



Developing Assessments of Deeper Learning: The Costs and Benefits of Using Tests that Help Students Learn

By Linda Darling-Hammond and Frank Adamson

The authors wish to thank the William and Flora Hewlett Foundation for funding this study as part of its commitment to deeper learning that prepares students for the ever-increasing demands of a rapidly changing world.

The Stanford Center for Opportunity Policy in Education (SCOPE) fosters research, policy, and practice to advance high quality, equitable education systems in the United States and internationally.

© 2013 Stanford Center for Opportunity Policy in Education. All rights reserved.

Citation: Darling-Hammond, L., Adamson, F. (2013). *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.



Stanford Center for Opportunity Policy in Education

Barnum Center, 505 Lasuen Mall
Stanford, California 94305
Phone: 650.725.8600
scope@stanford.edu
<http://edpolicy.stanford.edu>
@scope_stanford

Table of Contents

Executive Summary	i
Overview and Purpose	1
Concerns with Current US Tests	2
New Directions for Assessment	4
The Challenges for New Assessments.....	8
What Do We Actually Spend for Testing Today?	9
Considering Benefits as Well as Costs	12
Realizing the Benefits of High-Quality Assessments	13
Developing a Vision of a High-Quality Assessment System.....	13
Making High-Quality Assessments Affordable	14
Making High-Quality Assessments Feasible	15
Conclusion	18
Appendix A	19
Appendix B	21
Endnotes	27

Executive Summary

Despite a growing consensus that students must acquire higher-order thinking and performance skills to succeed in today's world, current US tests, which rely heavily on multiple-choice items, measure primarily low-level knowledge and skills. A recent RAND Corporation study found, for example, that fewer than 2% of mathematics items and only about 20% of English language arts (ELA) items on state tests ask students to analyze, synthesize, compare, critique, investigate, prove, or explain their ideas—the kinds of higher-order skills that students most need to become college- and career-ready.

Such skills are incorporated in the new Common Core State Standards (CCSS), which most states have adopted to ensure that all students graduate from high school prepared to succeed in the knowledge-based world they are entering. High-achieving nations have moved to place these skills at the center of their curriculum plans and assessments, using essays, open-ended problems, and performance tasks to evaluate students' abilities to think critically and solve problems.

The United States could take a major step in improving the direction of curriculum and assessments for deeper learning by drawing from the work of the Partnership for Assessment of Readiness for College and Careers (PARCC) and the Smarter Balanced Assessment Consortium (SBAC)—two multi-state consortia that were formed to develop next generation assessments of Common Core state standards. If the consortia are able to live up to their plans, the quality of assessment could improve substantially. An analysis of the content specifications for SBAC, for example, found that more than two-thirds of the assessment targets in ELA and mathematics intend to tap higher-level skills that are largely ignored in today's tests.

But high-quality assessments like those used in other countries tend to cost more than lower-quality assessments, in part because performance tasks and essays often require human scoring, whereas low-level skills can be measured with multiple-choice items that are cheap to score. Many states have traditionally budgeted only about \$20 per pupil for tests in math and reading, placing severe limits on the quality of learning their tests can measure. This represents less than two-tenths of 1% (.002) of average per-pupil spending on K-12 education. Given that most of us spend at least \$500 a year to assess the health of our automobiles, it is clear that this tiny investment carries a disproportionately large burden for the health of the education system.

However, states and districts together spend much more than this on all they do to increase scores on the end-of-year tests. Recent estimates put average state spending on ELA and math tests at \$25 to \$27 per pupil (with a range from about \$13 to \$105 per pupil), and spending on interim and benchmark testing at an additional \$17 to \$18 per pupil, not counting the costs of test preparation materials, personnel for test administration and analysis, or teacher time for scoring or professional development associated

with this testing. The combined costs of state and local testing in ELA and mathematics alone exceed \$50 per pupil on average.

This level of spending could support higher quality assessments that include the kinds of open-ended items and performance tasks that can measure more complex learning, scored both by teachers and by the evolution of more sophisticated artificial intelligence (AI) engines. Ironically, though, because these billions of dollars are largely pointed at boosting performance on narrow tests that do not measure or encourage the acquisition of higher-order skills, they do not result in the improvements to learning that would be possible if the same funds were spent differently.

From a cost-benefit perspective, this approach is penny wise and pound foolish. Although they may appear low in costs, today's testing programs are generally not organized to produce the benefits of deeper student learning found in high-performing countries. Instead we have a set of fragmented, disjointed efforts, unable to measure the most important learning goals, and not useful to teachers' efforts to understand how their students think and what could be done to support their success.

Current investments, which still total less than half of 1% of overall per-pupil spending, could support much higher-quality assessments, including performance tasks that tap critical thinking and problem solving skills, if they were refocused to do so. A wise use of resources would make it possible to develop a much more coherent system that not only provides assessments of deeper learning, but also offers formative supports for instruction and interim tools teachers could use to see how students are doing on tasks that reveal both how they think and what they know.

In order to realize the benefits of an instructionally helpful system, it will be important to make high-quality assessments both affordable and feasible to implement while strengthening teaching and learning at the same time. States that seek such a system can achieve their goals by:

- Understanding how the state and local components of a high-quality assessment system can operate together to strengthen learning.
- Taking advantage of the cost savings associated with multi-state consortia and productive uses of technology for online delivery and efficient scoring and reporting.
- Involving teachers in developing and scoring assessments in ways that also support teachers' professional learning and improved instruction, making these investments doubly beneficial.
- Combining state and local resources strategically to make sound, coherent investments in higher quality assessments.

