JULY 2010

EDUCATION WEEK

SPOTLIGHT

On Assessment

Editor's Note: Assessment is complicated in both practical and policy terms, raising myriad questions of how to best gauge student learning. This Spotlight looks at how schools and experts are approaching assessment.

CONTENTS:

Published January 27, 2010, in Education Week

Open-Ended Test Items Pose Challenges

By Stephen Sawchuk

Most experts in the testing community have presumed that the $350 million promised by the U.S. Department of Education to support common assessments would promote those that made greater use of open-ended items capable of measuring higher-order critical-thinking skills.

But as measurement experts consider the multitude of possibilities for an assessment system based more heavily on such questions, they also are beginning to reflect on practical obstacles to putting such a system into place.

The issues now on the table include the added expense of those items, as well as sensitive questions about who should be charged with the task of scoring them and whether they will prove reliable enough for high-stakes decisions. Also being

confronted are matters of governance—the quandary of which entities would actually "own" any new assessments created in common by states and whether working in state consortia would generate savings.

"The reality is that it does cost more to base a system on open-ended items, no question about it," said Scott Marion, the vice president of the Dover, N.H.-based Center for Assessment, a test-consulting group, who is advising several states. "If the model we're thinking about has got to be on-demand and high-stakes and used in systems with scores that are returned quickly, then it's going to cost a lot."

Higher Costs?

State dependence on multiple-choice testing under the federal No Child Left Behind Act has led to a backlash by those who say the tests, while cheap and technically reliable, come at a cost: not measuring complex cognitive skills.

Using a slice of money from the $4.35 billion Race to the Top Fund, created last year under the American Recovery and Reinvestment Act, U.S. Secretary of Education Arne Duncan has called for state consortia to craft richer item types aligned to common standards that would include constructed-response questions, extended tasks, and performance-based items in which students would apply knowledge in new ways. *(See Education Week, Aug. 12, 2009)*

His department is expected to open up a competition in March for the assessment aid from the stimulus law. Work is under way, meanwhile, on a national project to produce a "common core" of academic standards for adoption by states.

No Redos

Assessment experts caution that open-ended test items carry with them a number of practical challenges. For one, items that measure higher-order skills are generally more expensive to devise, depending on how extensive the items are and how much of the total test such items make up.

For instance, with their detailed prompts and scenarios, questions that require students to engage in extensive writing or to defend their answer choices often are "memorable," meaning the items can't be reused for many years and must be replaced in the meantime.

Wes Bruce, the chief assessment officer for the Indiana education department, recalled one prompt that required 5th graders to write about what would happen if a kangaroo bounded into the classroom.

> "While most people agree that some amount of analyzing student work is good professional development, every teacher probably doesn't have to do 100 papers to get the full value of it."
>
> JOAN HERMAN
> Director of the National Center for Research on Evaluation, Standards, and Student Testing

"All across the state, kids were talking about the prompt," he said. "From an assessment perspective, that's not good. Teachers will use it [subsequently] as an example for classroom work."

The scoring process for open-ended items is also far more complicated than sticking a bunch of test papers into a computer scanner. It relies on "hand scorers" who are trained according to a scoring guide for each question and a set of "anchor papers" that give examples of performance on the item at each level. Each open-ended item typically goes through multiple reviews to ensure consistent scoring.

Depending on the complexity of the item and how long it takes to score, the costs can increase dramatically. A short constructed-response item with four possible scores might take one minute to score, said Stuart R. Kahl, the president of Measured Progress, a nonprofit test contractor based in Dover, N.H. But an extended performance-based or portfolio item might take up to an hour, he said. With test scorers paid in the range of $12 to $15 an hour, such costs would add up.

For a mid-size state with about 500,000 students within the tested grades and subjects, the scoring of tests based even partly on constructed-response items would make up more than a fifth of the total annual contract cost, Mr. Kahl estimated.

Scoring Scenarios

For some, the idea of expanding human-scored items raises issues of reliability: Performance-based items are typically less mathematically reliable than those based entirely on multiple choice.

Todd Farley, a 15-year veteran of the test-scoring business who detailed his experiences in a recent book, *Making the Grades*, is among the skeptics. In the book, Mr. Farley alleges that the scoring guidelines for open-ended items were frequently counter-intuitive, and that as a "table leader"—an individual supervising other scorers' work—he occasionally changed other reviewers' scores.

Though test publishers interviewed for this story dismissed Mr. Farley's account, independent sources do point to areas of concern. At least two reports issued by the Education Department's office of inspector general last year, for instance, found lapses in Florida's and Tennessee's oversight of test contractors charged with scoring open-ended items.

In part to ameliorate the errors and costs associated with human scoring, test publishers are investing heavily in automated-response systems that use artificial-intelligence software to "read" student answers.

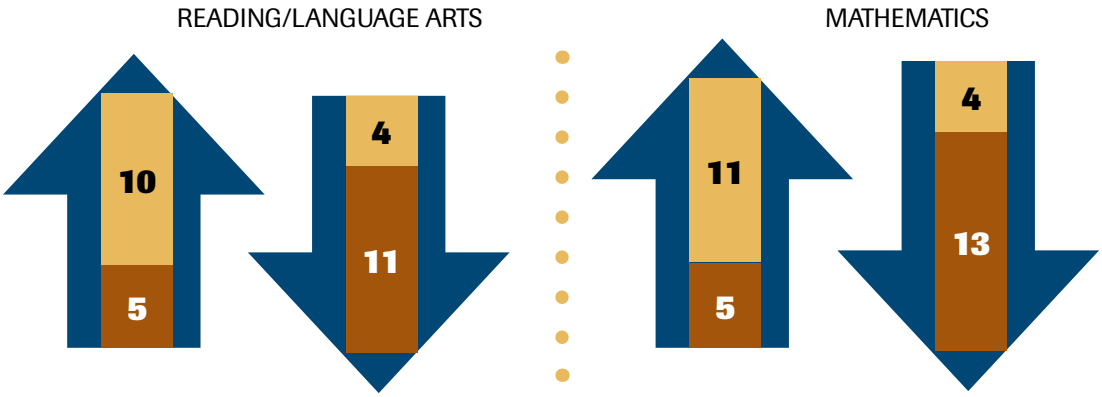
Such programs are already in use to minimize the number of human scorers needed for major tests such as the Graduate Record Examination. A handful of states, including Indiana, have piloted the technology for their own tests, and some experts, like Joan Herman, a director of the National Center for Research on Evaluation, Standards, and Student Testing, or CRESST, a group of assessment researchers headquartered at the University of California, Los Angeles, believe that the technology will be widespread within five years.

