

# Assessments for Deeper Learning:

WHAT'S  
NEXT  
AND WHAT  
WILL  
THEY  
COST?

by Linda Darling-Hammond

## States Are Creating High-Quality and Affordable Assessment Systems

Few factors are more powerful at influencing what is taught in our nation's classrooms than the effective pairing of learning standards with the curriculum and assessments that enact them. The United States is poised to begin a new era on both fronts as states phase in the Common Core State Standards (CCSS) and new assessments aligned to the standards.

Together, the CCSS and new assessments are intended to promote the instruction and learning our students need to succeed in a global economy. Studies show that employers' demands for workers with routine, repetitive skills—whether manual or cognitive—have dropped. Meanwhile, demand for employees with complex thinking and interactive skills has soared. Between 1970 and 2000, for example, the top skills demanded by Fortune 500 companies shifted from the 3 Rs—reading, writing, and arithmetic—to teamwork, problem solving, and interpersonal skills.

During this time, the United States has trailed other nations that are investing in assessments that measure deeper learning—including students' ability to analyze, synthesize, compare, connect, critique, hypothesize, prove, and explain their ideas. Along with reforms involving curriculum, instruction, and teacher quality, Singapore, New Zealand, Hong Kong, Australia, and other high-achieving jurisdictions have introduced ambitious assessments that call on students to demonstrate what they know using sophisticated written, oral, mathematical, physical, and multimedia products.

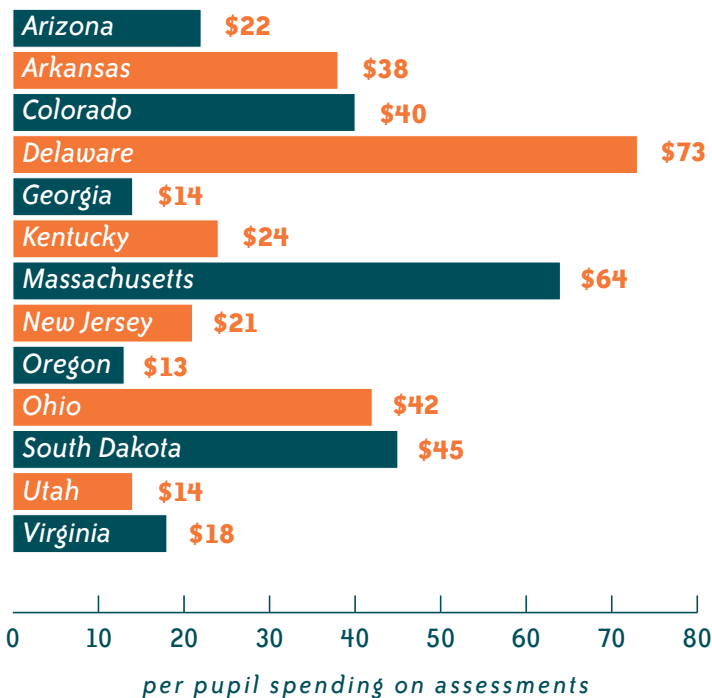
Successfully bringing deeper learning to scale will require a new generation of assessments that can measure higher-order thinking as well as real world skills. This article looks at where we're at in the development of such assessments, and how much they are likely to cost.

Meanwhile, U.S. students have been bubbling in answers on multiple-choice tests that measure lower level skills. As federal requirements under the No Child Left Behind Act (NCLB) led to extensive annual testing, states sought to control costs by using less expensive, machine-scored tests. Most dropped human-scored assessments requiring writing, research, and extended-problem solving. As a result, a recent RAND Corporation study found that on 17 states' tests, fewer than 2 percent of mathematics items and only 21 percent of English language arts items measured higher order thinking skills.<sup>1</sup> The vast majority of items called for memorization, recognition of information, and use of routine procedures.

Today, the CCSS are shifting the focus from rote learning and memorization to critical thinking skills and application. These standards are intended to be “fewer, higher, and deeper” than previous standards, setting the stage for a new focus on assessments of deeper learning. Two state consortia—the Partnership for Assessment of Readiness for College and Careers (PARCC) and the Smarter Balanced Assessment Consortium—were formed to develop next-generation assessments of these standards, which will be launched in 2014-15.



## Sample of States' Primary Assessment Contract Costs Per Pupil for Grades 3–9



Source: M Chingos, *Strength in Numbers: State Spending on K–12 Assessment Systems* (Washington, DC: Brookings Institution, 2012), [www.brookings.edu/research/reports/2012/11/29-cost-of-ed-assessment-chingos](http://www.brookings.edu/research/reports/2012/11/29-cost-of-ed-assessment-chingos).

## Evaluating the Quality and Costs of Next-Generation Assessments

As states evaluate and select assessments and support policies, two issues likely will guide their decisions. The first issue is how to determine the quality and efficacy of next-generation student assessments. Second is the cost of these assessments compared to the perceived and actual costs of their current assessment systems.