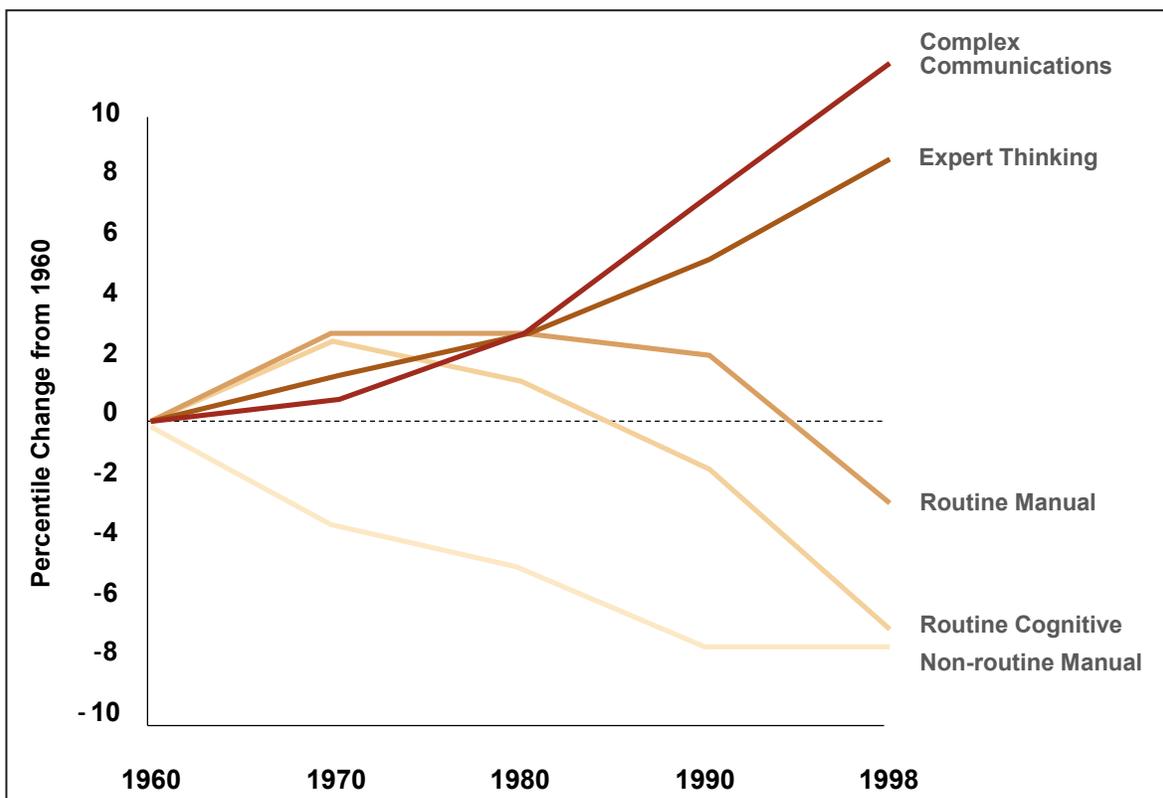
The question for policymakers has shifted from, "Can we afford assessments of deeper learning?" to, "Can the United States afford not to have such high-quality assessments?" The clear answer is that curriculum and instruction systems that include assessments of deeper learning are essential if students are to develop the skills they need for a knowledge society—a prerequisite for their success, and that of the nation.

Overview and Purpose

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, but whether they possess 21st century skills like problem-solving and critical thinking, entrepreneurship, and creativity.
— President Barack Obama, March 2009

As President Obama suggested, the changing nature of work and society in today's world places a premium not simply on students acquiring information but on their ability to analyze, evaluate, design, and create new solutions and products. By the year 2000, the top skills demanded by Fortune 500 companies had shifted from reading, writing, and arithmetic to teamwork, problem solving, and interpersonal skills.¹ Since 1970, employers' demands for workers with routine, repetitive skills—whether manual or cognitive—have dropped steeply, while demand for those with complex thinking and interactive skills has soared.² (See figure 1.)

Figure 1: The Demand for Job Skills



Source: Murnane & Levy (1996)

Responding to these societal needs, policymakers in nearly every state have adopted standards intended to ensure that all students graduate from high school ready for college and careers, with skills for the new economy. These skills include the abilities to analyze, synthesize, and apply knowledge to address new problems, design solutions, collaborate effectively, and communicate persuasively.

Such skills are incorporated in the new Common Core State Standards, just as they are in high-performing nations' curriculum and assessment systems. Policymakers in these nations understand that ambitious curriculum goals must be accompanied by high-quality student assessments that can support the improvements in student learning and school practice that they seek. As a result, Singapore, New Zealand, Hong Kong, the United Kingdom, a number of Australian states and Canadian provinces, and other high-achieving jurisdictions have introduced increasingly ambitious performance assessments that require students to find, evaluate, and use information, and to demonstrate what they know in tasks and projects that result in sophisticated written, oral, mathematical, physical, and multimedia products.³ (For examples, see Appendix A.)

These assessments—which call on students to design and conduct investigations, analyze data, draw valid conclusions, and report findings—seek to evaluate the 21st century skills identified as increasingly essential in a knowledge economy. They also require much deeper knowledge and greater application of learning than the traditional multiple-choice tests that have dominated in the United States over the last decade.

Concerns with Current US Tests

Achieving these goals will require a transformation in teaching, learning, and assessment so that all students develop the deeper learning competencies necessary for postsecondary success. The changes associated with these new expectations will be pronounced in the United States, where current tests tend to measure primarily low-level knowledge and skills. For example, a recent RAND Corporation study found that, on 17 states' tests, fewer than 2% of mathematics items and only 21% of ELA items reached the higher levels that ask students to analyze, synthesize, compare, connect, critique, hypothesize, prove, or explain their ideas.⁴ (In testing parlance, these are the skills measured at Depth of Knowledge levels 3 and 4. Levels 1 and 2 represent lower-level skills of recall, recognition, and use of routine procedures.⁵)

This study echoes the findings of other studies,⁶ and is even more worrisome, since these states were selected because their standards and tests were viewed as more rigorous than those of other states. The RAND study found that the level of cognitive demand was severely constrained by the dominance of multiple-choice questions, which the RAND authors found are rarely able to measure higher-order skills. Thus, the ambitious expectations found in state standards documents are frequently left unmeasured.⁷

Over the last decade, the nature of federal requirements under No Child Left Behind (NCLB) and the expansion of tested grades have led to the elimination of most state assessments requiring writing, research, and extended problem solving, and their nearly complete replacement with multiple-choice and short-answer tests. States abandoned performance assessments because of costs and the constraints on the types of tests that were approved. As a consequence, testing in most states is less focused on higher-order skills than it was in the 1990s, even though it now functions as the primary influence on curriculum and classroom instruction. Thus, while students in high-achieving nations are engaged in the kind of learning aimed at preparing to succeed in college and in the modern workplace, students in the United States are being drilled for multiple-choice tests that encourage recognition of simple right answers rather than production of ideas.

What and how tests measure matters, because when they are used for decision making, they determine much of what happens in the classroom. In the United States, students are tested far more frequently than in any other industrialized country, and test scores are used for more decisions about students, teachers, and schools. NCLB created a requirement for testing “every child, every year” in grades three through eight, plus at least once in high school. It also constrained the types of tests that could be used. By contrast, most countries test students at most once or twice before high school and some, like Finland, do not have any external tests before the 12th grade, other than tests that sample a small subset of students at a couple of grade levels.

Also in contrast with other countries, US tests are often used to determine the curriculum in which students can enroll and whether they are promoted or allowed to graduate; whether teachers are tenured, continued, or fired; and whether schools are rewarded or sanctioned, or even reconstituted or closed. With scores used to determine so many decisions, the incentives for teachers to “teach to the test” have become increasingly intense.⁸ In most other countries, tests are used to inform curriculum improvement and professional development, and to determine student pathways after middle or high school, but not to serve as arbiters of graduation, personnel decisions, or school sanctions and survival. Tests are taken seriously, but there is much more room for school-based assessment, scored by teachers, that counts in the system and enables assessments that promote deeper learning.⁹

... while students in high-achieving nations are engaged in the kind of learning aimed at preparing to succeed in college and in the modern workplace, students in the United States are being drilled for multiple-choice tests that encourage recognition of simple right answers rather than production of ideas.