"When that happens, it will open up entirely new windows for doing more complex, open-ended items on large-scale assessments," Ms. Herman said.

But such systems are not perfect, and experts including Mr. Bruce, in Indiana, noted that one of their limitations is that they typically don't generate interpretative information about where a student needs to improve.

And in an era in which most teachers are distanced from the assessment process, unions and other stakeholders argue that teachers should have a greater role in the development and scoring of assessments, which can serve an important function for gathering information on student performance.

Teacher scoring of assessments has generally been eschewed under the NCLB law, the current version of the Elementary and Secondary Education Act, in part because of fears that the law's accountability pressure would cause educators to inflate their students' scores. But other countries have



States Reporting Changes in Item-Type Use on NCLB Assessments Since 2002

Uphill Climb

A push for open-ended assessments comes after states reduced their use in the wake of the No Child Left Behind Act.

Multiple choice
Open/constructed

SOURCE: Government Accountability Office

tackled the problem by using a blind-scoring process, in which teachers meet in central locations to score exams with names removed.

Hiring substitute teachers and paying the scoring teachers for their time would be costly as well. But there are benefits in the form of professional development, said Marcia Wilbur, the executive director of curriculum and content development for the College Board's Advanced Placement program, which uses such a system to grade open-ended essays on AP exams.

Teachers, Ms. Wilbur said, spend a week looking at sample responses at each level of the scoring guides before they begin to score, and the guides often help them better understand students' learning progressions.

"Teachers will bare their souls about what they do in their classrooms while discussing the samples," she said. "It's not that they go back to their classrooms and teach to the test, but they have a better understanding of the skills and the features of the skills that students struggle with."

Another model, drawn from international practice in places such as Australia, England, Hong Kong, and New Zealand, would rely on teachers' scoring assessments that are embedded within classroom curricular activities. Kentucky used such a system in the days before the 8-year-old NCLB law.

Not only could such assessments provide better feedback on where instruction needs to be improved, but they also could be included in state gauges of achievement if there were a central auditing process to ensure appropriate administration and reliable scoring, CRESST's Ms. Herman said.

Aside from costs, the problem with systems that rely heavily on teacher scoring, Mr. Kahl argued, is that results cannot be scored as quickly or efficiently as those done by machine. That means it could prove tough to turn results around under the quick timeline envisioned by the NCLB law and current state accountability systems.

Both Ms. Herman and Mr. Kahl said it might be possible to build systems that coupled curriculum-embedded assessments, scored throughout the year, with information from more-typical "on demand" standardized tests. But such a model hasn't been tested in the United States.

"While most people agree that some amount of analyzing student work is good professional development, every teacher probably doesn't have to do 100 papers to get the full value of it," Ms. Herman added.

Consortia Challenges

Also uncertain is which entities would actually own test forms or items developed in common, and which would bear the responsibility for updating them. Under the current system, such decisions are now made through contracts with individual states.

A common test, Mr. Marion of the Center for Assessment said, argues for some kind of open-source setup, but the details are "dicey," he concedes. "The reality is that if this is paid for out of stimulus funds, the country should own them," he said, "but at some point, someone has to house these items."

Mr. Kahl, meanwhile, warns that one of the major selling points about state consortia—cheaper tests—might yield only limited savings for some states. Consortia would probably ease costs more for small states, where test development is a high percentage of overall assessments, but might not help larger states, where most testing costs involve factors such as printing, scoring, and reporting results.

Such questions of governance, finance, and sustainability have drawn the concern of policy experts, too.

Doubts and Hopes

Chester E. Finn Jr., the president of the Thomas B. Fordham Institute, warned in a recent article for the Washington think tank's newsletter that the federal competition could lock in features of an assessment system that would be difficult to change in the future.

He also worries that the Obama administration's ambitious goals for the assessment funding—which include generating information about both school and student performance as well as data about teacher effectiveness—could prove to be irreconcilable.

"If all the glitterati ... remains in the grant competition, anyone that wants to win the competition is going to have to pretend they can do all those things," Mr. Finn said in an interview. "But since we know that they can't all be done by the same assessment, in the same period of time at a finite price, something will get left in the dust."

But for all those challenges, state testing experts hope to see breakthroughs with the federal funding.

"They are expecting that there be some innovation in the assessment area," Mr. Bruce of Indiana said. "As a state that has been committed to using what at one point were innovative item types, and is still looking at ways to innovate in the scoring, it's exciting."

Coverage of the American Recovery and Reinvestment Act is supported in part by grants from the William and Flora Hewlett Foundation, at www.hewlett.org, and the Charles Stewart Mott Foundation, at www.mott.org.

Published March 3, 2010, in Education Week

Testing Experts Lay Out Vision for Future Assessments

More-analytical tasks would replace factual recall of multiple-choice.

By Catherine Gewertz
Washington

A group of high-powered policymakers and educators gathered here last week to build support for a new vision of educational assessment that is less a snapshot of students' one-time performance and more like good instruction itself.

Led by Stanford University professor Linda Darling-Hammond, a panel of experts outlined a comprehensive system that includes summative and formative tests of higher-order thinking skills, reflecting a marketplace that they say places increasing value on such skills.

They urged a move away from pages and pages of multiple-choice tests that demand factual recall, and toward the development of a set of deeper, more analytical questions, tasks, and projects that ask students to solve and discuss complex problems. One example is a problem that has been posed to Connecticut high school students: Figure out how to build a statue that could withstand the effects of acid rain, then describe, analyze, and discuss your findings.

Such assessments, Ms. Darling-Hammond said, can be "of, for, and as learning." They can "embody" content standards, she said, not just approximate them. Because teachers would help create and score the assessments, and the assessments would be pegged to good-quality content standards, an aligned teaching-and-learning system would take shape that would help teachers adjust instruction in real time and help district and state administrators plot longer-term education strategy, the experts said.

Common Standards

The portrait of assessment, fleshed out in a paper by Ms. Darling-Hammond that draws on assessment practices in the United States and abroad, was presented at a discussion organized by two Washington-based groups, the National Governors Association and the Council of Chief State School Officers. They have enlisted the support of 48 states to devise common content standards designed to ensure college and career readiness.

The common standards are an "essential" but "inadequate" step toward improving education, said Gene Wilhoit, the CCSSO's executive director, but they must be accompanied by improved assessment, new types of curriculum, and better teacher preparation and professional development.

Dane Linn, who oversees the common-standards work for the NGA, said a vital part of next-generation assessments is the role they must play in learning. "The assessments we end up with have to inform instruction," he said. If they don't change educators' practice, he said, "then what good are they?"