Given that most current state assessments still focus mainly on rote learning, the ability to follow procedures, and other basic skills—creating a barrier for reforms based on 21st century competencies—what should new assessments that accurately measure the most important kinds of learning look like? A group of 20 assessment experts suggested they should meet five criteria:

- Assess higher-order thinking skills, especially those skills that are transferable and allow students to apply knowledge to new problems and situations;

- Use high-fidelity assessment of skills as used in the real world, such that students are asked to show they can communicate, use technologies, and conduct research;
- Are internationally benchmarked to align assessment content and measurement practices with those used in leading nations;
- Use items that are instructionally sensitive and educationally valuable—that is, are able to reflect how well teachers are teaching and give them guidance on how to improve; and
- Are proven to be valid, reliable, and fair, as well as accessible to all learners.<sup>2</sup>

The consortia-developed assessments are being designed to meet each of these criteria to a much greater degree than the previous generation of tests. In areas that the consortia tests will not reach—such as long-term research and investigation tasks or the ability to communicate orally, visually, and with technology tools—some states are developing their own performance tasks. Others, through the Council for Chief State School Officers' Innovation Lab Network, are developing an assessment bank of such tasks that can be incorporated into proficiency-based assessments, graduation portfolios, and formative classroom activities.

Despite considerable interest among educators and policymakers for upgrading the quality of assessments to leverage instruction geared toward deeper learning, a number of challenges remain. Perhaps the biggest challenge facing state leaders as they move toward high-quality assessments of deeper learning is the perception that such assessments are not affordable.

While state-by-state comparisons are not always easy to make (both in terms of reliability of the data and because state assessment systems vary in their components), two recent studies have found similar cost estimates. One, a survey across grades in 37 states, found an average cost of \$24.52 per pupil for annual summa-

tive reading and math assessments required under NCLB.<sup>3</sup> Another survey of 45 states found an average of \$27 per pupil for testing in grades 3–9. This latter study also found a very wide range in spending, from \$13 per pupil in Oregon to \$105 per pupil in Hawaii.<sup>4</sup>

The costs of state end-of-year math and reading tests are not the whole picture, however. The true cost of assessment for many states is much higher. A recent Stanford University study found that when the costs for interim testing practices, test preparation materials, staff time for development and administration of tests, and professional development related to tests are included, states and localities together are on average spending more than \$50 per pupil on reading and math testing alone.<sup>5</sup> Unfortunately, in most cases these expenditures do not improve the overall quality of assessment, teaching, or learning because they mostly focus on items measuring low-level skills that now dominate the testing landscape.

These overall costs are considerably higher than the projected costs of the tests being developed by the new assessment consortia. For example, the Smarter Balanced summative assessments in English language arts and math are, together, estimated to cost \$22.50 per student. The suite of summative, interim, and formative assessments; time for teacher scoring; and supportive instructional materials is estimated to cost \$27.30 per student. The PARCC system, which uses computer-based administration, is estimated to cost \$29.50 per student for midyear performance-based assessment and an end-of-year sum-

mative assessment. This is less than one-half of one percent of average per pupil expenditures on education and benefits from economies of scale to keep costs down.

The Stanford study argued that states could get higher quality assessments—and spend less overall—if they combine state and local funds and focus these resources on a system of assessments that:

- Offers economies of scale through multi-state consortia, using online delivery and technology productively;
- Strategically aligns resources currently used for redundant, fragmented state and local testing; and
- Involves teachers more directly in developing and scoring open-ended assessments, using professional development time that can support the dual benefit of improved instruction and efficient use of resources.

### State Actions

State systems of assessments that support deeper learning are not totally new. Many states created such systems in the 1990s and they are used by many countries today.<sup>6</sup> During the 1990s, for example, Connecticut and Vermont used assessments in which students designed and conducted scientific experiments, often collaboratively, and analyzed and presented their results. Kentucky and Vermont's writing portfolios required students to plan, write, and revise extended pieces of work, while Oregon, Wyoming, and Wisconsin

created profiles of students' learning through sets of performance tasks.

New York State has long included teacher scoring in its Regents assessments, which include open-ended essays and tasks. Other states have also engaged teachers in scoring as a part of professional development time, knowing that teachers learn about their students and their instruction during the scoring and debriefing process.

Several states are already going beyond the current versions of the consortia assessments. Through the Innovation Lab Network, states are planning to augment the consortium assessments with more extended performance tasks that replicate, to the extent possible, the ways these kinds of abilities will be used in college and career contexts.