High-performing jurisdictions have been moving proactively to increase their teaching and assessment of inquiry and problem solving. Their educational investment strategies, which have yielded higher and more equitable levels of performance and rapidly increasing levels of educational attainment, are intended to support career and college readiness, and appear to do so. Where instruction focuses on assessment content, it is of paramount importance that tests actually test students on the deeper learning skills that they require in the 21st century. As a recent report from the National Research Council noted:

The extent to which [deeper learning] goals are realized in educational settings will be strongly influenced by their inclusion in district, state, and national assessments, because of the strong influence of assessment on instruction in the United States.¹⁰

New Directions for Assessment

The United States is poised to take a major step in the direction of curriculum and assessments for this kind of deeper learning with the adoption of new Common Core State Standards in more than 40 states. Two state consortia—the Partnership for Assessment of Readiness for College and Careers (PARCC) and the Smarter Balanced Assessment Consortium (SBAC)—were formed to develop next-generation assessments of these standards, which will be launched in 2014-15. As states are increasingly able to work collaboratively on problems of policy and practice, other initiatives, such as the Innovation Lab Network (ILN) of states and districts, coordinated by the Council of Chief State School Officers, are also developing strategies to create more intellectually ambitious assessments that are more internationally comparable.

Whereas items on current state tests represent mainly recall and recognition, the new Common Core assessments under development will have many more tasks that require students to analyze, critique, evaluate, and apply knowledge. The plans for the new consortia assessments will increase cognitive expectations by many orders of magnitude. An analysis of the Content Specifications for the SBAC found, for example, that 68% of the assessment targets in ELA and 70% of those in mathematics intend to tap these higher-level skills.¹¹

It seems clear from the sample tasks that have been released by the two consortia that the new tests intend to encourage instruction aimed at helping students acquire and use knowledge in more complex ways. (See Exhibits 1 and 2 on the following pages.)

Exhibit 1: Mathematics Performance Tasks

SBAC Sixth-Grade Task: Planning a Field Trip

Classroom Activity: The teacher introduces the topic and activates students' prior knowledge of planning field trips by:

- 1) Leading students in a whole-class discussion about where they have previously been on field trips or other outings with their school, youth group, or family.
- 2) Creating a chart showing the class's preferences by having students first list and then vote on the places they would most like to go on a field trip, followed by whole-class discussion about the top choices.

Student Task: Individual students:

- 1) Recommend where the class should go on a field trip based on their analysis of the class vote.
- 2) Determine the per-student cost of going on a field trip to three different locations, based on a chart showing the distance and entrance fees for each option, plus a formula for bus charges.
- 3) Use information from the cost chart to evaluate a hypothetical recommendation about going to the zoo.
- 4) Write a note to their teacher recommending and justifying which field trip the class should take based on an analysis of all available information.

PARCC High School Task: Golf Balls in Water

Part A: Students analyze data from an experiment involving the effect on the water level by adding golf balls to a glass of water in which they:

- 1) Explore approximately linear relationships by identifying the average rate of change.
- 2) Use a symbolic representation to model the relationship.

Part B: Students suggest modifications to the experiment to increase the rate of change:

Part C: Students interpret linear functions using both parameters by examining how results change when a glass with a smaller radius is used by:

- 1) Explaining how the y-intercepts of two graphs will be different.
- 2) Explaining how the rate of change differs between two experiments.
- 3) Using a table, equation, or other representation to justify how many golf balls should be used.

Source: Herman & Linn (2013). See also, http://ccsstoobox.agilemind.com/parcc/about_highschool_3834.html and <http://www.smarterbalanced.org/wordpress/wp-content/uploads/2012/09/performance-tasks/fieldtrip.pdf>.

Exhibit 2: English Language Arts Performance Tasks

PARCC Seventh-Grade Task: Evaluating Amelia Earhart's Life

Summary Essay: Using textual evidence from the *Biography of Amelia Earhart*, students write an essay to summarize and explain the challenges Amelia Earhart faced throughout her life.

Reading/Pre-Writing: After reading Earhart's *Final Resting Place Believed Found*, students:

- 1) Use textual evidence to determine which one of three given claims about Earhart and her navigator, Noonan, is most relevant to the reading.
- 2) Select two facts from the text to support the claim selected.

Analytical Essay: Students:

- 1) Read a third text called *Amelia Earhart's Life and Disappearance*.
- 2) Analyze the evidence presented in all three texts concerning Amelia Earhart's bravery.
- 3) Write an essay, using textual evidence, analyzing the strength of the arguments presented about Amelia Earhart's bravery in at least two of the texts.

SBAC 11th-Grade Task: Nuclear Power—Friend or Foe?

Classroom Activity: Using stimuli including a chart and photos, the teacher prepares students for Part 1 of the assessment by leading students in a discussion of the use of nuclear power. Through discussion:

- 1) Students share prior knowledge about nuclear power.
- 2) Students discuss the use and controversies involving nuclear power.

Part 1: Students complete research and pre-writing activities in which they:

- 1) Locate and take notes on a series of internet sources about the pros and cons of nuclear power in order to brief a congresswoman.
- 2) Respond to two constructed-response questions that ask students to analyze and evaluate the credibility of the arguments in favor and in opposition to nuclear power.

Part 2: Students individually compose a full-length, argumentative report for their congressperson in which they use textual evidence to justify the position they take (pro or con) on whether a nuclear power plant should be built in their state.

Source: Herman & Linn (2013). See also: <http://www.parcconline.org/samples/english-language-artsliteracy/grade-7-elaliteracy>. <http://www.smarterbalanced.org/wordpress/wp-content/uploads/2012/09/performance-tasks/nuclear.pdf>

Even these more ambitious assessments, conducted in one- or two-day sessions, do not measure all of the CCSS skills such as extended writing and research, oral communications, collaboration, uses of technology for investigation, modeling solutions to complex problems, and multimedia presentations. A growing number of countries evaluate these skills, requiring students to design and finish complex projects that may take many days or weeks to complete, and that require considerable student planning, perseverance, and problem solving. The products of this work are evaluated by teachers, with moderation that produces reliable scoring, and are included in examination results. (For examples, see Appendix A.)

Many high-achieving nations see these assessments as key to transforming their education systems. For example, as Singapore prepared to revamp its assessment system, then-Education Minister Tharman Shanmugaratnam noted:

[We need] less dependence on rote learning, repetitive tests and a ‘one size fits all’ type of instruction, and more on engaged learning, discovery through experiences, differentiated teaching, the learning of life-long skills, and the building of character, so that students can ... develop the attributes, mindsets, character and values for success.¹²

In the United States, researchers found that when similar performance assessments were launched in many states during the 1990s, instruction was strengthened and student performance was increased on measures of higher-order skills, especially when teachers were involved in scoring the assessments and reflecting together about how to improve curriculum and teaching.¹³ Unfortunately, most of these assessments were discontinued because of the countervailing requirements of NCLB, which increased costs because of annual testing requirements and constrained the forms of assessment that could be used.

Understanding the importance of assessments to support the teaching of these higher-level skills, a number of states and districts are planning to augment the consortia assessments with more extended assessments like those used before NCLB and abroad. These performance assessments will be used as formative tools to guide classroom instruction, as components of state assessment systems, in proficiency assessments that replace seat time expectations, as elements of end-of-course examinations, and in graduation portfolios.

