The eight-year-old No Child Left Behind Act established for the first time a federal benchmark for student achievement. When the Obama Administration took office last year, the new president promised to stay true to the goals of NCLB while upgrading what critics have termed simplistic, “fill in a bubble” testing to create a more comprehensive assessment of student learning.

“I am calling on our nation’s governors and state education chiefs to develop standards and assessments that don’t simply measure whether students can fill in a bubble on a test,” President Obama said in March 2009, “but whether they possess 21st century skills like problem-solving and critical thinking, entrepreneurship and creativity.”

In April of this year, U.S. Secretary of Education Arne Duncan set this process into motion by announcing that $350 million in funding under the Race to the Top program, established last year as part of the American Recovery and Reinvestment Act, would be used to fund the efforts of multi-state consortia to create a new generation of assessments. “States are leading the way in creating new standards designed to ensure that students graduate from high school ready for success in colleges”
Careers,” Duncan said at the time. “To fully realize this vision, states need new assessments that measure a broader range of students’ knowledge and skills.” The Department of Education issued a Request for Proposal calling for assessments that measure rigorous, globally competitive and consistent standards; provide accurate information about both current achievement and year-to-year growth; inspire great teaching by reflecting and supporting instructional best practices; and include all students, such as English-language learners and those with disabilities. Critics say the tests to date under No Child Left Behind have measured only rote memorization of facts and figures through their multiple choice questions and do not adequately track year-to-year growth or inspire teachers to do anything more than teach-to-the-test.

Federal Funding Categories
The department has divided its funding streams into two categories. Category A, which requires a minimum of 15 states per application, will award $150 million or more for assessment systems that measure against college- and career-ready benchmarks in English and math annually from grades 3 through 8, and at least once in high school, says Anne Whalen, senior adviser to Duncan. The number of states involved should help achieve a better cost for each state, encourage collaborations and spread best practices. In addition, the Category A funding may be used to develop formative interim assessments, develop a technology platform the assessments work off of (to develop, administer, score and report students’ work) and provide professional development for school leaders, Whalen says.

Category B funding, which includes up to $30 million and requires a minimum of five states per application, will be strictly for developing assessments around high school courses. No specific subjects are required, Whalen says, although the department will be taking a hard look at the scalability of applications. “We’re soliciting people to think about assessment systems—not necessarily just one test at the end of the year,” she says. “These tests must be able to assess complex demonstrations of student knowledge. A lot of the assessments that many states are administering today are predominantly multiple choice.” The department held 10 meetings between November and January in Boston, Atlanta, Denver and Baltimore/Washington, D.C. to solicit input from both the public and more than 40 assessment practitioners and researchers. The application package was due in June, and the department anticipates making awards in September to multi-state consortia, designed that way to spread best practices and to gain cost savings from using more states to help spread the cost, Whalen says.

Who Has Applied?
As of early June, two consortia, called the Smarter Balanced Assessment Consortium (with 35 states) and the Program for Assessment of College and Career Readiness (with 27 states, some overlapping with the first consortium) had planned to apply for Category A funding. A group of eight states led by the National Center on Education and the Economy planned to apply for Category B funding, and there could be others.

The department does not have to choose a single “winner” but could award funds to two or more consortia within the two categories, Whalen says. Although the state consortia are the only eligible applicants, they might contract with testing companies to help develop the granular details of the assessments. Monty Neill, interim executive director of Fair Test and chair of the Forum on Educational Accountability, expects they will. “[Testing companies] are set up to deliver these products to schools and to collect them,” he says. “Every test company is going to want to be involved, so they can continue to get contracts.”

The Smarter Balanced Assessment Consortium is hashing through how to

“No high-achieving nation tests every child, every year, in the way we’re currently doing.”

—Linda Darling-Hammond, professor, Stanford University
develop more open-ended assessments that still could be at least partially graded by computer and how best to include performance assessments for teachers and schools, according to Linda Darling-Hammond, a Stanford University professor who is also an adviser to the consortium. Darling-Hammond says more ambitious assessments carry additional costs.

“No high-achieving nation tests every child, every year, in the way we’re currently doing,” she says. “They have much more intellectually ambitious assessments [or measuring not just memory but what students can do with knowledge]. The groups are trying to look for creative ways to resolve those tensions.”

**Progress or Regress?**
Darling-Hammond sees the potential for the consortia to draw from “the leading edge of practice”—expecting of students more tasks that require research and inquiry and tasks that require them to express ideas—in terms of both designing the assessments and in using computer-scoring on open-ended items where the technology is workable and the subject matter lends itself, typically in math and science more than English and social studies. “There will certainly be more commonality among states, as well as more opportunity to compare results,” Darling-Hammond adds.

But Darling-Hammond sounds a note of caution that if such practices are not strongly infused into the new assessments, the process could result in a “reduction to the mean” in terms of the sophistication of testing methodology that would hold some states back from pursuing more comprehensive assessments. “There’s the possibility for progress and the possibility for regress,” she says. “You may be able to do more adventurous, more thoughtful, more innovative assessments less expensively. There are big savings when states collaborate. The downside is, you don’t want to hold back anyone from innovating further.”

