

Intergovernmental Approaches to Standards and Assessment

Lynn Olson

For the first time in history the nation has embraced a goal of bringing all children, not just some, to a high standard in reading and mathematics. But this ambitious target means less than meets the eye because academic expectations and tests vary so widely from state to state.

When Congress passed the No Child Left Behind Act in 2001, it set a target of having all students reach the “proficient” level in reading and mathematics by 2013-14. But the federal law left it up to states to develop their own English and mathematics standards, the tests to measure whether students have met those standards, and the scores that designate when a child has reached the “proficiency” bar.

The result has been widely varying standards and rigor from state to state, making the definition of “proficiency” virtually meaningless. In Tennessee, for example, 88 percent of eighth graders were judged “proficient” on the state reading test in 2005, while only 26 percent reached that level on the National Assessment of Educational Progress, a congressionally mandated exam that tests a representative sample of students in every state. A recent study by Policy Analysis for California Education, a research center run by Stanford University and the University of California, found the same pattern in a large majority of states, with students performing noticeably worse on NAEP than on state exams.

This lack of transparency and uniformity makes it hard for parents and the public to know how schools are performing. It has stretched thin the resources of textbook companies and test publishers who must develop instructional materials and assessments “customized” to meet each state’s unique version of Algebra I or high school chemistry. And it has revived a long simmering debate about why mathematics or reading should look different in Mississippi than in New York.

In the past year, there’s been a renewed call for common national standards and measures of performance. The bi-partisan Commission on No Child Left Behind recommended the development of voluntary national standards and tests based on the frameworks for NAEP exams. States could adopt these standards and assessments or have their own subject to review against the national model. The Washington-based Education Trust has proposed that states adopt standards and tests that would better prepare students for college and careers, with sanctions for states that show large discrepancies between passing rates on their own tests and on NAEP.

Yet despite a widespread desire for more uniform and rigorous standards and better assessments across states, there’s little appetite for having the federal government develop national standards and tests. Instead, proponents of more unified standards have begun searching for ways in which states, perhaps in partnership with the federal government, could more effectively share responsibility and authority for developing educational standards and ensuring the quality of educational measurement.

“We’ve got a well-documented problem,” said Chester E. Finn, Jr., the president of the Washington-based Thomas B. Fordham Institute. “The current structure



Lynn Olson has spent over eighteen years as a journalist covering public education, national education policy, and assessment issues. She is managing editor of Education Week and project editor of that publication’s “Quality Counts” report. Olson graduated from Yale University, and her professional awards include those issued by the Education Writers Association, the National Association of Secondary School Principals, and the International Reading Association.

for doing standards isn't working well at all. What we lack as a country is an obvious mechanism for solving this problem.”

In October 2007 the Nelson A. Rockefeller Institute of Government, the public-policy research arm of the State University of New York, brought together a group of some 40 state and federal education officials, policy advocates, testing experts, and educational researchers for a one-day symposium on intergovernmental approaches for strengthening K-12 standards and tests, with the support of the Chicago-based Joyce Foundation and Spencer Foundation. While the group did not attempt to reach a consensus, this paper summarizes some of their ideas. (*For a more detailed description of these ideas, see “Intergovernmental Approaches for Strengthening K-12 Accountability Systems: A Framework Paper,” by Allison Armour-Garb, the director of education studies for the Institute, at: http://www.rockinst.org/research/education/default.aspx?id=134&ekmense1=10_submenu_212_btnlink.)*

State-Led Collaboratives

One solution is to encourage states to come together voluntarily to agree to common standards and tests. The American Diploma Project (ADP), led by the Washington-based Achieve, Inc, a nonprofit group formed by governors and business leaders in the mid-1990s, is the largest multi-state collaboration to develop common standards and tests to date.

In 2004, in partnership with two other Washington-based organizations, the Education Trust and the Thomas B. Fordham Foundation, Achieve published a set of expectations for what high school graduates need to know in mathematics and English to succeed in college and the workplace. The benchmarks were based on a two-year research study with postsecondary institutions and employers in five states. Since then, 30 states have volunteered to join the ADP Network, a group of states committed to aligning their high school standards, assessments, and graduation requirements with what students will need to succeed in life after high school, based on the ADP benchmarks.