The Challenges for New Assessments

The challenge ahead will be for states and districts to prepare to implement new assessments, given the many changes they will entail. On the one hand, there is substantial consensus that US assessments must evolve to meet the new expectations for student learning. On the other hand, there are countervailing pressures regarding funding, time, and traditions that could stand in the way of assessment changes.

Especially given the current financial situation in most states, there is concern that new assessment designs need to be as cost effective and efficient as possible. At the same time, it is important for these assessments to encourage productive approaches to teaching and learning. High-quality assessments have tended to cost more than lower quality assessments, primarily because performance tasks and essays often require human scoring, whereas low-level skills can be measured with multiple-choice questions that are cheap to score.

Open-ended assessments — essay exams and performance tasks — are more expensive to score, but they can support more ambitious teaching and learning.

Across the United States, education spending averages just over \$10,000 per pupil and reaches \$18,000 per pupil in the highest-spending states. Yet many states have traditionally budgeted only about \$20 per pupil for tests in math and reading; as a result, they have had to restrict their tests to multiple-choice methods. This investment represents less than two-tenths of 1% of per-pupil spending on K-12 education. To get a sense of the relative size of this investment, think about the fact that most of us spend at least \$500 per year for routine check-ups on our automobiles, at least 20 times more than many states spend finding out if their students can read, write, communicate, and use mathematics well.

This tiny investment in testing has enormous influence on instruction, given the accountability policies that attach important consequences to scores. Multiple-choice tests, while inexpensive, produce few incentives to encourage instruction focused on higher-order thinking and performance skills. Open-ended assessments—essay exams and performance tasks—are more expensive to score but they can support more ambitious teaching and learning.

In addition, the apparently economical prices currently attached to state tests do not capture the considerable hidden expenditures on interim and benchmark assessments and other test preparation materials and activities that states, districts, and schools spend in current systems. Because of the nature of current summative tests, these materials focus on lower-level skills measured in limited ways. Also missing from the cost

calculus are the missed opportunities for assessing and developing higher-order thinking skills, which have consequences for students' learning and abilities.¹⁴

From a cost-benefit perspective, this approach is penny wise and pound foolish. Constraining our assessments to instruments that can only measure low-level learning, and then tying decision making that will drive virtually all instructional efforts to what they measure, is a recipe for low-quality schooling. Although they may appear to be low in costs, most states' current testing systems are not organized to produce the learning benefits found in high-performing countries that assess students differently.

What Do We Actually Spend for Testing Today?

Although cost concerns have narrowed the design of most state tests, districts and schools currently spend quite a bit more to boost achievement on those tests. Two independent studies, both published in 2012, estimated the average costs for state ELA and math tests at \$25 to \$27 per pupil, with a very wide range: A few states reported spending in the range of \$10 to \$15 per pupil and some reported spending more than \$50 per pupil.¹⁵ A Brookings Institution study found that the lowest spending states were Oregon, Georgia, and California. On the upper end, Massachusetts, which uses more open-ended items, spent \$64 per pupil, and tiny Hawaii spent just over \$100 per pupil.¹⁶

The Brookings study estimated that tests required under NCLB cost \$723 million annually, paid to vendors who administer the tests. However, with the addition of other subject areas and in-house spending, these costs more than doubled, to an estimated \$1.7 billion annually. This represents one-quarter of 1% of annual national K-12 education spending.

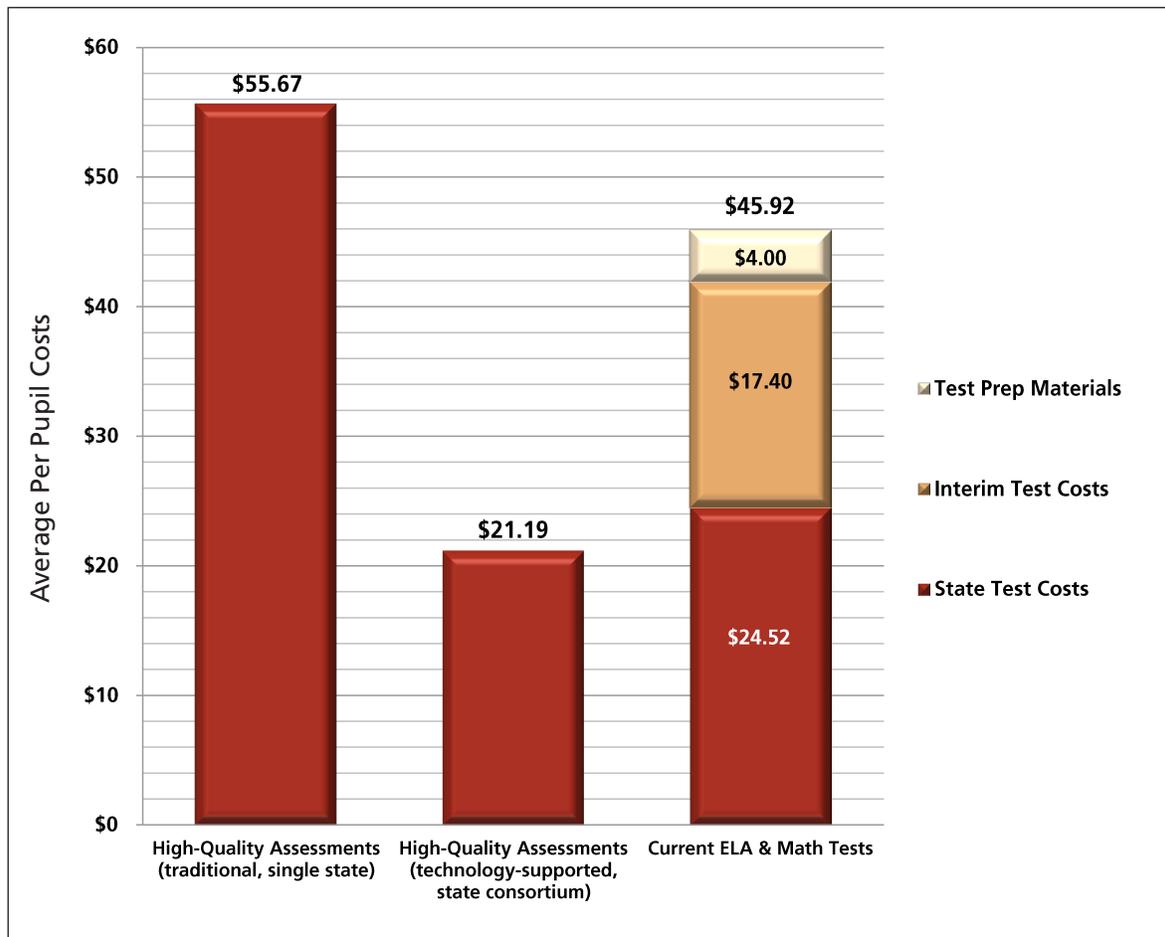
Beyond these costs, however, some states and nearly all local districts are making considerable investments in interim and benchmark assessments, test preparation materials and programs, and interventions to improve scores. Assessment Solutions Group (ASG) has sought to estimate some of these costs that can be readily captured. Both the ASG study¹⁷ and another recent study conducted by the American Institutes of Research¹⁸ found that all the districts they surveyed used interim or benchmark tests, usually provided by or constructed with a vendor.

