

Assessing Deeper Learning

Preparing students to succeed in a complex world will require many changes in education policy and practice, but perhaps the most urgent priority is changing the way students are assessed. The state tests that have had such a dominant influence on classroom practice have tended to measure a relatively narrow set of knowledge and skills, and have encouraged teachers to focus on those competencies. The rapidly advancing global economy and an increasingly complex society require that all students develop a broader set of competencies. New assessments that measure these abilities are vital.

Specifically, the assessments need to measure deeper learning: how well students can demonstrate deep understanding of core content, how well they use their knowledge to solve problems, whether they are able to communicate in a range of media, and how skilled they are at collaborating with their peers.

In addition to *measuring* students' deeper learning abilities, the new assessments that states, districts, and schools need to put in place can also *foster* deeper learning in the classroom. Assessments have long played a powerful role in shaping classroom practice. Teachers pay attention to what is tested, particularly if there are consequences tied to the results, and many teachers model their instructional practices on the test format. If assessments measure deeper learning abilities, teachers are more likely to teach the relevant skills.

Assessments make expectations for student learning concrete. Standards might state that students need to know particular content and be able to perform certain skills, but the assessments make clear what students have to do to demonstrate that they can meet the standards. They also signal to teachers the kind of classroom experiences students should have in order to develop such competencies.

Fortunately, the United States is now poised to put in place new assessments that will begin to address the need for better measures of student abilities. Two consortia of states, the Partnership for Assessment of Readiness for College and Careers (PARCC) and the SMARTER Balanced Assessment Consortium (SBAC), are now developing new assessments to measure the Common Core State Standards in English language arts and mathematics. These assessments, which are scheduled to be in place in 2014–15, are expected to include a range of measures, including performance tasks that students complete during the school year and tools for teachers to enable them to assess their own students periodically and adjust instruction accordingly.

Deeper Learning Competencies

Deeper learning prepares students to

- know and master core academic content;
- think critically and solve complex problems;
- work collaboratively;
- communicate effectively; and
- be self-directed and able to incorporate feedback.

The Comprehensive Assessment Consortia

In September 2010, the U.S. Department of Education provided \$330 million to two consortia of states, the Partnership for Assessment of Readiness for College and Careers (PARCC) and the SMARTER Balanced Assessment Consortia (SBAC), to develop comprehensive assessment systems to measure student achievement against the Common Core State Standards. Currently, PARCC consists of twenty-four states, and SBAC consists of thirty-one states. (At this time, states can be part of more than one consortium.)

PARCC is planning to develop assessments that will be delivered mostly on computer and will include computer-enhanced items as well as performance tasks. The summative assessment is expected to include an end-of-course assessment and a performance task. PARCC also plans to develop formative assessment tools for teachers.

SBAC is planning to develop a system that will include performance tasks and a computer-adaptive assessment administered at the end of the year, along with interim assessments that can be administered during the year but that are optional for schools. The assessment is also expected to include a range of item types, including those calling for extended student responses.

At the same time, new technologies make possible advances in assessment that were not feasible a decade ago. With technology, new assessments can measure a much broader range of student abilities in a much more efficient manner than traditional tests can.

This brief will examine what new assessments that measure deeper learning may look like, how other countries use assessments like these, how technology can play a role in measuring deeper learning more efficiently and effectively, and what it will take to make these assessments the norm in U.S. schools.

The State of State Assessments

While many states and districts experimented in the 1990s with new assessments that attempted to measure a broad range of student abilities, most of those efforts have receded in recent years. Currently, sixteen states—which together educate nearly half of the nation’s schoolchildren—use tests that are exclusively multiple choice, and many other states include only a handful of items that ask students to write their own answers.¹

To be sure, multiple-choice tests are useful: they efficiently measure knowledge and skills and provide important information about students’ abilities. But they cannot measure all of the skills that an increasingly complex world demands of students. They cannot measure whether students can formulate their own ideas about how to approach a problem, for example, or how well they work with peers to generate a solution over time. They cannot measure whether students can write effectively in a variety of media. To measure those abilities, a different type of assessment is needed, one that provides students with opportunities to explore projects, come up with solutions, and explain their reasoning.

