

How will the Common Core Initiative Impact the Testing Industry?

Thomas Toch

Executive Director, Association of Independent Schools of Greater Washington; and
Founder, Education Sector

Peg Tyre

Spencer Fellow at the Columbia University Graduate School of Journalism; and
Author, *The Trouble with Boys*

Common Education Standards: Tackling the Long-Term Questions

Thomas B. Fordham Institute

June 2010



Introduction

The emergence in recent months of common state standards in reading and math—and the U.S. Department of Education’s plans to fund new testing systems aligned with those standards—are likely to spell major changes for the testing industry. These changes are aimed at creating assessment systems that are a lot better for students, teachers, and the public education system as a whole than those created under (or swept up into) the federal No Child Left Behind Act (NCLB). But coming up with new assessment systems that help to propel more schools toward higher achievement requires two things: policymakers must give educators the incentives they’ve lacked under NCLB to measure advanced skills. And states must be willing to pay for higher quality tests once the federal funding ends.

The National Governors Association and the Council of Chief State School Officers have sponsored the development of common K-12 education standards in math and English/language arts—a project known as the Common Core State Standards Initiative (CCSSI)—in an effort to improve college readiness for the nation’s students and replace the patchwork of often-superficial state standards that have contributed to troubling gaps in educational opportunities for students living in different parts of the country.¹

Working closely with the Obama administration and many school reformers, the same organizations have called on states to band together to build new assessment systems that reflect the higher aspirations of the common standards. The Department of Education announced in April 2010 that in September 2010 it plans to award as much as \$320 million to groups of states

¹ See the Common Core State Standards Initiative website at: <http://www.corestandards.org/>.

to produce “comprehensive assessment systems” based on the new standards over the next four years, and up to another \$30 million for high school end-of-course tests.²

Testing-company executives say that the rise of state testing consortia would likely cause more disruption in an industry that has already struggled under NCLB. There would be far fewer tests to create under the consortia, especially if states relinquish the option under the Department of Education’s proposal to permit states to customize up to 15 percent of the questions on the new tests (NGA and CCSSO are also permitting states to customize up to 15 percent of their standards under their CCSSI project). With test development making up some 20 percent of the \$800 million to \$1 billion state accountability testing market today, the result of fewer tests under the consortium model would be another financial blow to the testing industry.³

But they also say the federal initiative could present new opportunities. For example, the industry could concentrate its resources on building smaller numbers of more sophisticated tests that move beyond the low-level skills that states have stressed under NCLB. The industry could focus on creating tests with more measures of advanced skills such as analyzing and synthesizing information—skills that today’s advocates of common standards are demanding.

Towards that end, the Education Department, key congressional voices, and a host of testing experts, education reform groups, and private foundations are pushing the industry to produce more “open-ended” questions such as writing prompts that require students to craft their own responses rather than merely select from among multiple-choice answers (questions that in many instances reward guessing). These kinds of test questions are important to achieving higher

² U.S. Department of Education’s Race to the Top Assessment Program, <http://www2.ed.gov/programs/racetothetop-assessment/index.html>.

³ These figures are based on interviews with testing industry experts.

academic standards and sounder classroom practice, they say, because “what gets tested is what gets taught.”

The bubble sheet and #2 pencil may, in time, become antiques. Those pressing for innovation want the testing industry to expand technology-based testing that gives teachers real-time results. And they want teachers to play a role in scoring performance test items—as they do in the Advanced Placement and International Baccalaureate programs—in the hopes of increasing teachers’ knowledge of and investment in high-quality student work. This demand for a new generation of assessments represents “a once-in-a-lifetime opportunity” for innovation, says Doug Kubach, president and CEO of Pearson Assessment and Information, the largest company in the testing industry and one that may get larger still under the federally funded consortia. (Why? Unless the consortia hire their own test makers, and that’s unlikely experts say, at least a few testing companies will wind up with big contracts.)