Even though they are still in draft form, the common standards have garnered the support of President Barack Obama, who has offered a better shot at $4 billion in Race to the Top Fund economic-stimulus money to states that embrace them. Last week, the president proposed tying Title I education dollars to adoption of those or other standards validated as rigorous enough to ensure college readiness. *(See Education Week, March 3, 2010)*

A special $350 million pot of Race to the Top Fund money is reserved for the development of common assessments. Six groups, or "consortia," of states, proposing differing approaches to assessment, have formed to compete for that money. In a private meeting after last week's panel discussion, leaders of those consortia met at the CCSSO's office to discuss ways they might work together on summative assessments. *(See Education Week, Feb. 3, 2010)*

In one more potent public symbol of the administration's support for common standards and assessments, the top education adviser in the White House, Roberto Rodriguez, appeared at the panel discussion and urged states to use the $350 million to build "transformative" assessment systems. As Congress begins reconsidering the reauthorization of the Elementary and Secondary Education Act, with the first hearing scheduled last week, Mr. Rodriguez said the administration views college and career readiness as a key objective in that legislation, but that aim requires revamped systems of assessment, professional development, and accountability.

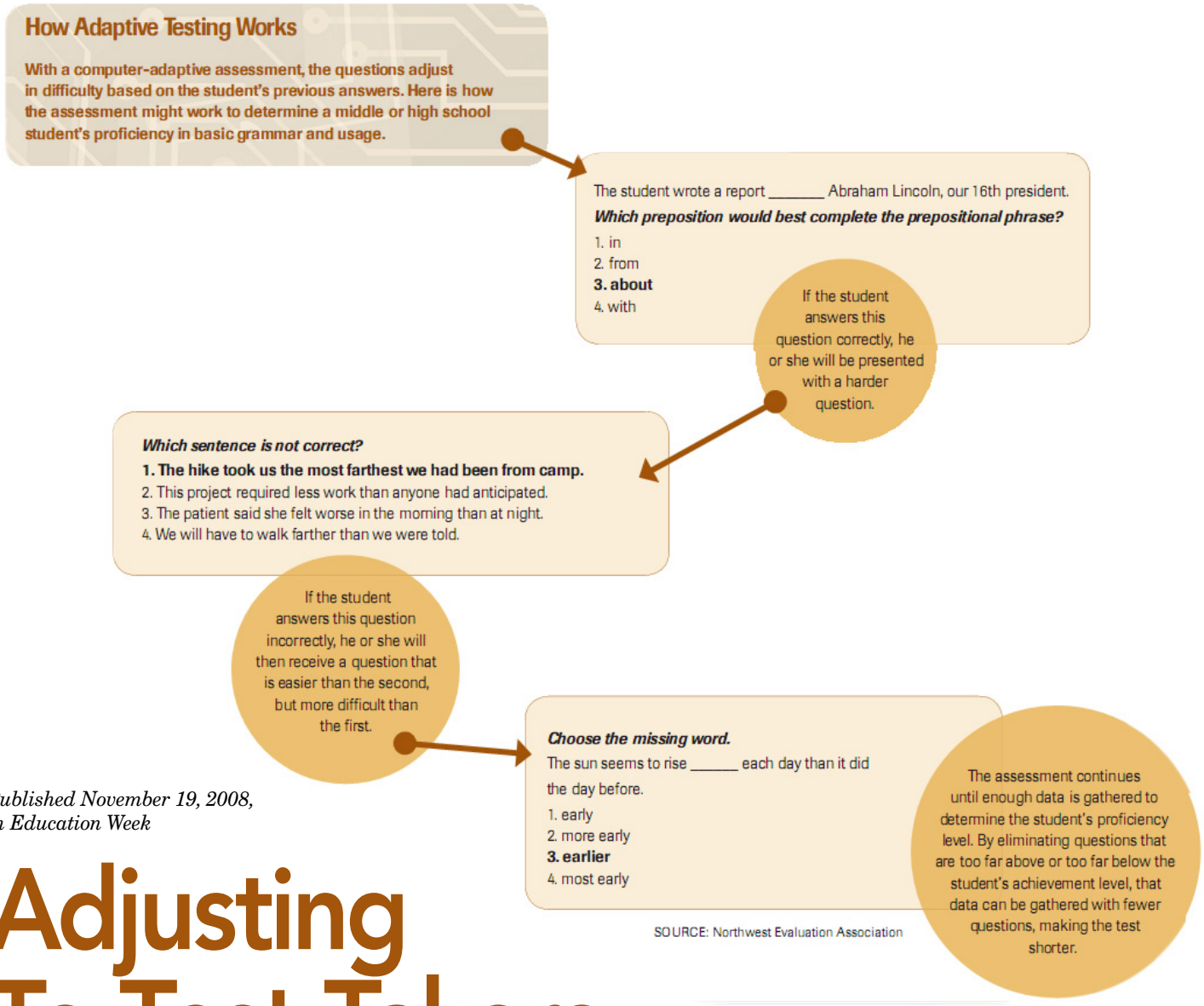
Offering a glimpse of the White House's priorities, he said that a good assessment system will measure individual student growth over time, include multiple measures of achievement, and provide summative information to inform both instruction and state and district policy. It will also integrate results into data systems to guide instruction and be well-integrated with curriculum and professional development.

Inseparable Pieces

Robert L. Linn, a widely respected authority on assessment who spoke on the panel, said that in designing new assessments, it is important to think of them as inseparable parts of systems that include the conception of standards and curriculum. If those are fused, he said, teachers can avoid the worst versions of "teaching to the test" because the tests are actually sound reflections of what the teachers know is important. "The test is bigger and closer to what you care about," said Mr. Linn, a distinguished professor emeritus of education at the University of Colorado-Boulder.

Another member of the panel, Edward Roeber, an adjunct professor of education, measurement, and quantitative methods at Michigan State University's college of education, said new assessments must be paired with revamped teacher preparation. Part of studying to become a teacher must be learning how to use formative assessment in the classroom to guide instruction, and few teachers now receive that training, he said.

Mr. Roeber also addressed a key area of interest among those monitoring the debate about new assessments: the price tag. His work on a soon-to-be-published study will show, he said, that if 30 states work together to design assessment systems that embody the qualities panelists were discussing, they could be crafted for about the same cost as what states spend now on tests used for the current version of the ESEA, the No Child Left Behind Act, a figure Ms. Darling-Hammond put at $1.4 billion per year.



Published November 19, 2008, in Education Week

Adjusting To Test Takers

Computer-adaptive testing addresses individual student needs, but cost and logistical challenges persist.

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.