Many schools have developed such assessments and have used them to good effect. They have established performance assessments, which ask students to develop a product that shows that they can apply their knowledge, and they have developed portfolios that measure students’ performance in the classroom over the course of a year.



But state-administered summative tests have had a disproportionate influence over classroom practice. Because schools have been held accountable for their performance on the state tests, and are subject to sanctions if they fail to improve their performance each year, many schools have placed more and more emphasis on the kinds of abilities measured by the state tests.

What would a test that measures deeper learning look like? Consider the following example, from a high school information technology test used in England:

Litchfield Promotions works with more than forty bands and artists to promote their music and put on performances in England. The number of bands they have on their books is gradually expanding. Litchfield Promotions needs to be sure that each performance will make enough money to cover all the staffing costs and overheads as well as make a profit. Many people need to be paid: the bands; sound engineers; and lighting technicians. There is also the cost of hiring the venue. Litchfield Promotions needs to create an Interactive Computer Technology solution to ensure that they have all necessary information and that it is kept up to date. Their solution will show income, outgoings and profit.

Candidates will need to: (1) Work with others to plan and carry out research to investigate how similar companies have produced a solution. The company does not necessarily have to work with bands and artists or be a promotions company. (2) Clearly record and display your findings. (3) Recommend a solution that will address the requirements of the task. (4) Produce a design brief, incorporating timescales, purpose and target audience.

Produce a solution, ensuring that the following are addressed: (1) It can be modified to be used in a variety of situations. (2) It has a friendly user interface. (3) It is suitable for the target audience. (4) It has been fully tested. You will need to: (1) incorporate a range of software features, macros, modeling, and validation checks—used appropriately. (2) Obtain user feedback. (3) Identify areas that require improvement, recommending improvement, with justification. (4) Present information as an integrated document. (5) Evaluate your own and others' work.²

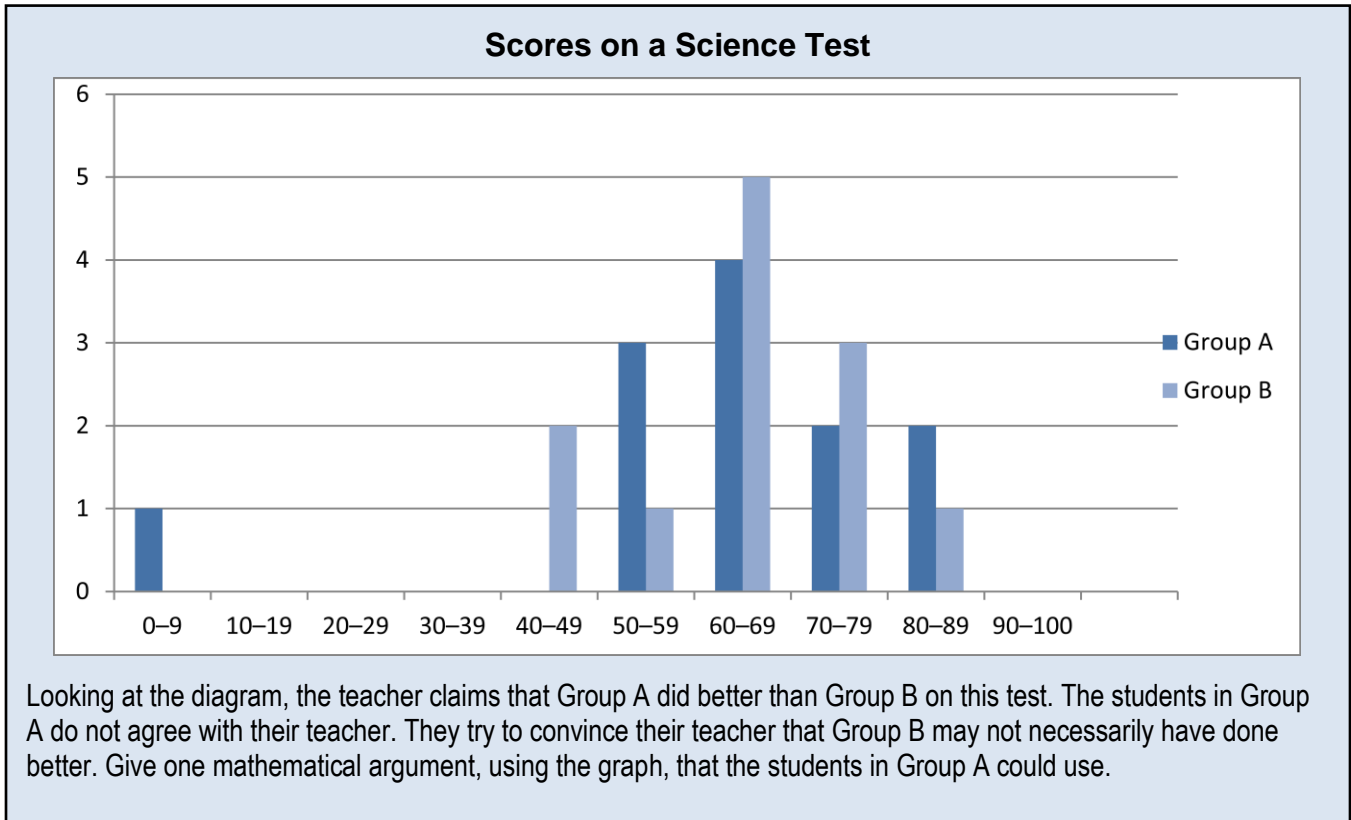
As this example indicates, other countries routinely employ assessments that measure whether students have a deep understanding of content, can think critically about relevant issues, and can communicate their understanding. These assessments are used to hold schools accountable for results as well as to inform instruction.