But the innovations that the feds and the reformers want are complicated, expensive, and, in some instances, seemingly at odds with one another. And it’s unclear whether cash-strapped states will be willing to invest over the long haul in the kinds of higher quality testing that will help give traction to higher standards. They certainly haven’t been willing to spend on testing under NCLB; instead, they’ve commissioned primarily multiple-choice measures of basic skills in no small part because these are the cheapest tests on the market.

And states will only be willing to set their standards high—and make their assessments rigorous and “cut scores” demanding—so long as Congress changes the model for judging schools. NCLB requires that states take action against public schools in which even a single category of students fails to achieve state standards as measured on a single test—the so-called snapshot

model. That part of the law makes many states overly cautious about what they are willing to require students to learn—and be tested on. Currently the Obama administration is calling for changes in NCLB to require schools to administer baseline tests for all students and then test them again to show that they are moving toward mastery of rigorous standards (the so-called growth model). Sanctions—under the Obama plan—would be targeted at schools that have made the least progress with struggling students. Congress’s response to that recommendation will go a long way toward determining the rigor of the next generation of state tests, with huge implications for the future of the testing industry.

Before NCLB

Before passage of the No Child Left Behind Act in 2001, the testing industry was highly profitable. For decades, major players such as Harcourt, CTB/McGraw-Hill, and Riverside sold achievement tests primarily to schools and school systems, reflecting the nation’s long tradition of local control in education. The districts and schools compared their students’ results on Harcourt’s Stanford series, Riverside Publishing’s Iowa Test of Basic Skills, and the predecessors to CTB/McGraw-Hill’s Terra Nova tests to representative national groups of students (so-called norm groups). The publishers subcontracted the actual scoring of the tests to another major industry player, Pearson Educational Measurement. Local school systems established their own educational agendas and rarely held educators responsible for their students’ performance on these “norm-referenced” tests. Because the publishers owned the tests, they could require local educators to keep questions confidential and produce new tests only infrequently (every six to eight years)—a model that led to big profit margins.⁴

⁴ See Thomas Toch’s *Margins of Error: The Education Testing Industry in the No Child Left Behind Era*.

But the testing landscape started to change following the publication of *A Nation at Risk*, the scathing 1983 indictment of public education that accelerated a nascent minimum-competency testing movement. By the late 1980s, national political leaders, frustrated by the slow pace of local reforms in the wake of *A Nation at Risk* and other manifestos, demanded more testing. President George H. W. Bush convened a national summit with the nation's governors that established the country's first national student achievement goals. Following the summit, increasing numbers of states began requiring students to meet statewide standards and commissioned tests pegged to the new standards.

Arkansas governor Bill Clinton had led the governors at the summit. When Clinton reached the White House in 1993, he convinced Congress to *require* states to set standards and test whether students were meeting them. The No Child Left Behind Act, signed into law by President George W. Bush in January 2002, went even further—more than doubling the amount of state testing, requiring that the results be broken down by subgroup in every school to draw attention to the plight of the poor and other groups of students that schools traditionally had not served well. The tests also counted for more: the law linked serious consequences—including school reorganization and even closure—to failing scores.

The law, it seemed, would be a boon to the testing industry. But in fact it wreaked havoc. Suddenly states were required to use standardized test scores to measure school performance, which in turn necessitated new testing tailored to every state's standards—an unprecedented mandate that overwhelmed the testing industry. During the 1990s, the testing companies had often borrowed questions for new state tests from their norm-references tests, without aligning very carefully the questions to new state standards. Under NCLB, that was no longer good

enough. And because many states demanded that test questions regularly be made public, test publishers had to do the difficult and costly work of making and field testing new banks of test questions every year. The cushy days of norm-referenced testing were over.

The demands of NCLB cut testing companies' profit margins. So did intensifying competition in the industry. While the old-line companies had as much as 90 percent of the state testing market in the past, a host of new players have entered the state testing market in the last decade, hoping to get a share of the NCLB lucre—companies like Measurement Inc. and Questar Educational Systems, nonprofits like Measured Progress and American Institutes of Research, a host of niche companies that target parts of the testing business such as test writing or scoring.