Michael Cohen, president of education policy group Achieve, says the RFP under Race to the Top presents a “game-changing opportunity.” A consultant to Program for Assessment of College and Career Readiness, Cohen sees a rare

---

**Maps to Develop Sophisticated Tests**

AS THE UNITED STATES WORKS TO DEVELOP WIDER, MORE COMPREHENSIVE AND sophisticated testing assessment systems, several countries that have gone down this road provide maps on how to do so.

In a white paper titled “Benchmarking Learning Systems: Student Performance Assessment in International Context,” Stanford University professor Linda Darling-Hammond explores testing systems in Finland, Sweden, Hong Kong, Singapore, Australia and the United Kingdom.

“While the use of performance tasks does require time and expertise, educators and policymakers in high-achieving nations believe that the value of rich performance assessments far outweighs their cost,” she writes. “These [new assessments] deeply engage teachers and students in learning, make rigorous and cognitively demanding instruction commonplace, and, leaders argue, increase students’ achievement levels and readiness for college and careers.”

Some larger nations like Canada, Australia and China have developed state- or provincial-level standards and blend state and local assessments.

This methodology “turns out to be an important way of enabling strong teacher participation and ensuring high-quality local assessments that can be moderated to ensure consistency in scoring,” Darling-Hammond writes. “In many cases, school-based assessments complement centralized ‘on-demand’ tests and may constitute up to 60 percent of the final examination score.”

Among other examples, Hong Kong uses school-based assessments that Darling-Hammond’s report describes as typically involving students in activities such as “making oral presentations, developing a portfolio of work, undertaking fieldwork, carrying out an investigation, doing practical laboratory work or completing a design project.”

These methodologies “help students to acquire important skills, knowledge and work habits that cannot readily be assessed or promoted through paper- and pencil-testing,” she adds, and later in the report stresses “the importance of assessment of, for and as learning, rather than as a separate disjointed element of the education enterprise.”

Students at New Tech High School in Sacramento, Calif., are engaged in the kind of challenging work, such as hands-on learning and higher-order thinking skills, that performance assessments support.
combination of “money on the table, lots of research to draw upon, and states that have demonstrated a commitment to and ability to work together.”

“This is a rare opportunity to … move to a next-generation assessment system that will still serve the accountability purposes that lawmakers will have in mind, but at the same time will provide data that is useful for teachers to improve instruction as well as signals on what kind of performances we expect,” he says. “You see states all over the map in terms of expectations for students. This is an opportunity to change that by building assessments that are common among large numbers of states.” The effort will face a number of challenges, Cohen acknowledges. It’s complicated, and states could have different ideas on everything from the broad architecture of the assessment to how test forms should be visually designed. “To do this on a large scale, with a lot of different states involved, in a four-year time period, with tests ready to go and useful for accountability purposes—you’ve got to be very thoughtful about how that’s done,” he says. “It will take vision and courage on the part of state leaders.”

**Engaging Teachers**

The process also will benefit greatly from front-end input from teachers and their unions to get them on board, Cohen believes, so they see the value of professional development. Teachers and unions are engaged in discussions on design and development of the tests.

The department’s RFP does not require such collaboration, Cohen says, but “it’s important, politically, to start out with teachers understanding and thinking well of what’s going on; otherwise, they’ll be a good source of opposition. It’s good educational practice, and it’s good politics.”

Such collaboration should overcome the understandable reluctance if all you do is say to teachers, ‘Great news, we’ve got more tests for you,’ ” Cohen adds. “A lot of [the teachers’] reaction] is going to depend on what tools we provide, whether or not we provide formative assessments with real-time information, sample lessons and sample instructions. All of that will be important to do.”

The trick will be subtly integrating the new assessments into the school curriculum rather than stopping the curriculum.

“This is a rare opportunity to move to a next-generation assessment that will deliver accountability that lawmakers want and also provide data that are useful for teachers to improve instruction.”

—Monty Neill, interim executive director, Fair Test

Race to the Top RFP provides adequate resources to support this broader range of assessments and is concerned that test scores will continue to be overemphasized. Instead of the multiple and more creative assessment methods that Cohen hopes to see, Fair Test is concerned the end result could be a change from one big test to a plethora of little quizzes, Neill says.

“We think that would be very likely instructionally deadly,” he says. “We’ve heard stories about teachers who are mandated as much as weekly to give a test to the kids on a computer. The kids hate it, and it’s not instructionally helpful.”

The department’s RFP does not adequately describe what’s meant by formative evaluations, Neill says, which could open up the process to misuse. “What we’re seeing is mini-knockoff standardized tests of terribly low quality that are being used and then being labeled as formative,” he says.

Neill adds that Fair Test has seen “anecdotal evidence” of such practices in places like Los Angeles, Boston, Chicago and New York. “It’s not always system-wide; sometimes schools are choosing to do it,” he says. “We don’t have the resources to investigate this as well as we would like.”

The department will need to trust teachers to evaluate open-ended portions of a comprehensive assessment, particularly in a subject like history, Neill says, and then spot-check to ensure teachers are within standardized grading guidelines and provide remediation if not. “Historians themselves disagree. Kids have to learn to figure out the evidence and apply it,” he says. “I would be surprised if there’s a computer program able to do that.”

Instead teachers would have to be directly involved in grading standardized tests rather than relying solely on a computer. “This raises the issue, are people ready to trust teachers?” Neill says. “There’s no way around this. We’re going to have to trust them to teach.”

*Ed Finkel is a freelance writer based in Evanston, Ill.*