Perhaps the best-known international example of an assessment that measures deeper learning is the Programme for International Student Assessment (PISA), a test of reading literacy, mathematics, and science administered in sixty countries by the Organisation for Economic Co-operation and Development, a Paris-based organization of developed nations. As OECD's documents make clear, PISA was not designed to measure students' mastery of a school curriculum but, rather, to evaluate what students can do with the information they have learned. To that end, the assessments measure "both students' capacity to apply knowledge and skills in key subject areas and ... their ability to analyze, reason, and communicate effectively as they pose, interpret, and solve problems."³

Evidence from OECD research suggests that students who have demonstrated these abilities are more successful after high school. A study of 30,000 Canadian students who took the PISA in 2000 found that six years later those who scored at level 5 (the highest level) in reading were twenty times more likely than those who scored at level 1 (the lowest level) to attend a university. The effects of high PISA scores far outweighed those of any other factor.⁴



How does PISA measure these deeper learning abilities? Consider the following example:



To perform well on an assessment like this, students must know their subject matter well (in this case, statistics); they must be able to think critically and understand when an answer is reasonable; and they must be able to communicate their understanding. As PISA results indicate, fifteen-year-olds in the United States are much less able than their peers in other countries to demonstrate these abilities. In 2009, U.S. students ranked twenty-fifth out of thirty-four industrialized countries in mathematics.

An effort is now under way in the United States to develop a version of PISA to measure U.S. students' deeper learning abilities and provide results at the school level, not just at the national level. PISA for Schools, which is expected to be piloted in about 100 schools in the fall of 2011, will provide, for the first time, an indication of how students in schools that attempt to develop students' deeper learning competencies succeed in doing so. The assessment will also provide a means for comparing student results with those of their peers in other countries.

The Role of Technology

Technology offers an opportunity to measure deeper learning in new and innovative ways. Increasingly, states and school districts are administering tests on computers; the two assessment consortia are planning to deliver nearly all of their assessments in this manner. Computer-based testing has a number of benefits.

One advantage is that computer-based assessment makes new test formats possible. For example, students can search for original documents or background information to help them solve problems. Computers also can enable students to take part in simulations, in which they can collect data and solve



problems as they move through a virtual environment—much as many young people do every day on video games.