These tests carry noticeable costs. For example, one widely used online interim test cost \$12.50 per student per year.¹⁹ Districts also pay for data management systems and, often, test preparation materials designed to improve scores. Test preparation programs, ranging in cost from about \$2 to \$8 per pupil, provide practice for the standardized tests.²⁰

Based on data from 189 districts and one state that sponsor statewide interim tests, ASG found that interim testing costs ranged from about \$6 to \$60 per pupil in 2011, with an

average of \$17 to \$18 per pupil. These estimates do not include formative assessments; test preparation materials; staff time for development, administration, analysis of the data; or professional development related to the assessments, as districts could not easily estimate these costs. By ASG estimates, the average costs of state tests and local benchmark assessments in ELA and mathematics are about \$42 per pupil. Adding the average costs of test preparation materials would bring this estimate to \$46 per pupil. (See figure 2.)

**Figure 2: Average Costs for State and Local Tests
(Per-Pupil Costs for ELA and Math)**



Sources: Topol et al. (2010); Topol (2012).

Topol and colleagues note:

All in all, most states and their districts are spending \$35 - \$55 per student on testing [for ELA and mathematics], not counting any of the related human resource and other time that goes into the testing and test preparation, professional development, data analysis, interventions and supplemental education services designed to raise scores, all of which are pointed at relatively narrow kinds of learning that have a dubious relationship to the skills and abilities students are now being called upon to acquire. In total, these investments amount to many billions of dollars of educational investment that may not be leveraging the kinds of instruction the Common Core standards and understandings of 21st century skills require.²¹

Appendix B illustrates these costs for three states, selected to represent the range of current spending on tests. California is currently one of the lowest-spending states, Kentucky is a mid-spending state, and Massachusetts is one of the highest-spending states. In brief, as Table 1 shows, combined state and local spending in these three states on NCLB-required ELA and mathematics testing ranges from \$31 to \$87 per pupil, with Kentucky near the national average, at \$40. The state-provided data for Kentucky does not include the additional costs of interim testing (e.g. data systems, materials, teacher scoring) reported in the surveys for the other two states. Based on the reports of districts that could calculate these costs, a conservative estimate would be an additional 20% on top of the interim testing costs that are paid to test vendors. This would bring Kentucky's total for ELA and math testing at the state and local level to approximately \$43 per pupil, very close to the ASG national estimate of \$42 per pupil.

None of these estimates include the amounts that states and localities spend for test preparation materials (about \$4 per pupil on average), formative assessments and tools, local teacher scoring, and professional development associated with testing, or curriculum development associated with assessments. Adding these costs places the average costs of state and local testing in the areas of ELA and mathematics at more than \$50 per pupil.

Costs for testing in the full range of subjects can be significantly higher, depending on the number of tests, the vendor chosen, and the extent to which the tests use open-ended questions. In Kentucky, for example, which has a system of end-of-course examinations at the high school level, plus statewide use of several ACT testing products, the total costs for state and local testing are approximately \$67 per pupil. In Massachusetts, costs would certainly exceed the \$87 per pupil that is spent for ELA and math.

Table 1: Estimated Costs Per Pupil for State Tests Plus Local Interim Testing

	California		Kentucky		Massachusetts	
Interim Testing Costs	\$14.71		\$15.04*		\$23.39	
	All Tests	ELA + Math	All Tests	ELA + Math	All Tests	ELA + Math
State Testing Costs	\$19.64	\$16.63	\$51.68	\$24.95	NA	\$63.75
State + Local Costs	\$34.35	\$31.45	\$66.72	\$39.99	NA	\$87.39

*Kentucky’s estimate for interim testing costs, provided from a state data set, does not include costs beyond those allocated directly for vendor test fees (e.g., data management systems, teacher scoring).

Considering Benefits as Well as Costs

Because of pressures to “teach to the test” in high-stakes accountability systems, the additional costs for interim and benchmark testing have become viewed as mandatory. However, in many cases they may not improve the quality of assessments or leverage higher quality instruction, because they are focused on raising scores on current state tests, which measure mostly low-level skills.

Many analysts have found that intensive teaching to narrow tests has reduced time for instruction in topics and important subjects that are untested; has focused instruction in tested areas on the multiple-choice and short-answer formats of the test, rather than supporting more intellectually rigorous approaches; and has reduced the emphasis on writing, oral communications, extended problem solving, research, and investigation—all abilities that are critical for college- and career-readiness.²² The tangible expenditures on testing, as well as the costs to instruction, have not been considered in discussions of what kinds of assessment might be affordable as learning goals change.

This raises an important question: *Might our nation’s schools be able to improve test quality and better invest valuable resources with a more integrated, higher quality assessment system, especially if resources currently spent in uncoordinated, fragmented ways can be coordinated to support a curriculum, teaching, and assessment system that focuses on higher-order skills?*

Realizing the Benefits of High-Quality Assessments

In order to realize the benefits of high-quality assessments, it will be important to figure out how to make them both affordable and feasible to implement in state and local school systems. Doing this will depend on:

- Having a vision of a high-quality assessment system and how it can operate to strengthen learning;
- Taking advantage of cost savings associated with consortia and productive uses of technology;
- Involving teachers in scoring assessments in ways that also support teacher learning and improved instruction;
- Being strategic about combining state and local resources to make sound, coherent investments in high-quality assessments.

Developing a Vision of a High-Quality Assessment System

Over a number of years, the Council of Chief State School Officers (CCSSO) has worked with key stakeholders to develop a set of principles for student assessment systems. These principles suggest that the student assessment process should be considered as a system that supports a variety of purposes, such as informing learning and instruction, determining progress, measuring achievement, and providing information for accountability. A CCSSO report²³ outlined key elements of high-quality assessment systems:

- 1) Standards guide an integrated system of curriculum, assessment, instruction, and teacher development.** Formative tools, along with interim and summative assessments, are connected to a common curricular vision tied to professional learning. Thus, everything that informs the work of schools is well aligned and pulling in the same direction.
- 2) A balanced set of assessments offers evidence of student performance on challenging tasks that evaluate applications of knowledge and skills.** State and local assessments work together to evaluate a broad array of competencies that generalize to higher education and careers. Components of the assessment system evaluate students' abilities to find, analyze, and use resources; communicate in multiple forms; use technology; collaborate with others; and frame and solve complex problems.
- 3) Teachers are integrally involved in the development of curriculum and the development and scoring of assessments.** The assessment systems are designed to increase the capacity of teachers to prepare students for the demands of college and careers by involving them in moderated scoring of the assessments, which enables them to deeply understand the standards and develop stronger curriculum and instruction.