The competition for market share has been so intense that companies have found themselves “winning” money-losing bids. In one instance, Pearson, now the nation's largest test publisher, won a \$19 million contract to do Washington State's NCLB testing from 2004 to 2008 that ended up costing the company \$26 million to complete.

At the same time, state lawmakers have sought to spend as little as possible on NCLB testing—an average of \$20 per student, or less than one-quarter of 1 percent of combined federal, state, and local per-pupil spending, despite testing's central role in today's education system.⁵ An NCLB provision introduced by the late Minnesota senator Paul Wellstone has provided states with federal funding to help pay for NCLB testing (if federal testing aid falls below specified levels—\$400 million in 2009-10—the law's student testing requirements are suspended). But states can spend the money on many things other than making tests, such as “improving the dissemination of information on student achievement and school performance,” and the Boston-

⁵Thomas Toch *Margins of Error: The Education Testing Industry in the No Child Left Behind Era*, p. 11.

based Eduventures concluded in an analysis of state spending under the law that “states in many cases have opted to allocate most of those [federal] dollars to the development of state standards or initiatives aimed at aligning instruction to state standards”—important activities, but not test-building.⁶

Harcourt Assessment, founded in the 1920s and long one of the nation’s largest educational testing companies, collapsed under the increasing demands and intensifying competition under NCLB. It delivered some Illinois’s 2006 test results in 2007, it lost contracts, and the once-lucrative company went into the red. Its owner, Dutch publisher Reed Elsevier, sold the company in 2008 to testing giant Pearson for \$950 million. And Harcourt is not the only company that has struggled under NCLB. There have been numerous reports of scoring delays and errors, and companies have come under fire for administering tests with poorly vetted questions and items that are poorly aligned with state standards. In June 2010, Florida’s chancellor for elementary and secondary education apologized to the state’s educators after administration and scoring problems by its testing contractor, Pearson, put the state’s NCLB testing program six weeks behind schedule.⁷ “The entire industry has fared terribly under NCLB,” says Barry Topol, the chief financial officer at Harcourt Assessment from 2005 to 2008 and, more recently, co-founder of the Assessment Solutions Group, a testing consulting company.

Because many states have responded to NCLB’s demands by lowering their standards and spending as little as possible to comply with the law’s testing requirements, the testing industry has found itself producing increasing numbers of tests in the NCLB era that measure mainly low-

⁶ See Thomas Toch’s *Margins of Error: The Education Testing Industry in the No Child Left Behind Era*.

⁷ Kathleen McGrory and Hannah Sampson, “Glitches Delay FCAT Scores,” *Miami Herald*, June 6, 2010, <http://www.miamiherald.com/2010/06/05/v-print/1665809/glitches-delay-fcat-scores.html>.

level skills via multiple-choice questions—the fastest and cheapest tests to make, administer, and score. But many experts say open-ended questions are best suited to measuring the advanced skills that today’s reformers say students need. Other industrialized countries with high-performing education systems are moving toward greater use of essays, experiments, and other open-ended assessments (also called constructed-response or performance measures), largely because with such tasks, “teaching to the test” is more likely to foster higher-order learning.

The Government Accountability Office in 2009 reported that thirty-eight of forty-eight states administered tests that consisted entirely or almost entirely of multiple-choice items, and that the number of open-ended questions had declined in thirteen states in math since 2002.⁸ At the same time, the performance of U.S. students has declined on an important international benchmark of students’ more advanced skills such as problem solving and the ability to apply knowledge to new tasks: the Programme for International Student Assessment, or PISA.⁹

Testing the Common Core

President Obama and Secretary of Education Arne Duncan want the states to do better when it comes to standards and testing. States seeking the \$4 billion in federal stimulus funding under the Department of Education’s Race to the Top program have to commit to using new standards (most people assume it will be the NGA/CCSSO common core) and to joining a testing consortium that measures students’ grasp of the new standards.

⁸ U.S. Government Accountability Office report, *No Child Left Behind Act: Enhancements in the Department of Education’s Review Process Could Improve State Academic Assessments*, <http://www.gao.gov/new.items/d09911.pdf>, p. 18. The number of open-ended questions declined in reading tests in 11 states during the same period, while the number of such questions increased in five states in math and five states in reading.