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 difficult, 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," 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.

> "It increases the information you learn about the student, the student's growth, and how the class is doing."
>
> MATT CHAPMAN
> President and CEO, NWEA

"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."

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

NCLB Constraints

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.

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.

Published September 17, 2008, in Education Week

Test Industry Split Over 'Formative' Assessment

Purpose of Informing Instruction Obscured by Market, Critics Say

By Scott J. Cech

There's a war of sorts going on within the normally staid assessment industry, and it's a war over the definition of a type of assessment that many educators understand in only the sketchiest fashion.

Formative assessments, also known as "classroom assessments," are in some ways easier to define by what they are not. They're not like the long, year-end, state-administered, standardized, No Child Left Behind Act-required exams that testing professionals call "summative." Nor are they like the shorter, middle-of-the-year assessments referred to as "benchmark" or "interim" assessments.

Or they shouldn't be, at least according to experts inside and outside the testing industry, who believe that truly "formative" assessments must blend seamlessly into classroom instruction itself.

"It makes me want to scream and run out of the room," said Ray Wilson, the executive director of assessment and accountability for the 33,000-student Poway Unified School District in Poway, Calif., referring to off-the-shelf commercial products labeled "formative assessment" that major test-makers try to sell him. "I still contend that so long as a teacher doesn't have primary control [over assessment content]," he added, "you will never have truly formative assessment."

"I had not heard, frankly, that there was some resistance to companies selling formative assessment," said Robert Block, the associate vice president for K-12 learning and development at the Princeton, N.J.-based Educational Testing Service, which sells test questions in bulk to schools as "formative assessment item banks."

But he pointed out that school districts' requests for proposals often ask specifically for formative assessments.

"It has become the standard," he said of the testing industry's practice of labeling some assessment products as "formative."

"I'm not sure if it's good or bad—it's just what the market is looking for."

'AssessMints'

The schism was in full view at the Council of Chief State School Officers' National Conference on Student Assessment in Orlando last June. In the main exhibit hall, lined with competing assessment-makers, Peter D. Hofman, the vice president of marketing and business development at the Dover, N.H.-based testing company Measured Progress, handed out boxes of mints labeled "AssessMints" to strolling psychometricians and state-assessment directors.

"These are the only legitimate formative-assessment products," declared Mr. Hofman, whose company does not sell formative-assessment materials.

He left unsaid his implicit message: The room's rival exam firms—like almost every other testing company in the country—sell something that's not what it purports to be.

It's not just semantic consistency that's at stake: Formative assessments are a more than half-billion-dollar business in the United States, according to the latest analysis by Outsell Inc., a Burlingame, Calif.-based research and advisory firm for the information, education, and publishing industries.

That total, including tests, test delivery, scoring, scoring analysis, professional development, and other services, but not such bundled products as textbooks, accounted for about 30 percent of the $2.1 billion in overall assessment revenue generated in the United States in the 2006-07 academic year—the most recent year for which statistics are available.

What's more, said Laurence Bloom, an affiliate analyst for Outsell, the latest estimates project that formative-assessment revenue will climb at a rate of 10 percent to 20 percent per year through 2010, making it the fastest-growing segment in the assessment market.

That's a lot of money being spent on something that experts say can't really be sold—only practiced.

More Than Word Choice

Measured Progress is one of the few large testing companies taking a pass on the lucra-

tive market. That's mostly because Stuart R. Kahl, the company's president and chief executive officer, takes his cues, as many other testing experts do, from a 1998 research-literature review of Paul Black and Dylan Wiliam, then at King's College, University of London. The review, which predated the formative-assessment market, concluded that the research to date showed achievement gains using formative-assessment strategies that were "among the largest ever reported for educational interventions."

"We use the general term 'assessment' to refer to all those activities undertaken by teachers—and by their students in assessing themselves—that provide information to be used as feedback to modify teaching and learning activities," they wrote. "Such assessment becomes formative assessment when the evidence is actually used to adapt the teaching to meet student needs."

Richard J. Stiggins, the executive director of the Portland, Ore.-based Assessment Training Institute, mostly stays away from the term. "There's not universal agreement on what it means—I've actually stopped using the word," said Mr. Stiggins. He said that he now usually refers to formative assessment as "assessment for learning" to cut down on ambiguity.

Mr. Stiggins said he doesn't believe testing companies are shading the truth, but he added, "Merely selling people a test on the assumption that it will be used formatively won't cut it."

In-House Differences

No one whom *Education Week* asked would identify a specific product or company that he or she thought was misusing the term. Many experts were leery of pointing to a company, in part because there are often differences of opinion within a company itself.

"It's really hard to do that because even within companies, there's a wide range," said Mr. Kingston. His own university, for example, uses the word formative when referring to an assessment its staff created.

Another example is Mr. Stiggins. Although he has made the assertion "formative assessment isn't something you buy—it's something you practice" something of a mantra in the many talks he gives to educators around the country, his company is owned by the ETS, which sells item banks under the label "formative."

Within any testing company, said Mr. Kingston, "you have your curriculum experts, you have your psychometric experts, you have your marketing and business people, and they all approach the needs that are out there in different ways."

Mr. Kahl, the CEO of Measured Progress, said he sympathizes—to a point. "I know of individuals within those companies that feel [ambivalent about some products]," he said. "On the the other hand, you look at the [product] catalogs and you see things in there labeled 'formative assessment,' and you've got to wonder who's driving the truck in those big companies."

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

Published June 18, 2010, in Education Week

COMMENTARY

A Better Way to Assess Students and Evaluate Schools

By Monty Neill

Most Americans agree: We need a better way to assess students and evaluate schools. The latest Phi Delta Kappa/Gallup poll found that only one out of four respondents thought the No Child Left Behind law, the current version of the Elementary and Secondary Education Act, had helped schools in their community. Even U.S. Rep. George Miller, D-Calif., an original sponsor of that legislation and the chairman of the House Education and Labor Committee, agrees that NCLB may now be, as he put it, "the most negative brand" in the country.

As state testing intensified under the law and punitive sanctions were imposed, score gains on the National Assessment of Educational Progress slowed or halted for reading and math at all grade levels for almost all groups. Gap closing among demographic groups likewise slowed or stopped. Too much standardized testing damaged learning, particularly for the nation's neediest children. The test-and-punish approach distracted attention from more valuable reforms.

Yet, the underlying problems that propelled passage of NCLB remain. The nation still needs rational and effective approaches to school improvement, including strong curricula, skilled teaching, and equitable opportunities to learn. Society must address the consequences of poverty that undermine learning. Accountability systems and assessments should support high-quality teaching and learning.