- 4) Assessment measures are designed to improve teaching and learning. Successful systems emphasize the quality of assessments over the quantity. They invest in a set of formative and summative approaches that support the learning of ambitious intellectual skills in the classroom. Assessment as, of, and for learning is made possible by:
- a) the use of curriculum-embedded assessments, which provide models of good curriculum and assessment practice, enhance curriculum equity across classrooms, and allow teachers to evaluate student learning in support of teaching decisions;
 - b) close examination and scoring of student work as sources of ongoing professional development;
 - c) use of learning progressions that allow teachers to see where students are on multiple dimensions of learning and to strategically support their progress.
- 5) Accountability systems are designed to evaluate and encourage multiple dimensions of student success. Both student assessments and school accountability are based on multiple measures. Beyond test data, school indicators may include student portfolios and exhibitions of competence, student participation in challenging curricula, progress through school, graduation rates, college attendance, citizenship, a safe and caring climate, and school improvement. When evaluating schools, many nations include information from school inspections in which experts examine teaching, learning, and school operations up close in order to diagnose school needs and guide improvement efforts.

Making High-Quality Assessments Affordable

Fortunately, recent advances in a number of areas have made renewed efforts to implement performance-based assessments more financially viable. Cost savings are associated with advances in computer-based scoring of open-ended items and tasks plus more economical approaches to teacher scoring, as well as the savings of computer-administered assessments, and the economies of scale that states realize as part of a consortium.

In a recent study, ASG found that the cost of higher quality assessments—that is, tests that replace half of the usual multiple-choice items with open-ended items and performance tasks—can be significantly reduced by:

- 1) participating in a state assessment consortium,
- 2) using online delivery of assessments, and
- 3) scoring open-ended items and tasks through both computer-based scoring and by using teachers who are paid professional development stipends.²⁴

Based on data ASG collected, the \$55 estimated as the cost for a state to administer high-quality assessments would be reduced by more than half if all of the cost-saving approaches were applied. (See figure 2.) Thus, increasing assessment quality appears to be an achievable goal.

In these estimates, the greatest savings are realized by being part of a consortium and by delivering assessments online, as both of the two assessment consortia plan to do, rather than in paper-and-pencil format. Computer-based scoring of some items also realizes savings. For complex items, the costs of AI scoring (computer-based scoring that uses artificial intelligence engines) and scoring by teachers are currently about the same when teachers are paid professional development stipends. These professional development stipends are less expensive than vendor scoring.²⁵

AI scoring may become more cost-efficient over time while still allowing for items that are sufficiently complex to measure the standards. Getting to that point will require additional research tied to extensive experience in human scoring of complex tasks, because analysis of this scoring experience provides the means for training an AI engine. In the meantime, there are many benefits to including teachers in the scoring of performance-based assessments, as described in the next section.

Making High-Quality Assessments Feasible

If we believe that assessment and accountability systems should encourage the kinds of learning students will need for later success, it will be important to figure out how to develop and use what Lauren Resnick has termed “tests worth teaching to.”²⁶ To realize the benefits of high-quality assessments and to make them feasible in the United States, at least two things are needed:

- **Strategies for involving teachers in scoring to increase the benefits and reduce the costs of high-quality systems.**

High-achieving systems in many countries increase the benefits of performance assessments and offset the costs by engaging teachers in developing, reviewing, scoring, and using the results of assessments. Comparability in scoring is achieved through the use of standardized rubrics, as well as training, moderation, and audit systems that support consistency in scoring.²⁷

The moderation process in which teachers learn to calibrate their scoring to a standard is a strong professional learning experience, as it enables teachers to understand deeply the standards and nature of student performance they require. This sparks conversations about revisions to curriculum and teaching. As teachers become more skilled at using new assessment practices and developing curriculum, they become more effective at teaching the standards.

Thus, the assessment systems increase the capacity of teachers to prepare students for the demands of college and careers in this new century and global society.

These teacher-scoring strategies are comparable to those that are currently used in the Advanced Placement and International Baccalaureate programs in the United States as well as those used in states that launched leading-edge assessment systems during the 1990s, such as Connecticut, Kentucky, Maine, Maryland, New York, and Vermont. Research about the outcomes of these systems has found that they can be reliably scored and support growth in teaching and gains in student outcomes on assessments of higher-order skills.²⁸

Like European and Asian countries, New York State has long maintained its Regents examination program, which includes open-ended essays and tasks, by setting aside days for teacher scoring. 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.²⁹

States could make scoring of open-ended assessments more affordable by creatively using existing professional development days and utilizing existing incentives, such as continuing education units that could be offered for engagement in scoring and associated discussions of curriculum and instruction. This would link professional development and assessment budgets in a cohesive project that helps teachers improve instruction, assessment, and delivery of standards.

As noted earlier, technology can be used to make new assessments more efficient and affordable in a number of ways, including: by delivering the assessments, by scoring the results in some cases, and by delivering responses to teachers who are trained scorers so they can evaluate them from an electronic platform. Such a platform can also support calibration of scorers and moderation of scores. These technologies are already being used in the International Baccalaureate and Hong Kong assessment systems, both of which include open-ended tests along with classroom-based papers and projects in their examination systems.

In order to gain the cost benefits of machine scoring and the teaching and learning benefits of teacher-moderated scoring, a mixed system could be developed. Computer-based scoring can be utilized for constructed-response tasks where useful, while teachers score a proportion of these tasks for anchoring and learning purposes. Meanwhile, teachers can be engaged to score tasks that require more nuanced professional judgments, which will also support instruction focused on more challenging forms of learning.

- **Strategies for pooling resources to enable more integrated systems of formative and summative assessment that better represent higher-order skills.**

As we have seen, although considerable state and local resources are currently spent on testing, these investments generally do not leverage higher quality teaching or assessment, because they are focused almost exclusively on boosting scores on tests of lower-level skills. Furthermore, investments from states, districts, and schools are fragmented and uncoordinated, so they do not support alignment of effort or economies of scale.

The goal for states should be to create the kind of teaching and learning systems that many other countries have built to provide an integrated approach to curriculum, instruction, assessment, and teacher development. The new assessment consortia—PARCC and SBAC—could offer a first step in this direction, as they seek to build coherent systems of formative supports, along with interim and summative assessments that include greater focus on higher-order skills.

To take advantage of these possibilities and further enhance the quality of learning, states will need to think about their investments in standards as more than the line item for testing in the state budget. They will need to connect interagency planning and to pool resources for curriculum, assessment, and professional development so that these elements are mutually reinforcing.

Conclusion

Testing in the United States has been shaped by pressures to test frequently and inexpensively; as a result, studies have found that most current state tests focus almost exclusively on multiple-choice questions that measure low-level skills.

Educators and policymakers know that assessments need to evolve to support college and career readiness. To meet the demands of a knowledge-based economy and the expectations represented in new Common Core State Standards, assessments must better represent more complex competencies.

Currently, the average state-testing system in reading and mathematics costs \$25 to \$27 per pupil. However, the pressures of meeting accountability requirements have caused states and localities to add additional interim and benchmark tests, as well as spending for data systems and test preparation. In combination, these expenditures now average more than \$50 per pupil.

