

IN PERSPECTIVE

 www.edweek.org/go/states

Adjusting to Test Takers

By Katie Ash

For Jeannine Ugalde and her middle school students, using assessment data to inform classroom instruction is a regular part of the school day.

Drawing on the results of computer-adaptive tests given periodically throughout the year, the 7th and 8th grade humanities teacher and her students at Oak Valley Middle School in San Diego set classroom goals that target the areas the youngsters struggle with the most.

"It's hard work coming up with goals, so they feel a lot more ownership of their education," said Ms. Ugalde, whose 1,050-student school is part of California's Poway Unified School District. "And now they specifically say, 'Remember, you said we were going to work on this.'"

Jerry Chen, one of Ms. Ugalde's 7th graders, explained.

"The data can help me make goals because I learn where my weaknesses in school are, so I know what to make my goal," he said. "[The information] helps me find where I am and helps me set goals according to my needs."

As more schools gain the technological knowledge and hardware to implement computer-based tests, districts are showing a growing interest in computer-adaptive testing, which supporters say can help guide instruction, increase student motivation, and determine the best use of resources for districts.

A computer-adaptive assessment is one that uses the information it receives during the test to determine which question

to present the test-taker with next. For example, when a student answers a question correctly, he or she will be presented with a harder question, but if the answer is wrong, the next question will be easier. Consequently, a 5th grader taking the test could answer questions at the 6th or 3rd grade level, depending on his or her responses.

“*[In a fixed-form test], low-ability students are going to get questions that are too difficult, and they're going to be frustrated. [With an adaptive test], everybody will be equally challenged.***”**

DAVID J. WEISS
Professor of Psychology
University of Minnesota-Twin Cities

This method of testing shortens the test by not asking high-achieving students questions that are too easy for them, and likewise not giving struggling students questions that are too hard.

In essence, "each student gets questions that are appropriate just for them," said David J. Weiss, a professor of psychology at the University of Minnesota-Twin Cities and an expert on computer-adaptive assessments.

"[In a fixed-form test], low-ability students are going to get questions that are too dif-

ficult, and they're going to be frustrated," he said. "[With an adaptive test], everybody will be equally challenged."

In addition to shortening the length of the test, the approach creates a fairer psychological test-taking environment for each student, said Mr. Weiss, who has studied computer-adaptive tests since the 1970s.

'Kid-Centric' Assessments

The Northwest Evaluation Association, a nonprofit organization based in Lake Oswego, Ore., that partners with school districts to provide testing services, has created one of the most widely used computer-adaptive assessments. Called the Measures of Academic Progress, or MAP, it is used by more than 2,340 school districts in the United States, including the Poway Unified schools, and in 61 other countries.

By adjusting in difficulty based on the student's performance, MAP makes assessment "kid-centric," said Matt Chapman, the president and chief executive officer of the NWEA. And after administering the assessment, the teacher can see not only if the student passed a certain benchmark, but also pinpoint exactly where the student's achievement level is—whether below, at, or above grade level.

"It increases the information you learn about the student, the student's growth, and how the class is doing," Mr. Chapman said of the method.

Teachers, then, can use the information "to inform instruction at the classroom level,"

How Adaptive Testing Works

With a computer-adaptive assessment, the questions adjust in difficulty based on the student's previous answers. Here is how the assessment might work to determine a middle or high school student's proficiency in basic grammar and usage.

The student wrote a report _____ Abraham Lincoln, our 16th president.

Which preposition would best complete the prepositional phrase?

1. in
2. from
- 3. about**
4. with

If the student answers this question correctly, he or she will be presented with a harder question.

Which sentence is not correct?

- 1. The hike took us the most farthest we had been from camp.**
2. This project required less work than anyone had anticipated.
3. The patient said she felt worse in the morning than at night.
4. We will have to walk farther than we were told.

If the student answers this question incorrectly, he or she will then receive a question that is easier than the second, but more difficult than the first.

Choose the missing word.

The sun seems to rise _____ each day than it did the day before.

1. early
2. more early
- 3. earlier**
4. most early

The assessment continues until enough data is gathered to determine the student's proficiency level. By eliminating questions that are too far above or too far below the student's achievement level, that data can be gathered with fewer questions, making the test shorter.

SOURCE: Northwest Evaluation Association

said Ginger Hopkins, the vice president of partner relations for the NWEA. "It allows teachers to adjust whole-group instruction and create flexible grouping," for students at similar achievement levels, she said.

Some schools, such as the 150-student K-8 Trinity Lutheran School in Portland, Ore., have used the test to regroup classrooms based on the most efficient use of resources. Facing declining enrollment and budget constraints, the school has used MAP to reconfigure students into multiage classes, said Principal Jim Riedl.

The Chicago International Charter Schools—a group of 12 charter schools teaching 7,500 students—is using the assessment in part to evaluate the efficacy of each of the four educational management organizations, or EMOS, the network employs.

