seems to be a bad time to be applying high stakes." Such advice seems apt when the possibilities of legal battles over tests are very real. "It's not a matter of whether we'll have lawsuits, it's a matter of when," says one official from a state that is implementing performance assessments and tying results to accountability measures.

Caution Urged

The presence of such challenges means that although many state policymakers

are interested in performance assessment, not all are buying — at least not yet. At this juncture, some state policymakers are "hoping that other states will come up with the model that will allow them to switch" to a more performance-based assessment system, says John Myers, education program director with the National Conference of State Legislatures.

A likely scenario for many state testing programs, experts say, is that they will move cautiously on performance assessments, beginning with small pilot projects, waiting for more research on the costs and benefits of the newer tests, and continuing to rely to some extent on traditional standardized tests. But it's not clear how high-stakes state testing programs that are

part performance-based and part multiple-choice will affect classroom instruction. "If one is high stakes and the other isn't, why pay attention to the one that isn't?" asks Aschbacher.

Moving slowly on state performance assessments may ultimately be a good thing, say some experts who worry that the new tests may be oversold. "One of the things that troubles me greatly is that we're setting up performance assessments and paper-and-pencil tests against one another," says Stiggins of the Northwest Lab. "Each test has a contribution to make. We can't throw away any of the tools at our disposal." □

John O'Neil is Contributing Editor to *Educational Leadership*.

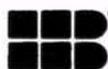
Want to work wonders in your district?

Choose the integrated curriculum of



- Universal Themes in Fiction and Nonfiction
- Related Activities Across the Content Areas
- Cooperative Learning
- Authentic Assessment
- Parental Involvement
- For Grades 1-4

Plus ¡Qué maravilla!, a parallel program in Spanish



HAMPTON-BROWN BOOKS
P.O. Box 223220 Carmel, CA 93922

Make **WONDERS!** happen.

Call 1-800-933-3510

for more information and a free sample.

Copyright © 1992 by the Association for Supervision and Curriculum Development. All rights reserved.