These resources, which still total less than 1% of overall per-pupil spending, could support much higher quality assessments, including performance tasks that include critical thinking and problem solving skills, if the assessments were refocused to do so. To shift our current systems of assessment, states and districts will need to support higher quality assessments through a combination of:

- Cost savings, such as the economies of scale enabled by state consortia, online delivery, and efficient scoring of open-ended tasks by teachers and computers.
- Strategic reallocation of resources that are currently used for state and local testing in fragmented ways and that are not focused on improved assessment quality.
- Use of professional learning time and incentives to support teacher engagement in assessment scoring, development, and use, which provides the double benefit of improved instruction and more efficient use of resources.

The question for policymakers has shifted from, “Can we afford high-quality assessments of deeper learning?” to, “Can the United States afford not to have high-quality assessments?” The answer is that assessments of deeper learning are needed to provide the impetus for students to develop skills for the knowledge economy, as a prerequisite for global competitiveness, and for the long-term well-being of the nation.

Appendix A: Examples of International Assessments of Higher-Order Skills

Around the world, high-achieving countries, such as Finland, Singapore, Australia, New Zealand, and Canada, use multiple sources of evidence to evaluate the skills and knowledge needed to meet the demands of today's dynamic, technological era. Students engage in a variety of tasks and tests that are both embedded in the classroom curriculum and on-demand, providing many ways to demonstrate and evaluate their learning. Increasingly, these tasks require them to exhibit skills of investigation, analysis, communication, and collaboration that are called for in contemporary college and careers.

Exhibit A-1: Extended Experimental Investigations

Science Assessment in Queensland, Australia

In Queensland science courses, as in other Australian states and Singapore and Hong Kong, students must complete an extended experimental investigation that they design, conduct, and evaluate. In Queensland, the task is defined as follows:

Within this category, instruments are developed to investigate a hypothesis or to answer a practical research question. The focus is on planning the extended experimental investigation, problem solving, and analysis of primary data generated through experimentation by the student. Experiments may be laboratory or field based. An extended experimental investigation may last from four weeks to the entirety of the unit of work. The outcome of an extended experimental investigation is a written scientific report. For monitoring, the discussion/conclusions/evaluation/recommendations of the report should be between 1,500 and 2,000 words.

To complete such an investigation the student must:

- develop a planned course of action
- clearly articulate the hypothesis or research question, providing a statement of purpose for the investigation
- provide descriptions of the experiment
- show evidence of modification or student design
- provide evidence of primary and secondary data collection and selection
- execute the experiment(s)
- analyze data
- discuss the outcomes of the experiment
- evaluate and justify conclusion(s)
- present relevant information in a scientific report.

Source: Darling-Hammond & Wentworth, 2010, p. 11.

Exhibit A-2: Collaborative Projects Using Technology

Graduate Certificate in Secondary Education (GCSE) Task in Interactive Computer Technology

In England, students choose a number of domains in which to be examined as part of the high school assessment system. Most of these examinations, which are linked to high school courses, include a project-based component that typically counts for 60% of the total examination score. The project below has been used as part of the Interactive Computer Technology (ICT) examination.

Litchfield Promotions works with more than 40 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 ICT 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 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.
- 5) Produce a solution, ensuring that the following are addressed:
 - It can be modified to be used in a variety of situations.
 - It has a friendly user interface.
 - It is suitable for the target audience.
 - It has been fully tested.

Candidates 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 their own and others' work.

Source: Darling-Hammond & Wentworth, 2010, p. 30.

Appendix B: Spending for Interim Testing at the Local District Level

To estimate the expenditures made by local school districts for interim assessments, Assessment Solutions Group conducted surveys of a small sample of local districts in two states: California and Massachusetts. The samples were selected to include small, medium, and large districts that were geographically dispersed across each state and represented urban, suburban, and rural communities. The surveys were conducted through an interview process that also collected administrative record data where available. The protocol used to collect the data is included at the end of this appendix. ASG had also previously secured data on interim assessments for all of the districts in Kentucky, which is analyzed here by district deciles.

Although the ASG survey also asks about formative assessments and professional development costs, the data were deemed insufficiently precise and comparable across districts to be used for this analysis. Consequently, the analysis includes only the following categories of costs:

- Spending for interim or benchmark tests paid to test vendors;
- Costs for internal district development and scoring of interim tests;
- Assessment materials and data manuals;
- Expenditures for data management systems to store and analyze testing data; and
- Spending for item banks and other supports for the interim testing program.

In the following tables, we display these local interim testing data for sampled districts, along with state testing costs, drawn from an earlier ASG study.³⁰ Where available, total state testing costs are shown separately from costs for ELA and math testing required under NCLB.

California

California's state testing program is one of the least expensive in the country, costing about \$17 per pupil for assessments in ELA and mathematics, and just under \$20 per pupil for the entire state testing program, including science and social studies. Beyond the state program, an ASG survey of seven small, medium, and large districts found that all of them also were engaged in interim tests focused on improving scores on the state tests. These include a wide range of tests, from assessments developed by companies like NWEA and Pearson to locally developed tests and materials to support test preparation. Not including staff time for test administration, analysis of results, and professional development, the costs for this testing ranged from about \$7 to \$29 per pupil. At an average of nearly \$15 per pupil, the interim tests appear to cost nearly as much as the state tests themselves.

Total state and local costs for testing in ELA and math are estimated at \$31 per pupil. Adding in other state tests, the total is approximately \$34 per pupil.

California: State Testing Costs Plus District Costs for Interim and Benchmark Tests

	Total Students	Assessed Students*	Assessment Costs	Other Costs	Total Costs	Per Pupil Cost
CA State Tests	6,217, 002	5,233,539	\$78,554,442	\$24,234,000†	\$102,788,442	\$19.64
ELA & Math**	6,217, 002	5,233,539	\$66,029,442	\$21,010,000†	\$87,039,442	\$16.63
District A	5,472	3,952	\$88,248	\$4,250	\$92,498	\$23.41
District B	17,427	15,519	\$292,987	\$153,800+	\$446,787	\$28.88
District C	664,233	597,810	\$3,900,000	NA	\$3,900,000	\$6.52
District D	34,472	19,193	\$134,000	\$15,000	\$149,000	\$7.76
District E	131,784	105,490	\$222,025	\$517,500	\$739,525	\$7.01
District F	42,000	42,000	\$511,000	\$270,000	\$781,000	\$18.60
District G	18,500	18,500	\$160,000	\$40,000	\$200,000	\$10.81
7 District Average**						\$14.71
Estimated State plus Local Costs - All Subjects						\$34.35
State plus Local Costs - ELA & Math Only						\$31.34

* Where districts could not estimate the number of students served by interim tests, the number of total students is recorded as the number of assessed students.