“ We think that [a computer-adaptive assessment] offers certain benefits, and it has certain drawbacks. Like all testing, it really depends on what you're trying to do, and what you're trying to learn.”

SCOTT MARION

Vice President
National Center for the Improvement of Educational Assessment

"It's very high-stakes for the EMO providers, and I'm not saying for our schools that are underperforming, it's not scary," said Beth D. Purvis, the executive director of the group of

schools. "But frankly, if you're an underperforming [school], you should be scared. It is a fair way of saying, 'Get your act together.'"

Engaging Learners

Although many teachers were skeptical about the assessments when they were introduced in Poway Unified in 2001, "as students' performance on the state test provided us with evidence that it was working, the schools came on board one by one," said Ray Wilson, the recently retired executive director of assessment and accountability for the 33,000-student district, who oversaw the school system's adoption of MAP.

The assessments have especially helped students on the upper and lower ends of the performance spectrum, said Linda Foote,

the instructional technology specialist for the district.

In fixed-form tests, “students who are high-performing could look good all year without much effort,” she said, “and the struggling students could work and grow dramatically” but still appear to be underperforming, she added.

With MAP, “we don’t honor students who aren’t working and dishonor students who are,” Ms. Foote said.

The assessments also play a key role in increasing student motivation, she said.

“One of the most stunning pieces is that the kids, because they finally have the connection between the work that they do and the progress they make,” Ms. Foote said, are “much more willing ... to do the work.”

The approach makes students active participants in their learning, she said, and “takes the mystery out of education.”

NCLB Constraints

Still, only one state—Oregon—uses a computer-adaptive assessment for reporting purposes under the federal No Child Left Behind Act.

Because NCLB requires that students be tested on grade level, most computer-adaptive tests, which may present students with questions above or below grade level depending on how they answer previous questions, are not allowed for accountability purposes.

However, Oregon’s test—the Oregon Assessment of Knowledge and Skills, or OAKS—is an adaptive test that stays on grade level, said Anthony Alpert, the director of assessment and accountability for the state department of education.

The test was piloted in 2001, the year the NCLB legislation passed Congress, and has since grown in popularity among school districts in the state, said Mr. Alpert. Today, almost all of Oregon’s state testing is through the computer-adaptive online assessment, he said.

“The real brilliance of the founders of the online system in Oregon was making it an optional approach and letting districts be innovative in making decisions in how they’re able to participate,” he said. As an incentive to encourage districts to use OAKS, each student is allowed to take the assessment three times a year, and the state records only the student’s highest score.

States Slow to Embrace Online Testing

By Katie Ash

Although most assessment experts agree that in the future, state tests will routinely be administered by computer, progress toward that goal has been slow, expensive, and fraught with logistical challenges.

“I don’t think any state has found it easy to go to computer-based testing,” said Stuart R. Kahl, the president and chief executive officer of Measured Progress, a Dover, N.H.-based nonprofit organization that provides assessments to school districts.

Although many districts have implemented computer-based testing at the district and school levels, using computers for statewide assessments is much more difficult, Mr. Kahl said.

“Districts that buy in to local online testing tend to have success, and then they expect everything to be just as smooth for statewide testing,” he said. But heightened security concerns and the increased number of students who need to take the test make computer-delivered statewide assessments harder to carry out, he said.

One big difference between using computer-based tests at the local level and the state level is the limited testing window required for statewide assessments, Mr. Kahl pointed out.

To keep students from sharing or discussing questions on the assessment, it’s important for all students to complete the test within a very short time frame. But with such a narrow window, schools need more computers so students can take the test at the same time, and beefed-up network capacity to handle the spike in bandwidth required for the assessment applications to run smoothly.

And even if schools have the hardware and the bandwidth to support computer-based statewide assessments, it doesn’t mean they’re ready to administer the tests, Mr. Kahl said. “There’s so much more that has to happen.”

Still, the advantages of computer-based assessments—quicker feedback, better-organized data, and a much less labor-intensive administration than with paper-and-pencil tests—are encouraging states to find solutions to those obstacles.

Cost Barriers

Michael Griffith, the senior financial analyst for the Denver-based Education Commission of the States, is working with the state of Texas to come up with a cost estimate for moving its statewide assessment online.

“A lot of people get excited, and there are a lot of possibilities,” he said, “but there are a lot of costs involved and logistical problems that we’re finding with the state of Texas that run into the cost issue.”

For example, buying enough computers for all students to be able to take the test on the same day is a major expense, said Mr. Griffith. Expanding the testing window can help lower that cost, but it also decreases the level of test security and ties up computer labs for a longer amount of time, which can disrupt classes.

Another option might be for schools to administer the test at an outside location, such as a testing center, said Mr. Griffith, but whether that solution would be any more cost-effective or easier to coordinate is up for debate.