Assessment functionally defines what we value in learning. As the old saying goes, "What you test is what you teach." With curriculum and instruction, it is a necessary component of the learning process. Assessment and evaluation inform the community about attainment of goals, including those beyond academic outcomes. They signal problems that must be addressed and provide information on how to improve.

A healthy assessment and evaluation system would include three key components: limited large-scale standardized testing; extensive school-based evidence of learning; and a school-quality-review process.

Large-scale tests. When it comes to assessment, the United States is an international outlier. As Stanford University's Linda Darling-Hammond has shown, many nations with better and more equitable educational outcomes test far less than we do. They typically test just one to three times before high school graduation, and use multiple-choice questions sparingly, if at all. Excessive testing wastes educational resources and fosters the use of cheap, low-level tests, while adding high stakes narrows and dumbs down the curriculum. The results provide little instructional value to students, teachers, schools, or districts.

Higher-quality tests would help. But based on the U.S. Department of Education's published criteria for awarding the $350 million it will give to state consortia for test development, only modest improvements are likely to come from that program, far less than the qualitative leap schools need. Tests will continue to be administered too frequently.

Congress should return to the requirements of the 1994 version of the ESEA to test once each in elementary, middle, and high school. This would bring the United States in line with other nations, while freeing up re-

sources for new assessment and evaluation approaches.

Local and classroom evidence of learning. The primary public source of data about student achievement should be the work students do in the classroom. That kind of evidence reveals the range, depth, and quality of student learning. The United States has avoided taking this path, however, trekking instead through the wastelands of high-stakes standardized testing. This is largely because authorities have distrusted and not been willing to invest in teachers, unlike more successful nations, such as Finland. The pending ESEA reauthorization brings with it the chance to change direction and avoid another lost decade.

Classroom-based assessment by skilled teachers is of great value. Teachers assess frequently, but research shows that many have limited assessment skills. Thus, they need ongoing training to develop their assessment capabilities. In places as disparate as Nebraska, with its former STARS program of local, state-approved assessments, and New York state, where the New York Performance Standards Consortium replaces state tests with a mix of school- and consortium-based performance assessments, attention to assessment has been contributing to improved teaching, forging a stronger community of educators, and producing improved results by a variety of measures from independent exams to college enrollment and success.

Classroom-based assessments can be adapted to students' varying needs while maintaining high standards. Assessing extended work, such as research projects, far more readily ensures evaluation of higher-order thinking skills than can large-scale standardized exams.

Of course, teachers cannot create every high-quality assessment they need. States should gather tasks that have been approved by skilled educators into "libraries" which teachers can access as they need. Using already-vetted instruments will contribute to ensuring the quality of classroom-based evidence of student learning.

In this country and around the world, a wide range of classroom- and school-based evidence—from exams, projects, "learning records," and portfolios—is audited and moderated. Essentially, a random sample from each classroom is rescored by trained readers to verify a teacher's initial scoring. This produces useful feedback to the originating teacher, score adjustments where needed, and professional development for the readers. Research in other nations and in this country shows that this process can be done with a degree of consistency more than sufficient for statewide comparability. What is standardized is not individual student work but the criteria for gathering and evaluating work products.

Schools would produce an annual report, including evidence of educational successes and ongoing problems, along with improvement plans. Documentation of student learning across the curriculum would then become publicly available. Such reports could be discussed by the school's community and reviewed by higher governmental authorities.

School-quality reviews. Often called "inspectorates," these are the central tool for school evaluation in places such as England (which tests at a few grades), Wales (which tests only at grade 5, with no stakes), and New Zealand (which has only a NAEP-like national exam). Clearly, this is a very different mind-set: Instead of test results, the core of evaluation is a comprehensive review every four to five years covering the range of attributes parents and communities want for their schools. School-quality reviews have been proposed by the politically diverse signers of the Broader, Bolder Agenda. In the United States, these quality reviews would be complemented by limited large-scale testing and annual school reports, providing comprehensive evidence in which each component serves as a check on the others.

During inspections, skilled professionals, perhaps accompanied by parents and community members, conduct three- to five-day visits. The teams come prepared with other data (assessment results, graduation rates, school-climate surveys, opportunity-to-learn information, and so forth). They sit in on classes, review student work, and interview students, teachers, and other staff members. They prepare a draft report and discuss it with school personnel. The final report is a public document that includes an evaluation and recommendations for improvement. This approach is similar to college and school accreditation processes.

Schools with severe problems would be reviewed more frequently. States could specify how and when recommendations become mandates, some of which could require new resources, outside assistance, or strong interventions.

Since nations using a more balanced, comprehensive, improvement-focused assessment and evaluation system have produced better educational results with fewer harmful side effects, it makes good sense to restructure the current test-based U.S. system. The model outlined here can provide better assessment, comparability, and accountability. These improvements are needed by all schools, especially those which primarily serve low-income children.

> "Without a healthy assessment and evaluation, the reform enterprise will fail again."

Comprehensive data analysis can identify educational problems and solutions. Equitably distributed resources, strong collaborative leaders, professional learning communities of teachers, rich and challenging curriculum, strong parental involvement, and safe learning environments are also necessary.

But without healthy assessment and evaluation, the reform enterprise will fail again.

Monty Neill is the interim executive director of FairTest, the National Center for Fair & Open Testing, in Boston. FairTest developed this proposal with help from allies, particularly the Massachusetts Coalition for Authentic Reform in Education.

Published February 11, 2010, in Education Week

COMMENTARY

A Seamless System of Assessments

By Robert Rothman

Is testing a waste of time? Teachers seem to think so. In a 2006 survey, 71 percent of them said that students took too many standardized tests, and 62 percent called testing a "necessary evil."

Yet when Oregon introduced its online testing system, which allows students to take the tests up to three times a year, teachers embraced it. They apparently did not think the testing burden was either excessive or evil.

Why? Because the Oregon test delivers near-instantaneous results that show teachers how students perform on particular content strands, such as geometry or measurement. Teachers can use the information in real time to adjust instruction and devote additional attention to students in areas where they need help. For these teachers, tests are not separate from instruction; they are integral to it.

By contrast, the accountability tests most states use, which loom increasingly large, provide little useful information to teachers. The results often come back too late, in many cases after the class has left for the summer. And the results often tell little about students' strengths and weaknesses, much less what instructional strategies teachers should employ to raise performance.

Accountability tests can provide important information, but because of their outsized influence in schools, they have been asked to do too much. We expect them to inform students and parents about academic progress, teachers about what to do to improve instruction, community members about school success, and states and the nation about whether students and schools are meeting goals. These expectations, all important, are more than any assessment—even the most sophisticated—can bear. And most of the tests now in use are far from sophisticated.