Another advantage is that computer-based tests can be scored almost instantly, allowing results to be provided to schools in time to inform instruction and programs within the same school year. No longer will students and teachers have to wait until after the year is over to receive their test results.

At the same time, because young people are increasingly adept at using computers, computer-based assessment allows students to take tests in a format that is comfortable for them.

Researchers from the Harvard University Graduate School of Education have developed a sophisticated set of “virtual performance assessments” to measure eighth graders’ performance in science. In one example, “A New Frog in Town,” students choose avatars to enter a virtual environment in which frogs mysteriously develop six legs. The students must move through the environment to gather evidence and test hypotheses about what might be causing the mutation. The virtual environment enables students to collect a broad range of evidence—for example, water and frogs from a variety of sources—and conduct tests to determine their characteristics. The computer also collects information on what choices the student makes and his or her ability to back up claims with evidence and reasoning—abilities that are often difficult to measure with conventional tests.⁵

Are Deeper Learning Assessments Feasible?

One reason state officials have been reluctant to implement assessments that measure deeper learning is the perception that such assessments are expensive and difficult to administer. It is true that conventional multiple-choice tests are inexpensive; typically, such tests cost about \$10 per pupil, just one-tenth of 1 percent of overall average per-pupil spending in the United States. Assessments that include open-ended questions are typically more expensive because they require human scoring.⁶

However, a recent report produced for the Stanford Center for Opportunity Policy in Education (SCOPE) found that high-quality assessments that measure deeper learning can be produced and administered relatively efficiently. By forming consortia to develop assessments jointly, using technology to deliver and score assessments, and relying on teachers to score the assessments, states can significantly reduce the cost of high-quality assessments.⁷

Technology can also make complex assessments more administratively feasible. By administering assessments on computer, states can reduce the logistical challenge of managing test booklets and other materials. To the extent that they rely on computer scoring, states can reduce the costs associated with human scoring and minimize the possibility of human error.⁸ Computerized testing can also enable test results to feed directly into data systems, making them available to school, district, and state administrators almost instantly.



Federal Policy Recommendations

The federal government has made a strong start toward supporting the development of assessments that measure and drive deeper learning by providing \$330 million to two consortia of states to develop new comprehensive assessment systems. Forty-five states have signed up to belong to one or both consortia, making it likely that these assessments will be widely used across the country. However, the federal government can further ensure that states develop and implement assessments that measure deeper learning through additional legislation. Specifically, the federal government should:

- **Require that state tests assess deeper learning competencies.**

Forty-five states and the District of Columbia have adopted new standards that set as their goal college and career readiness for all students. These standards expect students to develop the deeper learning competencies they will need to succeed after high school. The federal government can support this development by requiring states to align their assessments to standards for college and career readiness and ensuring that the assessments measure the full range of content and skills included in the state standards, not just those that are easiest to measure.

- **Provide professional development funds toward the improvement of teachers' assessment knowledge and skills.**

Deeper learning assessments require teachers to play a greater role in assessing student performance than they do now. Teachers need to know how to develop and review assessment tasks, how to score them accurately and reliably, how to develop and employ effective classroom assessments to track student knowledge and skills over time, how to interpret assessment results, and how to modify instruction based on assessment results. The involvement of teachers in assessment programs in other countries suggests that these practices are valuable in improving instruction, but teachers need support to be able to perform these practices effectively. The professional development needs are particularly acute for high school teachers, who typically teach many more students than elementary teachers do.

The federal government could direct states to include plans for developing teachers' assessment knowledge and skills as part of their plans for professional development funds under Title II of the Elementary and Secondary Education Act. The Higher Education Act could also offer grants to higher education institutions that make assessment literacy an integral part of teacher-preparation programs.

- **Ensure that assessments fairly include all students, including English language learners and students with disabilities, and that they provide accurate information about student knowledge and skills.**

Over the past eight years, states have made great strides in including students with special needs in assessment programs, and achievement by students who have often been underserved has improved. However, new assessment techniques and technologies, such as universal design for learning and computer-based assessments, could enable more students to participate in regular assessments.