A subset of 13 of those states have gone beyond their initial commitment and have agreed to develop a common end-of-course exam for Algebra II, which will be given for the first time in spring 2008. Each of the participating states has agreed to give the same test, based on the same standards, and with the same cut score. But it will be up to each state how to use the results. Every state, for example, intends to use the test to help place students in credit-bearing courses in college, but some states will require all students to take the exam, while others will only encourage school districts to administer it. And while some states may count the results toward course grades in high school, others initially will limit use of the results to improving the curriculum.

Michael Cohen, the president of Achieve, who led an unsuccessful effort under the Clinton Administration to develop voluntary national tests in English and mathematics, now believes that such voluntary, interstate initiatives are the way to go.

“The work that Achieve has been doing with a critical mass of states, I think, will enable us to create an existence proof that it’s possible for states to have a common standard and for some large number of them to have a common assessment,” he said.

“I think that our ability to make progress on this agenda is enhanced without the federal government being heavily involved in it,” he added. “I can’t see the federal government overseeing or creating any of this without sparking truly mind-numbing fights over who’s in charge that are utterly divorced from what really needs to happen.”

On a smaller scale, the states of New Hampshire, Rhode Island, and Vermont have joined forces to create the New England Comprehensive Assessment Program. Working across state borders, the three states have developed common standards, tests, and cut scores in mathematics, reading, and writing. This spring, they will administer the first common science tests to students in grades 4, 8, and 11.

At a more informal level, the State Collaboratives on Assessment and Student Standards, coordinated by the Washington-based Council of Chief State School Officers, brings together states on a voluntary basis to discuss and research issues related to testing and accountability that are of mutual interest, such as the development of tests for English-language learners. “States are willing and want to work together on these issues,” said Phoebe Winter, an independent consultant. “It’s just having a vehicle to do it.”

A National Accreditation Agency

Another model for developing common standards and tests comes from England. There, the Education Act of 1997 created a public, non-departmental body called the Qualifications and Curriculum Authority. The QCA is charged with developing and maintaining England’s national curriculum in consultation with multiple stakeholders. But rather than developing course syllabi and assessments itself, the QCA awards multiple examination bodies the right to devise and administer tests based on the national curriculum and then regulates their quality.

A national qualifications framework, revised in 2006, lays out the responsibilities of organizations involved in developing and delivering the national curriculum assessments, and sets out QCA’s monitoring role for ensuring high-quality, consistent, and rigorous standards. Each year, for example, QCA publishes a report that highlights the percentage of secondary school exams dispatched to testing centers on time, the percentage of exam booklets without errors, and the percentage of exam results issued in a timely fashion.

Since 2004, the QCA has been involved in an ambitious effort to reform curricula, assessments, and qualifications for 14- to 19-year-olds in England that, as with the American Diploma Project, is designed to better prepare young people for further education and skilled careers.

Eva Baker, the co-director of the National Center for Research on Evaluation, Standards, and Student Testing at the University of California at Los Angeles, suggested that a similar structure could work in the United States. The existence of mul-

multiple examination bodies, with sufficient oversight to guarantee the comparability and worthiness of the exams, she asserted, could infuse the U.S. system with more choices and options for students while maintaining rigor.

Possible Federal Models

While previous federal attempts to set curriculum standards have failed, the federal government has been more successful in working with private industry to set standards for technology and business processes. The National Institute of Standards and Technology is a federal agency that conducts research, cosponsors public-private R&D efforts, provides technical assistance to small manufacturers, and gives an annual award for excellent performance and quality in the private sector. NIST also coordinates federal, state, and local technical standards with those developed by industry in order to eliminate duplicative or unnecessarily complex regulations.

A federal agency modeled after NIST, suggested Richard P. Nathan, the co-director of the Rockefeller Institute, could promulgate voluntarily developed education standards, support improvements in educational measurement, provide technical assistance to states and districts, and recognize promising innovations in educational accountability.