In addition, teachers who administer computer-based assessments need to be trained on how to use the system so they feel comfortable with the technology, said Mr. Griffith.

Lastly, the technology needed to run computer-based assessments is not just a one-time expense, Mr. Griffith pointed out.

“You need computers that are fairly new,” he said. “And once you get it going, you’ve got to refresh your computers once every five years, so it really has ongoing costs.”

Prospects Vary

To defray the initial cost of moving to a computer-based statewide assessment system, some states—such as Virginia—have chosen to provide state assessments via computer to a subgroup of students, while most students still take a traditional paper-and-pencil assessment.

But using a mixed model cannot reap all the benefits of a 100 percent computer-based assessment delivery, said Mr. Kahl, of Measured Progress.

“A mixed model has the added expense of delivering [the assessment] in both modes,” he said.

It also prevents states from being able to tap in to the true potential of computer-based assessments, such as incorporating interactive items and multimedia, he said.

“I think what’s happening now is some states are in the position to get started, other states are saying ‘we’re not ready,’ and there are many states that aren’t even dabbling in it,” Mr. Kahl said.

Mr. Griffith, of the ECS, agreed.

“With the amount of testing that’s being done with [the federal No Child Left Behind Act], to be able to collect and organize your data, and report it back that much quicker—that’s become very attractive for states,” he said. “But because there is a cost to get this started, I don’t think you’re going to see a lot of states moving forward with this in the next few years.”

Although no other state uses computer-adaptive assessments for reporting purposes under NCLB, Utah Governor Jon M. Huntsman Jr., a Republican, last month signed into law a bill that allows a pilot test of such assessments in his state.

This school year, three districts are participating in the pilot, said Patti Harrington, the Utah superintendent of public instruction. The decision to use computer-adaptive testing came from recommendations from a panel on assessment that included superintendents, legislators, parents, and state- and district-level administrators, as well as Ms. Harrington.

“We like it because it does so much more than just say whether or not a student reached a certain point,” she said. “It emphasizes growth and can be used throughout the year to inform instruction. And because it’s computer-based, teachers have the results immediately.”

Addressing the Drawbacks

But computer-adaptive assessments are not the best way to evaluate students in every situation, experts point out.

“We think that [a computer-adaptive assessment] offers certain benefits, and it has certain drawbacks,” said Scott Marion, the vice president of the National Center for the Improvement of Educational Assessment, based in Dover, N.H. “Like all testing, it really depends on what you’re trying to do, and what you’re trying to learn.”

Computer-adaptive tests don’t provide as detailed diagnostic information as standards-based assessments, said Mr. Marion.

“A score is not diagnostic. ‘You’re pretty good in this area’ is not really diagnostic,” he said. “And in the standards-based world, that’s where we really run into trouble.”

Like Mr. Marion, Neal M. Kingston, an associate professor at the University of Kansas, in Lawrence, and a co-director of the Center for Educational Testing and Evaluation, also has concerns about computer-adaptive assessments.

“Item-response theory, or the mechanism used to determine which items are easier or harder, ... assumes there’s a universal definition of hard and easy,” he said. In some subjects—reading, for instance—that assumption may hold, he said, but for other subjects—such as high school math, which may combine algebra and geometry questions—that assumption isn’t always correct.

“The adaptive-testing model assumes that everyone has taken [courses] or learned [subjects] in the same way,” which is not always the case, Mr. Kingston said.

Subjects such as social studies, in which the curriculum varies greatly from place to

place, present a particularly difficult challenge for computer-adaptive tests, which are often created on a national or state level, he said.

“I think that the companies that provide these kinds of tests have an obligation,” Mr. Kingston said, “to provide more technical information than I have seen them do in the past about the appropriateness of the models that they’re using to determine whether it’s the same test in rural Kansas as it is in the center of New York City.”

Both Mr. Marion and Mr. Kingston say they recognize the potential of computer-adaptive testing, even as they voice caution.

“It’s a pretty nice framework for making certain types of tests for certain purposes,” said Mr. Marion, “but the promises—from what I’ve seen—far exceed the practice.” ■

Staff Writer Stephen Sawchuk contributed to this report.

Coverage of new schooling arrangements and classroom improvement efforts is supported by a grant from the Annenberg Foundation.

Reprinted with permission from Education Week, Vol. XXVIII, Issue 13, November 19, 2008, by The Reprint Dept., 800-259-0470. (11370-1208).

For web posting only. Bulk printing prohibited.

Editorial & Business Offices:
Suite 100, 6935 Arlington Road
Bethesda, MD 20814
(301) 280-3100
FAX Editorial (301) 280-3200
FAX Business (301) 280-3250

Education Week is published 45 times per year by Editorial Projects in Education Inc. Subscriptions: U.S.: \$79.94 for 45 issues. Subscriptions: Canada: \$135.94 for 45 issues.