### Issues to Consider

Whether a state is part of one of the assessment consortia or not, there are several things state board members can do to address both the cost and quality of new assessments:

- Evaluate proposed assessments in light of the five criteria for high-quality assessment described earlier;
- Understand the gap between what planned assessments will measure and what needs to be measured to promote deeper learning and prepare a 21st century workforce;
- Consider how a system of state and local assessments can support these deeper learning goals;

- Before making decisions about how to allocate new resources, evaluate comprehensive data on current test expenditures at the state and local levels, taking into account all current assessment-related practices, and consider how to make coherent, high-leverage investments in new systems;
- Explain the benefits of more advanced assessments to parents and community members. Most parents, along with their children, are weary of multiple-choice tests and will be pleased to learn that their child's teacher will have better information to help customize instruction. But state boards, along with their state agency and school districts, should consider leading a communications plan to ensure all stakeholders understand the advantages of the new tests—and prepare them for the likelihood of initially lower scores; and
- Put in place a professional development plan to support teachers as they transition to assessments of deeper learning; this should include using teachers in scoring of the new assessments, which will help them understand and teach the standards.

If schools are to foster the transferable learning that is described in the Common Core State Standards and required of young people in contemporary society, assessments will need to support curriculum and teaching focused on deeper learning, along with traditional basic skills. New assessments under development are a step in the right direction, but states will need to take leadership in

figuring out how to design systems that will use them well. ■

*Linda Darling-Hammond is Charles E. Ducommun Professor of Education at Stanford University, where she has launched the Stanford Center for Opportunity Policy in Education and the School Redesign Network and served as faculty sponsor for the Stanford Teacher Education Program.*

<sup>1</sup>K. Yuan and V. Le, *Estimating the Percentage of Students Who Were Tested on Cognitively Demanding Items Through the State Achievement Tests* (Santa Monica, CA: RAND Corporation, 2012), [www.rand.org/pubs/working\\_papers/WR967.html](http://www.rand.org/pubs/working_papers/WR967.html).

<sup>2</sup>L. Darling-Hammond, J. Herman, J. Pellegrino et al., *Criteria for High-Quality Assessments* (Stanford, CA: Stanford University, Stanford Center for Opportunity Policy in Education (2013), [edpolicy.stanford.edu/publications/pubs/847](http://edpolicy.stanford.edu/publications/pubs/847)).

<sup>3</sup>B. Topol, J. Olson, and E. Roeber, *Getting to Higher-Quality Assessments: Evaluating Costs, Benefits, and Investment Strategies* (Stanford, CA: Stanford University, Stanford Center for Opportunity Policy in Education, 2012), [www.assessment-group.org/uploads/3.2\\_Affording\\_High\\_Quality\\_Assessments\\_ASG\\_2012\\_FINAL\\_CA\\_1.pdf](http://www.assessment-group.org/uploads/3.2_Affording_High_Quality_Assessments_ASG_2012_FINAL_CA_1.pdf).

<sup>4</sup>M. Chingos, *Strength in Numbers: State Spending on K-12 Assessment Systems* (Washington, DC: Brookings Institution (2012), [www.brookings.edu/research/reports/2012/11/29-cost-of-ed-assessment-chingos](http://www.brookings.edu/research/reports/2012/11/29-cost-of-ed-assessment-chingos)).

<sup>5</sup>L. Darling-Hammond and F. Adamson, *Developing Assessments of Deeper Learning: The Costs and Benefits of Using Tests That Help Students Learn* (Stanford, CA: Stanford University, Stanford Center for Opportunity Policy in Education (2013), [edpolicy.stanford.edu/publications/pubs/745](http://edpolicy.stanford.edu/publications/pubs/745)).

<sup>6</sup>For examples, see L. Darling-Hammond and F. Adamson, *Beyond Basic Skills: The Role of Performance Assessment in Achieving 21st Century Standards of Learning* (Stanford, CA: Stanford University, Stanford Center for Opportunity Policy in Education, 2010), [scale.stanford.edu/system/files/beyond-basic-skills-role-performance-assessment-achieving-21st-century-standards-learning.pdf](http://scale.stanford.edu/system/files/beyond-basic-skills-role-performance-assessment-achieving-21st-century-standards-learning.pdf).

## Resources

### The Gordon Commission on the Future of Assessment in Education

[www.gordoncommission.org](http://www.gordoncommission.org)

### NASBE Study Group Report—Reform at a Crossroads: A Call for Balanced Systems of Assessment and Accountability

[www.nasbe.org/wp-content/uploads/SG\\_Assessment\\_Report\\_20091.pdf](http://www.nasbe.org/wp-content/uploads/SG_Assessment_Report_20091.pdf)

### Partnership for Assessment of Readiness for College and Careers (PARCC)

[www.parcconline.org](http://www.parcconline.org)

### Stanford Center for Opportunity Policy in Education (Scope):

[edpolicy.stanford.edu](http://edpolicy.stanford.edu)

### Smarter Balanced Assessment Consortium

[www.smarterbalanced.org](http://www.smarterbalanced.org)

### Innovation Lab Network

[www.ccsso.org/resources/programs/innovation\\_lab\\_network.html](http://www.ccsso.org/resources/programs/innovation_lab_network.html)