⁹ See the PISA website at: http://www.pisa.oecd.org/pages/0,2987,en_32252351_32235731_1_1_1_1_1,00.html.

The department has very high expectations for the testing consortia. It is demanding testing systems that benchmark schools against international standards, compare states' performance, track students' growth in classrooms, make greater use of technology in test taking and scoring, test every student every year, and help teachers in their daily work to a far greater degree than testing has under NCLB.

At this writing, there are two leading contenders to build the comprehensive testing systems that the federal government is funding: the Smarter Balanced Assessment Consortium, a coalition of forty-five states led by Oregon, Utah, and Wisconsin, among others, and the Partnership for Assessment of Readiness for College and Career, an alliance of twenty-six states led by Florida and Massachusetts that is working closely with Achieve, a Washington, D.C.-based organization founded by governors and corporate leaders to promote higher education standards.¹⁰

As the Department of Education guidelines require, both consortia would establish new assessments of school performance for accountability, increase online testing, link testing more closely to instruction, and compare scores across states. The Smarter Balanced consortium would also give teachers a central role in designing and scoring tests and would include science experiments and other performance tasks in students' standardized evaluations and school accountability calculations. The Partnership consortium would administer and score performance measures with computers.

Presumably the consortia would have the ability to demand dependable tests that measure the kinds of advanced skills reformers say are crucial. In the short term, they will have the federal

¹⁰ For a summary of the two consortia models, see: <http://dese.mo.gov/stateboard/meetings/April/documents/CommonAssessment.pdf>.

money to spend on testing development. In the longer term, they will represent a large enough share of the national testing market to make substantial demands on the industry.

The creation of consortia and the pooling of testing would eliminate the current need under NCLB for every state to create unique tests, and thus cut down on development costs. (Currently about 20 cents of a state's testing dollar goes to test development.) Assuming the same amount gets spent on testing, the "savings" could be used to improve test quality. Free from the Sisyphean task of coming up with hundreds and hundreds of new test items to meet individual state standards, the limited number of working psychometricians could focus on coming up with deeper tests of knowledge that are well-aligned with the common core.

Standardized tests with a sufficient number of non-multiple-choice items to measure advanced skills and encourage teachers to teach such skills in their classrooms would cost about \$55 per student, compared to the \$20 per student for multiple-choice tests, according to a new study by Topol and his colleagues for the Stanford Center for Opportunity Policy in Education. But the study, *The Cost of New Higher Quality Assessments: A Comprehensive Analysis of the Potential Costs for Future State Assessments*, calculates that the cost of such tests would drop to \$40 per student for states participating in a national testing consortium.¹¹

Some states are already achieving such savings. Following NCLB's enactment, Rhode Island, Vermont, and New Hampshire formed the New England Common Assessment Program, or NECAP. And since 2005, that small consortium has produced tests with an ample number of open-ended questions at two-thirds of what the states would have paid had they developed

¹¹ To read the Stanford Center study in its entirety, see:
http://edpolicy.stanford.edu/pages/pubs/pub_docs/assessment/scope_pa_topol.pdf.

similar tests individually.¹² (Maine joined the consortium in 2009.) Similarly, nearly two dozen states have administered common tests for English Language Learners under NCLB through a consortium called World-Class Instructional Design and Assessment (WIDA), which first began testing in 2004-05.¹³

There's probably no room in a market dominated by a few large consortia of states for the eight or nine comprehensive testing companies in today's industry (firms that do everything from write test questions to score and report test results) plus the two dozen niche players that have emerged under NCLB. Most industry insiders believe that one or two large firms will prevail in a consolidated marketplace, very likely including Pearson, the dominant company in today's market.

But how many companies survive will depend partly on how much test customization states demand. If many states use nothing but consortium-developed test questions, which would save them the greatest amount of money, there will be fewer opportunities for niche players. But if states avail themselves of the federal education department's allowance for states in consortia to customize up to 15 percent of their tests, there's likely to be sufficient additional work for companies that are not developing the core consortium tests.