Fortunately, there is a growing chorus suggesting that states and the federal government reduce their reliance on accountability tests as the primary measure of student and school performance and instead develop comprehensive assessment systems, based on standards, that include a variety of measures that can effectively serve all of the purposes we want tests to serve. These measures won't simply mean more testing; they can mean better information that can lead to improvements in teaching and learning.

The emphasis on assessment systems is fueled in part by a confluence of factors that make it more likely that these visions can become reality in the next few years. The state-led effort to develop common-core standards for college and career readiness will invariably be followed by new assessments that measure progress toward those standards. *(See Education Week, April 21, 2010)* These assessments are likely to include components, such as curriculum-embedded projects, that measure competencies traditional accountability tests do not tap.

At the same time, the U.S. Department of Education has launched a $350 million competition for consortia of states to develop assessments to measure the common-core standards. The designs that are being suggested, while in their preliminary stages, represent bold departures from current practice.

Finally, advances in technology make possible new types of assessment formats, such as simulations, that were not available a decade ago, and also make the use of existing formats more feasible. Technology also facilitates the development of information systems that can link assessments and assessment results more closely to instructional tools to support teachers, and directly to the classroom.

As a result, what we might be seeing in the next few years is a sophisticated information system that informs students and parents about progress toward standards for college and career readiness; teachers about the knowledge and skills students can demonstrate and what they need to work on (as well as about what they themselves need to do to address those areas); and communities and states about school progress toward standards. And, if the measures are done right, the assessments themselves can support learning by modeling the kinds of activities students need to be able to know how to perform.

To envision how such a system might shift school practice, consider what has happened in the retail industry. In the past, retail stores would close their doors for a day each year to take inventory. Now, thanks to the accurate and instantaneous information bar codes allow, retailers can keep track of their inventory in real time, 365 days a year. This is not to say that students are commercial products, or that we want to slap bar codes on their foreheads. But a comprehensive assessment system could provide continuous, coherent, and high-quality information on student performance that teachers, school leaders, and district and state administrators could use to improve teaching and learning.

In such a system, assessment is neither excessive nor evil. Nor is it a waste of time. On the contrary, assessment—and the information it provides—is a vital tool to improve instruction, learning, and school practice.

Robert Rothman is a senior fellow at the Alliance for Excellent Education, in Washington.

Published February 24, 2010, in Education Week

COMMENTARY

Next-Generation Assessment Systems

By Stanley N. Rabinowitz

An unprecedented confluence of factors—economic, political, and educational—is causing many states to rethink their student-assessment programs. But careful thought and expert guidance will be needed if they are to avoid the problems of the past and take advantage of promising new developments.

Most state assessment programs, regardless of their history or goals, were revised early in the last decade to meet the requirements of the federal No Child Left Behind Act: All assess reading/language arts and mathematics in grades 3-8 and high school, and science in elementary, middle, and high school. Collectively, they are increasingly homogeneous, largely multiple-choice, with some sprinkling of constructed-response and direct writing.

Change is in the offing. The upcoming reauthorization of the Elementary and Secondary Education Act, of which No Child Left Behind is the current version, is likely to result in changes to assessment requirements, allowing more flexibility and providing greater support for innovative assessment models. Dissatisfaction with current limited options, coupled with the Common Core State Standards Initiative's potential impact—assessments covering fewer but clearer objectives, and advancing more-rigorous content and skills requirements—will necessitate a broad reconceptualization of assessment. This inevitably will mean a shift away from state standardized testing as the only game in town, and a move toward the development of innovative state assessment systems.

This next-generation model will include differentiated roles for assessment at the federal, state, and school levels; the use of multiple measures; and assessments that support accountability programs focused on both growth and current status. It also will be likely to take greater advantage of technology, and will benefit from U.S. Department of Education initiatives and dollars, represented by the Obama administration's Race to the Top Fund and other grant programs.

What follows is an overview of key concepts states should consider as they move forward.

Multiple Measures. The growing call for more performance-based assessment is reminiscent of past practice, when programs such as the New Standards Project and the Vermont and Kentucky writing and math portfolios of the early 1990s seemed to usher in a new era. But significant changes in the political context of accountability, along with technological advances such as the use of computer simulations, should encourage states to look to the future, not the past.

To do this, we need to create a new vision of technical evidence for performance-based assessments, building on advances in such work as alternative assessments for students with disabilities. And we must ensure that teachers get the necessary training to incorporate performance-based instruction into their teaching, and to use classroom-embedded assessments to diagnose students' strengths and weaknesses.

Balance. When researchers call for the development of a balanced assessment system, they typically mean balancing roles and responsibilities. But other factors also must be balanced:

• *Cost*. What is the most cost-efficient way to get teachers, students, administrators, policymakers, and parents the information they need about individual students and the system itself?

• *Constraints.* How do we overcome technical limitations in the next-generation assessments and properly train teachers and others to develop and use new techniques and understandings?

• *Risk*. How do we support true reform without overburdening students, teachers, and other key constituencies?

Complementary Components. What constitutes the "system" in our next-generation assessment system? Having a system means there are several assessment components, each playing a complementary, not duplicative, role. Our system would have both formative (diagnostic, classroom-embedded) and summative (used for accountability, mostly on-demand) assessments. Each type would be designated for its primary role. And having both would allow teachers to use formative assessments honestly: to focus on how well students are advancing, without worrying that short-term deficiencies will affect their own evaluations or reflect badly on their students.

Valuing efficiency, our ideal system would include multiple-choice, constructed-response, and performance tasks, as appropriate, to best measure the knowledge and skills called for in challenging new content standards. Indicators would need to justify these various options' inclusion: What is the incremental validity of each component (in other words, what would be lost if it were excluded)?

Roles and Responsibilities. In our system, each governing level would play a different primary role. The federal role would be limited but targeted, ensuring that all students achieved sufficient levels of reading, math, and science proficiency. States would build on that foundation by adding other content areas, such as social studies, and including additional indicators—those based, for example, on local economic conditions, resources, and values and beliefs. The primary local-level function would be curriculum and instruction, ensuring that all teachers are prepared to meet, and are supported in meeting, the needs of all students, including those most at risk, such as English-language learners, students with disabilities, and stu-

dents living in poverty.

Assessments at each level would be consistent with that level's "system" responsibilities. One beneficial outcome of having differentiated roles with specialized assessments would be that the technical requirements could vary based on the purpose and the stakes, allowing different types of validity evidence, for example, to drive local diagnostic tasks, as opposed to state and federal accountability assessments.