Others have suggested that the National Assessment Governing Board, which currently sets policy for the National Assessment of Educational Progress, could play this standards-setting role across states. NAEP is currently administered to a nationally representative sample of students in every state at key grades. And it is, in large part, the disparities between how students perform on NAEP versus state tests that have highlighted the gross differences in state standards and assessments. But participants at the Chicago meeting had little enthusiasm for turning a broader standards-setting and monitoring role over to NAGB, in part because it could threaten the integrity of NAEP as an information tool.

Another model, proposed by Thomas Toch of the Washington-based Education Sector, a nonprofit think tank, would create an independent national oversight agency to audit state testing programs and test publishers, “in the spirit of the Consumer Product Safety Commission.” The agency, which Toch dubbed the “National Testing Quality Commission,” would be designed to protect students and society from the harms caused by poorly designed tests and improper test use, in the same way that the Consumer Product Safety Commission protects the public from risks of injury or death from commercially developed products.

Like the CPSC, the commission could promote standards for tests and test use, regulate the tests and testing services produced and sold by private companies, ban inappropriate tests or test practices, collect and disseminate information about the quality of tests and testing programs, and maintain a hotline or other mechanisms for the public to report suspected problems.

“To me, the question is, ‘How do we ensure on an ongoing basis that standardized tests are valid and reliable, that they test what we want them to test, that they distinguish high-achieving students from low-achieving students?’ ” said Toch.

He favors the notion of an entity like the Consumer Product Safety Commission “in part because of the emphasis on the consumer nature of standardized testing in an era of educational accountability and school choice,” he added. “There’s very little information out there for parents, taxpayers, and the general public.”

Encouraging Innovation

Another question is how to design incentives that would encourage states to participate in any attempt to develop common and rigorous standards and tests. Many observers suggest that the No Child Left Behind Act’s emphasis on getting all children to proficiency by 2013-14, combined with sanctions for schools that fail to meet their targets, has spurred a race to the bottom rather than a pursuit of more ambitious standards.

One idea is to have the federal government finance a competitive grants process for states that agree to come together to develop common standards and tests. The government also could encourage states to participate by providing regulatory flexibility for those that do. The Education Trust, for example, has proposed that states willing to benchmark their standards and tests to the entry demands of college and the workplace be given extra time to get most students to that target.

At present, participants noted, far too little money is spent to develop the kinds of rich, robust tests needed to measure challenging standards. John Merrow, the executive producer and host of Learning Matters, argued that the private sector “spends ten times more testing flea powder and kitty litter than we spend on testing kids.”

That led to the notion of having private foundations underwrite a national competition to develop high-quality standards and assessments, similar to the federal competitions that now exist to develop new military weapons systems. Under such a scenario, three or four competitors might be charged to come up with initial designs, with only the best moving forward.

“The development work, with respect to standards and testing, ought to be privately financed by a consortium of foundations, maybe the Gates Foundation and a half-dozen others, putting together a \$100 million kitty, or whatever the requisite sum is, for getting this done in the next five years,” said Finn of the Fordham Institute. Once such standards and tests are past the initial development phase, he argued, “then we can see what the appropriate role of federal funding may be in implementation.”

“Competition among test developers, among private companies, could hopefully lead to good quality,” agreed Cohen of Achieve, “particularly if it’s a more focused competition, rather than 50 different tests at every subject and grade level. At present, we don’t have a mechanism to organize that kind of competition.”

In a reflection of the growing interest in developing common American standards, the James B. Hunt Jr. Institute for Educational Leadership and Policy in Chapel Hill, N.C., has contracted with the Washington-based National Academies to study the costs and consequences of having 50 different state standards and tests. It

also has asked what the best options would be for developing common education standards, noting the benefits and consequences associated with each option. The reports will be widely distributed.

Window of Opportunity

Despite the renewed push for more common, rigorous standards, it clearly won't happen overnight.

“For much of the nineteenth century,” noted Finn of the Fordham Institute, every railroad company in America had its own gauge for building train tracks, so companies could not move railroad cars or engines from one company's lines to the next. “I don't know how many decades it took before they got common, standard gauges on the railway system,” he said, “but it took a long time.”

Given Congress's failure to reauthorize the federal No Child Left Behind Act in 2007, discussions about the appropriate federal and state role in developing better standards and tests are sure to continue in parallel with the reauthorization debate. This may provide a new window of opportunity to make progress toward higher standards for all students. ■