Subcontracting within the industry could also keep more companies in the game. There's a great deal of that work going on under NCLB; the Educational Testing Service, the general contractor in California, has hired giant Pearson to handle the logistics of that state's testing. Under a consortium model, smaller companies might write test questions, or score some types of

¹² See Thomas Toch's *Margins of Error: The Education Testing Industry in the No Child Left Behind Era*.

¹³ See the WIDA website at: <http://www.wida.us/standards/elp.aspx>.

questions. Others might develop and administer tests for subgroups—e.g., special education students or English Language Learners from Russia.

Companies large and small might find opportunities offering additional interim and diagnostic assessments to schools and school districts on top of what is required by the Department of Education’s testing RFP. Currently states have different rules for disclosing what is on their tests—some make the whole test publicly available, others release portions or individual questions. Depending on who actually owns the consortia tests and test questions—and if and to what degree that material is made public—small companies might sell supporting products such as tutoring materials.

From Here to There

The Partnership consortium says it would launch its first tests in the spring of 2014; the Smarter Balanced tests would debut a year later. They’ll need at least that much time.

Getting state experts within a consortium to agree with what should be tested will be tricky.

“Although common standards will make the processes somewhat easier, there is still a selection process when coming up with questions,” says H. Gary Cook, a former Wisconsin testing director and testing industry executive who works with the WIDA consortium. “Getting consensus on *exactly* what you test can be a painstaking and time-consuming process.”

The large numbers of states at the table will make those discussions tougher. So will the Obama administration’s requirement that the consortia candidates include in their proposals detailed information about the number and types of questions they’ll have on their tests, and what those questions will cost. “Instead of the states being the clients and the testing companies being the

proposers, it feels like the U.S. Department of Education has become the client and the states are the proposers,” says Brian Gong, former director of Kentucky’s innovative testing department and now executive director at the National Center for the Improvement of Educational Assessment, a nonprofit consulting firm that has provided technical advice to both major consortia seeking federal funding.

It will also be difficult to build more constructed-response questions into standardized tests that are administered nationally. As the recent report from the Assessment Solutions Group suggests, such questions are more expensive, more challenging, and more time-consuming to create, administer, and score. There’s plenty of evidence—from the Advanced Placement and International Baccalaureate programs, the New York State Regents exams, and other state and foreign tests, where such questions are widely employed—that essays and other open-ended questions can be used reliably on a large scale. But ensuring, for example, that such tests align from one grade and one year to the next, and that such questions are graded to the same standards, is more difficult and costly than with multiple-choice questions. Some in the testing industry question whether states are willing to invest in the technical expertise required by more ambitious tests, given their reluctance to spend under NCLB. “The feds’ \$160 million [the upper limit the Department of Education plans to spend per consortium] is more than enough to create the [new generation of] tests,” says one insider. “It’s not enough to sustain them.”

Many hope that a shift from paper-and-pencil testing to online assessments will improve large-scale testing. A much greater use of technology will result in a “less cumbersome assessment process and faster delivery of results and allow more real-time adjustment in instruction, get greater use out of longitudinal data, and provide for a wider range of accommodations for special

education students and English Language Learners,” declared Pearson president and chief executive officer Doug Kubach in testimony to the House Committee on Education and Labor last December. Experts suggest that a move to technology-driven assessment is only a matter of time. “The paper-and-pencil test is probably a thing of the past,” says John Tanner, former CCSSO testing director and now president of Test Sense, a consulting company.

But we’re not there yet. Not even close. And the same combination of inertia and parsimony that has limited the usefulness of NCLB testing could slow the shift to technology. An online testing model could certainly be cheaper, considering that 25 percent of the testing industry’s revenue comes from plants like Pearson’s in Iowa City where companies print, ship, warehouse, score, and collate paper tests and reports. But the testing industry doesn’t yet have a strong online presence. Though some testing companies like the Northwest Evaluation Association have made progress with online testing, “none of it is as good as the companies say it is,” says the Center for Assessment’s Scott Marion, a consultant to many state testing departments. The *Miami Herald* recently reported that an online test administered by Pearson statewide in Wyoming in 2009-10 was plagued with problems, prompting the state to sue the company for \$9.5 million.¹⁴ Adds Topol, the former Harcourt official: “Companies don’t have the systems expertise.” A shortage of computers in schools needed to administer statewide testing online is, by itself, a major barrier.