Choice. While most assessment and accountability systems expect all students to master all content at a "major" level, regardless of students' interests or abilities, our system would incorporate an element of choice at the state, school, and student levels. Students could pick the content area or areas on which they wished to focus deeply, for example, and would be assessed accordingly, so long as they also demonstrated sufficient knowledge in other areas to support responsible citizenship and employment.

Some states have already moved in this direction by supporting magnet and charter schools, adding new content areas and indicators to state assessment and accountability programs, and requiring career "majors" for students at the high school level.

The bottom line when it comes to choice is that all schools and students should be great at something that society values, and sufficiently good at all else, to maximize motivation and success while keeping students' options open for later-developing needs and interests.

Too often, assessment-reform plans are dismissed from the start because of the identification of what are thought to be barriers to their development and implementation. The real world certainly has real constraints. But before we say "we can't," we should ask ourselves the following questions:

- Is the barrier real, or is it just perceived?
- Is it real, or is the change just going against the traditional approach?
- Is it real, or is the change just difficult?
- Is it real, or does the change just require new technical tools?
- Is it real, or is the change just expensive?

We may find surprising opportunities if we dare to address perceived barriers to assessment reform honestly.

Our assessment and accountability systems should reflect what we value most for our students, schools, and society, and what we think it means to be a well-prepared student, worker, and citizen. Once these are clear, we should be willing to fight and to pay for their reflection in our system for measuring academic progress.

As states begin this important endeavor, they should take the following steps:

- Develop a vision statement that incorporates the values the system will represent.
- Devise an implementation plan with goals, key dates, milestones, responsibilities, and necessary resources.
- Secure sufficient funding to implement the plan.
- Develop a system for evaluation and feedback.

There is no time to waste. Much needs to be done, and the quality of American education is at stake.

Stanley N. Rabinowitz directs the Assessment and Standards Development Services program at WestEd, which manages the federal Assessment and Accountability Comprehensive Center. He also serves on the validation committee for the Common Core State Standards Initiative. This essay is adapted from his closing remarks at the 2009 Reidy Interactive Lecture Series.

Published July 24, 2007, in Education Week

COMMENTARY

IN TESTING, the Infrastructure Is Buckling

By Thomas Toch

While most public school students enjoy the idle days of summer, the nation's testing companies are working around the clock to help states get the results of millions of standardized state tests to parents before the start of the new school year, a deadline under the federal No Child Left Behind Act that many states may not make.

The tests are the linchpin of Washington's efforts to promote higher standards in public education by cracking down on schools where students don't measure up. But No Child Left Behind is overwhelming the nation's testing infrastructure, and the result has been troubling: Instead of encouraging schools to raise the level of rigor in classrooms, the law is giving them powerful incentives to do just the opposite.

The testing system is beset by a host of problems: a shortage of the experts who ensure test quality, intense competition among testing companies that has led to below-cost bidding, underfunded state testing agencies, and the sheer scale of the NCLB testing requirements.

Together, 23 states added more than 11 million tests in the 2005-06 school year to comply with the law, pushing the total number of NCLB tests to 45 million. Test booklets have to be sent to and collected from nearly every public school in the country, and the results scored and reported back to the parents of every tested student under super-tight NCLB timelines—a massive logistical challenge.

Evidence that the system is buckling under this pressure isn't hard to find: Beset by misprinted tests, faulty student information, scoring glitches, and other troubles, Illinois earlier this year released its 2006 No Child Left Behind results just days before students sat for the state's 2007 tests; more recently, Florida announced that it had misreported the results of 200,000 reading tests.

But arguably the most damaging consequence of the testing crisis has taken place off the public stage: The problems plaguing testing have led states to gravitate to tests under the No Child Left Behind law that mainly measure low-level skills. They are using tests with a surfeit of questions that require students to merely recall or restate facts rather than do more demanding tasks like applying or evaluating information, because such tests are cheaper and faster to produce, give to students, and score.

The problem is that these dumbed-down tests encourage teachers to make the same low-level skills the priority in their classrooms, at the expense of the higher standards that the federal law has sought to promote. Teachers and principals are rational people. If their reputations, and even their jobs, are tied to their students' test scores, as is true under No Child Left Behind, they are going to feel tremendous pressure to stress the rote skills that the exams test most often.

Testing-industry leaders say that states are backing away from or abandoning outright open-ended questions, which stretch students by requiring them to produce their own answers, because they are more costly and more time-consuming to use than multiple-choice questions. As a result, close to half the students tested under NCLB nationwide in the just-completed school year saw only multiple-choice questions.

In addition to lowering teachers' sights for their students, such tests produce an inflated sense of student achievement. Scores on reading tests that measure mostly literal comprehension are going to be higher than those on tests with a lot of questions that measure whether students can make inferences from what they read.

The same is true in math. In a study by the University of Colorado at Boulder testing expert Lorrie Shepard, 85 percent of 3rd graders who had been drilled in computation for a standardized test picked the right answer to the problem 3 x 4 = ___, but only 55 percent answered correctly when presented with three rows of four X's and asked how many that represented.

> "Despite testing's tremendous importance to school reform, under the law states typically spend about one-quarter of 1 percent of combined federal, state, and local school revenues on their statewide testing programs, or about $20 of the more than $8,000 spent per student."

Workforce experts, of course, say American students will need higher-level skills to compete successfully for good jobs in the new global economy.

By the measure of how much money states spend on No Child Left Behind testing, it's hardly surprising that we're getting simple-minded tests. Despite testing's tremendous importance to school reform, under the law states typically spend about one-quarter of 1 percent of combined federal, state, and local school revenues on their statewide testing programs, or about $20 of the more than $8,000 spent per student.

Next year, things are likely to be worse, when states have to administer another 11 million standardized tests after an NCLB science-testing requirement goes into effect.

But so far, U.S. Secretary of Education Margaret Spellings has sidestepped the testing problem. Testing under the law is a state issue, she has said, and ensuring that tests measure high-level skills goes "beyond what was contemplated by NCLB."

But the Bush administration can't have it both ways. It can't say it wants high standards for all students and then sit on its hands when it becomes clear that a key part of the No Child Left Behind reform plan is working against that goal.

Thomas Toch is a co-director of Education Sector, a Washington-based think tank.

Published October 17, 2007, in Education Week

COMMENTARY

Five Assessment Myths and Their Consequences

By Rick Stiggins

America has spent 60 years building layer upon layer of district, state, national, and international assessments at immense cost—and with little evidence that our assessment practices have improved learning. True, testing data have revealed achievement problems. But revealing problems and helping fix them are two entirely different things.