Currently, the federal government is supporting the development of new assessments for students with severe disabilities by providing funds to two consortia of states, Dynamic Learning Maps and the National Center and State Collaborative, which are creating alternate assessments for students



with significant cognitive disabilities. These consortia are separate from PARCC and SBAC, which are developing comprehensive assessment systems that will serve the vast majority of students with special needs. The government could continue support for the assessments these consortia are developing as well as encourage states and state consortia to incorporate new technologies that expand participation in assessment. However, in no way should such changes lower expectations for students; *all* students should be expected to graduate from high school prepared for college and careers.

- **Provide states with financial support for assessment operating costs.**

The \$330 million the federal government provided to state consortia to develop new assessments tied to standards for college and career readiness is a major investment. Assessment development is a costly and labor-intensive process—new items and tasks must be written and field-tested, for example—and states and state consortia might be reluctant to take on the challenge, particularly in the current budget climate, without the federal incentive. Pooling their resources as consortia, an effort made possible by common standards, can also save money in the development process.

Once the initial development work is completed, though, states will continue to face ongoing costs to operate the new assessments. If the assessments include tasks that require human scoring, they will likely be more expensive than existing tests. The federal government currently provides support for states to develop and implement assessments through a formula grant program and a competitive grant program under Title VI of the Elementary and Secondary Education Act. These programs were created by the No Child Left Behind Act to ensure that states had the capability to administer assessments the law required. The federal government should maintain these programs and update them to ensure that the funds can be used to support the development and implementation of higher-quality assessments.

Conclusion

Assessments are powerful levers in education, and current state assessments have in many ways impeded the development of deeper learning. Better assessments—ones that measure students’ abilities to understand core content, to think critically and solve problems, to communicate effectively, and to collaborate with their peers—can help ensure that students have opportunities to develop those competencies. The assessments under development by two consortia of states represent a promising start.

However, assessments are not enough to develop students’ deeper learning competencies. Students also need instruction that fosters deeper learning, as well as opportunities in and out of school to develop their competencies. And teachers need curriculum and instructional tools as well as supportive policies that facilitate their work rather than restrict it.

Hundreds of schools currently are able to create structures and environments to support deeper learning, and they are showing great success. But for all students to reap the benefits, federal and state policies need to change. Assessment is an ideal place to start.

This brief was written by Robert Rothman, a senior fellow at the Alliance for Excellent Education.

Support for this paper was provided by the William and Flora Hewlett Foundation. The opinions expressed are those of the author(s) and do not necessarily reflect the views of the William and Flora Hewlett Foundation.



Endnotes

- ¹ “Quality Counts 2010: Fresh Course, Swift Current,” *Education Week* 29, no. 17 (2010).
- ² L. Darling-Hammond, *Benchmarking Learning Systems: Student Performance Assessment in International Context* (Stanford, CA: Stanford Center for Opportunity Policy in Education, 2010).
- ³ Organisation for Economic Co-operation and Development, *PISA 2009 Results: Executive Summary* (Paris: Author, 2010), p. 3.
- ⁴ Organisation for Economic Co-operation and Development, *Pathways to Success: How Knowledge and Skills at Age 15 Shape Future Lives in Canada* (Paris: Author, 2010).
- ⁵ J. Clarke-Midura and C. Dede, “Next Generation Assessments for Measuring Complex Learning in Science,” in Rennie Center for Education Research and Policy and PACE, *The Road Ahead for State Assessments* (Boston: Rennie Center for Education Research and Policy, May 2011).
- ⁶ According to an estimate prepared by the Assessment Solutions Group, a typical state spends \$19.93 per pupil on a battery of tests in reading, writing, and mathematics in grades three through eight and grade ten. A battery of “high-quality” assessments, which would include substantially more open-ended tasks, would cost the typical state \$55.67 per pupil. B. Topol, J. Olson, and E. D. Roeber, *The Cost of New Higher-Quality Assessments: A Comprehensive Analysis of the Potential Costs for Future State Assessments* (Stanford, CA: Stanford Center for Opportunity Policy in Education, 2010).
- ⁷ Ibid.
- ⁸ Clarke-Midura and Dede, “Next Generation Assessments.”