The state testing consortia will also have to figure out how to accomplish the federal government’s challenging, and sometimes conflicting, testing goals. The Education Department has called on the testing consortia it funds to test every student every year at many grades and to

¹⁴ McGrory and Sampson, “Glitches Delay FCAT Scores.”

provide “real time” results to improve teaching. But it also wants more performance test questions and a greater role for teachers in scoring tests, features likely to add time and cost to the testing enterprise and that would thus be easier to achieve if there were less testing, not more. At the same time, the department wants to know whether U.S. students measure up to their counterparts abroad, and it wants to be able to compare the performance of one teacher to another. To many experts, it’s an overly ambitious agenda. “They’ll have trouble doing everything,” Marion warns.

At the very least, experts say, a single set of end-of-year tests of the sort that states have used under NCLB is unlikely to accomplish the multiple aims of federal officials and school reformers. Indeed, the Department of Education is calling on consortia to build systems of tests. But the contours and complexity of these new systems are not yet clear. There are likely to be middle-of-course tests, the results of multiple tests combined into single scores, and tests delivered in new ways, among other changes.

Ultimately the extent to which standardized testing and the testing industry change in the wake of the new common standards—and the impact of such changes on the nation’s classrooms—depends on how much of an incentive states have to embrace the new common standards and the greater rigor they represent. NCLB calls for high standards and rigorous tests, yet many states responded by lowering standards. The law gives them incentives to do so. The Obama administration has urged Congress to eliminate such incentives when it reauthorizes NCLB—notably the requirement that schools be declared “in need of improvement” if one or more groups of students don’t measure up to state standards, rather than if schools fail to increase students’ overall learning during the school year. If Congress incorporates the Obama changes

into its reauthorization of the Elementary and Secondary Education Act, there may well be sweeping changes in education testing and in the testing industry. If it doesn't, the impact of the emerging common standards and state testing consortia that are designed to support the standards is likely to be much diminished.

References

- Common Core State Standards Initiative. <http://www.corestandards.org/>.
- McGrory, Kathleen, and Hannah Sampson. "Glitches Delay FCAT Scores." *Miami Herald*, June 6, 2010, <http://www.miamiherald.com/2010/06/05/v-print/1665809/glitches-delay-fcat-scores.html>.
- Missouri Department of Elementary & Secondary Education. <http://dese.mo.gov/stateboard/meetings/April/documents/CommonAssessment.pdf>.
- OECD Programme for International Student Assessment (PISA). http://www.pisa.oecd.org/pages/0,2987,en_32252351_32235731_1_1_1_1_1,00.html.
- Toch, Thomas. *Margins of Error: The Education Testing Industry in the No Child Left Behind Era*. Washington, DC: Education Sector, 2006.
- Topol, Barry, John Olson, and Ed Roeber. *The Cost of New Higher Quality Assessments: A Comprehensive Analysis of the Potential Costs for Future State Assessments*. Stanford, CA: Stanford University, Stanford Center for Opportunity Policy in Education, 2010. http://edpolicy.stanford.edu/pages/pubs/pub_docs/assessment/scope_pa_topol.pdf.
- U.S. Department of Education. Race to the Top Assessment Program. <http://www2.ed.gov/programs/racetothetop-assessment/index.html>.
- U.S. Government Accountability Office. Report to the Chairman, Committee on Health, Education, Labor, and Pensions, U.S. Senate. *No Child Left Behind Act: Enhancements in the Department of Education's Review Process Could Improve State Academic Assessments*. GAO-09-911. Washington, DC: Government Accountability Office, September 2009. <http://www.gao.gov/new.items/d09911.pdf>, p. 18.