As a member of the measurement community, I find this legacy very discouraging. It causes me to reflect deeply on my role and function. Are we helping students and teachers with our assessment practices, or contributing to their problems?

My reflections have brought me to the conclusion that assessment's impact on the improvement of schools has been severely limited by several widespread but erroneous beliefs about what role it ought to play. Here are five of the most problematic of these assessment myths:

Myth 1: The path to school improvement is paved with standardized tests.

Evidence of the strength of this belief is seen in the evolution, intensity, and immense investment in our large-scale testing programs. We have been ranking states on the basis of average college-admission-test scores since the 1950s, comparing schools based on districtwide testing since the 1960s, comparing districts based on state assessments since the 1970s, comparing states based on national assessment since the 1980s, and comparing nations on the basis of international assessments since the 1990s. Have schools improved as a result?

The problem is that once-a-year assessments have never been able to meet the information needs of the decisionmakers who contribute the most to determining the effectiveness of schools: students and teachers, who make such decisions every three to four minutes. The brief history of our investment in testing outlined above includes no reference to day-to-day classroom assessment, which represents 99.9 percent of the assessments in a student's school life. We have almost completely neglected classroom assessment in our obsession with standardized testing. Had we not, our path to school improvement would have been far more productive.

Myth 2: School and community leaders know how to use assessment to improve schools.

Over the decades, very few educational leaders have been trained to understand what standardized tests measure, how they relate to the local curriculum, what the scores mean, how to use them, or, indeed, whether better instruction can influence scores. Beyond this, we in the measurement community have narrowed our role to maximizing the efficiency and accuracy of high-stakes testing, paying little attention to the day-to-day impact of test scores on teachers or learners in the classroom.

Many in the business community believe that we get better schools by comparing them based on annual test scores, and then rewarding or punishing them. They do not understand the negative impact on students and teachers in struggling schools that continuously lose in such competition. Politicians at all levels believe that if a little intimidation doesn't work, a lot of intimidation will, and assessment has been used to increase anxiety. They too misunderstand the implications for struggling schools and learners.

Myth 3: Teachers are trained to assess productively.

Teachers can spend a quarter or more of their professional time involved in assessment-related activities. If they assess accurately and use results effectively, their students can prosper. Administrators, too, use assessment to make crucial curriculum and resource-allocation decisions that can improve school quality.

Given the critically important roles of assessment, it is no surprise that Americans believe teachers are thoroughly trained to assess accurately and use assessment productively. In fact, teachers typically have not been given the opportunity to learn these things during preservice preparation or while they are teaching. This has been the case for decades. And lest we believe that teachers can turn to their principals or other district leaders for help in learning about sound assessment practices, let it be known that relevant, helpful assessment training is rarely included in leadership-preparation programs either.

Myth 4: Adult decisions drive school effectiveness.

We assess to inform instructional decisions. Annual tests inform annual decisions made by school leaders. Interim tests used formatively permit faculty teams to fine-tune programs. Classroom assessment helps teachers know what comes next in learning, or what grades go on report cards. In all cases, the assessment results inform the grown-ups who run the system.

But there are other data-based instructional decisionmakers present in classrooms whose influence over learning success is greater than that of the adults. I refer, of course, to students. Nowhere in our 60-year assessment legacy do we find reference to students as assessment users and instructional decisionmakers. But, in fact, they interpret the feedback we give them to decide whether they have hope of future success, whether the learning is worth the energy it will take to attain it, and whether to keep trying. If students conclude that there is no hope, it doesn't matter what the adults decide. Learning stops. The most valid and reliable "high stakes" test, if it causes students to give up in hopelessness, cannot be regarded as productive. It does more harm than good.

Myth 5: Grades and test scores maximize student motivation and learning.

Most of us grew up in schools that left lots of students behind. By the end of high school, we were ranked based on achievement. There were winners and losers. Some rode winning streaks to confident, successful life trajectories, while others failed early and often, found recovery increasingly difficult, and ultimately gave up. After 13 years, a quarter of us had dropped out and the rest were dependably ranked. Schools operated on the belief that if I fail you or threaten to do so, it will cause you to try harder. This was only true for those who felt in control of the success contingencies. For the others, chronic failure resulted, and the intimidation minimized their learning. True hopelessness always trumps pressure to learn.

Society has changed the mission of its schools to "leave no child behind." We want all students to meet state standards. This requires that all students believe they can succeed. Frequent success and infrequent failure must pave the path to optimism. This represents a fundamental redefinition of productive assessment dynamics.

Classroom-assessment researchers have discovered how to assess for learning to accomplish this. Assessment for learning (as opposed to of learning) has a profoundly positive impact on achievement, especially for struggling learners, as has been verified through rigorous scientific research conducted around the world. But, again, our educators have never been given the opportunity to learn about it.

Sound assessment is not something to be practiced once a year. As we look to the future, we must balance annual, interim or benchmark, and classroom assessment. Only then will we meet the critically important information needs of all instructional decisionmakers. We must build a long-missing foundation of assessment literacy at all levels of the system, so that we know how to assess accurately and use results productively. This will require an unprecedented investment in professional learning both at the preservice and in-service levels for teachers and administrators, and for policymakers as well.

Of greatest importance, however, is that we acknowledge the key role of the learner in the assessment-learning connection. We must begin to use classroom assessment to help all students experience continuous success and come to believe in themselves as learners.

Rick Stiggins is the founder of the Educational Testing Service's Assessment Training Institute, in Portland, Ore.

Published by Editorial Projects in Education, Inc.
6935 Arlington Road, Suite 100
Bethesda, MD, 20814
Phone: (301) 280-3100
www.edweek.org

WEB LINKS

EDUCATION WEEK RESOURCES

Assessment Resources

NOW FEATURING INTERACTIVE HYPERLINKS.
Just click on your website and go.

Assessment and Accountability Comprehensive Center
http://www.aacompcenter.org/

Common Core State Standards Initiative
http://www.corestandards.org/the-standards

CRESST – National Center for Research on Evaluation, Standards, and Student Testing
http://www.cse.ucla.edu/

MAP -- Measures of Academic Progress
http://www.nwea.org/products-services/computer-based-adaptive-assessments/map

National Center for Improvement of Educational Assessment
http://www.nciea.org/

Northwest Evaluation Association
http://www.nwea.org/

EDUCATION WEEK

SPOTLIGHT

View the complete collection of Education Week Spotlights, including:

- STEM in Schools
- ELL Assessment and Teaching
- Reading Instruction
- Response to Intervention
- Technology in the Classroom
- and more!

@www.edweek.org/go/spotlights