**Includes STAR tests in ELA and Math, CAHSEE (High School Exit Exam), and CELDT (English Language Development tests).

†District Apportionments +Includes estimated costs for local scoring of interim writing assessments

++Unweighted average of district costs

Kentucky

The portion of Kentucky’s state testing program that includes NCLB-required tests in ELA and mathematics is right at the national average, costing about \$25 per pupil. It is augmented by other tests, including an extensive end-of-course exam system, which brings the state testing total to \$63 per pupil.

Beyond the state program, data collected by the Kentucky Department of Education for all 174 Kentucky districts shows that 93% of them reported costs for interim testing in 2011. Not including staff time for test administration, analysis of results, and professional development, the costs for this testing ranged from about \$1 to \$57 per pupil. At an average of \$15 per pupil, these local costs are about comparable to those for California. However, these estimates do not include “other” costs that ASG collected from its surveys, such as costs for data management systems and teacher scoring. Thus, these numbers underestimate local costs relative to the data from the other two states.

With this conservative estimate, total state and local costs for testing in ELA and math are estimated at \$87 per pupil.

Kentucky: State Testing Costs Plus District Costs for Interim and Benchmark Tests

	Assessed Students	Assessment Costs*	Other Costs ⁺	Total Costs	Per Pupil Cost
KY State Tests**	392,260	\$24,724,148	NA	\$24,724,148	\$51.68
ELA & Math	392,260	\$9,786,887	NA	\$9,786,887	\$24.95
10th Decile	21,260	\$851,737	NA	\$851,737	\$40.06
9th Decile	40,738	\$1,217,646	NA	\$1,217,646	\$29.89
8th Decile	29,851	\$652,459	NA	\$652,459	\$21.86
7th Decile	59,904	\$1,095,338	NA	\$1,095,338	\$18.28
6th Decile	81,558	\$1,083,203	NA	\$1,083,203	\$13.28
5th Decile	25,555	\$270,293	NA	\$270,293	\$10.58
4th Decile	31,436	\$236,871	NA	\$236,871	\$7.54
3rd Decile	27,435	\$139,955	NA	\$139,955	\$5.10
2nd Decile	33,222	\$92,977	NA	\$92,977	\$2.80
1st Decile	26,236	\$33,684	NA	\$33,684	\$1.28
Average Cost	377,195	\$5,674,163	NA	\$5,674,163	\$15.04
Estimated State plus Local Costs - All Subjects					\$66.72
State plus Local Costs - ELA & Math					\$39.99

* Includes cost data from the 161 of Kentucky’s 174 districts (93%) that reported any interim testing costs.

+ Data regarding other costs were not available for the state or districts in Kentucky. Many districts engage in local hand scoring of assessments, as well as support data systems for managing the data, but these data were not available in the state data set that was used to construct this table.

**Includes English language arts, mathematics, and science tests (together totaling \$27.29 per pupil), plus end-of-course examinations in 5 subjects, ACT Explore, ACT and WorkKeys. These estimates do not include the costs of scoring the Kentucky writing portfolio, which was recently discontinued.

Massachusetts

Massachusetts has one of the most expensive testing programs in the country. NCLB-required tests in ELA and mathematics include open-ended items in both reading and mathematics and cost about \$64 per pupil, based on data on vendor costs collected in a Brookings Institution study.³¹

Beyond the state program, an ASG survey of nine small, medium, and large districts found that all of them also were engaged in interim tests focused on improving scores on the state tests. Not including staff time for test administration, analysis of results, and professional development, these testing costs range from about \$8 to \$51 per pupil. At an average of about \$23 per pupil, both state and interim testing costs in Massachusetts appear higher than those in the other two states.

Total state and local costs for testing in ELA and math are estimated at \$40 per pupil. Adding in other state tests in ELA and math, the total is approximately \$79 per pupil. In this case, we do not have data about the total costs of tests that go beyond ELA and mathematics.

Massachusetts: District Costs for Interim and Benchmark Tests

	Total Students	Assessed Students	Assessment Costs	Other Costs	Total Costs	Per Pupil Cost
MA State Tests* ELA & Math		509,312	\$32,469,904	--	\$32,469,904	\$63.75
District A	24,200	24,000	\$1,233,000	--	\$1,233,000	\$50.95
District B	2,530	2,530	\$37,100	\$19,000	\$56,100	\$22.17
District C	12,355	9,000	\$154,500	\$43,500	\$198,100	\$22.01
District D	55,000	55,000	\$1,090,685	--	\$1,090,685	\$19.83
District E	5,398	5,398	\$87,455	--	\$87,455	\$16.10
District F	22,787	22,787	274,242	\$67,750	\$341,992	\$15.01
District G	5,600	5,600	\$14,200	\$54,585	\$68,785	\$12.28
District H	4,900	4,900	\$18,987	\$30,000	\$38,987	\$10.00
District I	6,222	6,222	\$48,000	--	\$48,000	\$7.71
9 District Average	138,992	135,637	\$2,958,269	\$214,835	\$3,173,104	\$23.39
Estimated State plus Local Costs - ELA & Math						\$87.39

*Estimated costs are based on Chingos (2012) p. 27, estimating spending for test vendor contracts in grades 3-9. These costs do not include Massachusetts high school exit exam, or any other high school tests.

ASG Survey on Assessment and Other Spending Data

1. RESPONDENT INFORMATION									
State:									
District:									
Contact Information:									
Name:									
Title:									
email:									
Telephone:									
K-12 Student Population	7,100								
	Grades K-2	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grades 9-12	Total
Grade Band Student Population	1,800	600	600	600	600	600	600	1,700	7,100

2. ASSESSMENT DATA:									
INTERIM ASSESSMENT TOTAL ANNUAL COSTS:									
Name/Vendor:									
Brief Description									
Assessment Mode: (PPT/CBT)									
Number of Administrations Per Year:									
List Domains Assessed:									
Vendor Created: (yes/no)									
		Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grade 9-12	Total
List number of students assessed by grade									
Comments:									
FORMATIVE ASSESSMENT TOTAL ANNUAL COSTS:									
Name/Vendor:									
Brief Description:									
Assessment Mode: (PPT/CBT)									
Number of Administrations Per Year:									
List Domains Assessed:									
Vendor Created: (yes/no)									

		Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grades 9-12	Total
List number of students assessed by grade									
Comments:									
OTHER ASSESSMENT TOTAL ANNUAL COSTS SUPPORTING SUMMATIVE ASSESSMENTS:									
Type/Name of Assessment									
Brief Description									
Assessment Mode: (PPT/CBT)									
Number of Administrations Per Year:									
List Domains Assessed:									
Vendor Created: (yes/no)									
		Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grades 9-12	Total
List number of students assessed by grade									
Comments:									

3. OTHER SPENDING:									
Test Preparation Materials Annual Cost									
List Domains:									
Vendor Created: (yes/no)									
		Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grades 9-12	Total
List number of students by grade									
NCLB Supplemental Education Services (SES) Annual Costs:									
For Tutoring, Remediation, Student Interventions		Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grades 9-12	Total
Costs by Grade									
Comments:									
Other Spending: List									
Instructional Resources related to test results									
Instructional Management Systems									
Professional development related to testing and use of assessment results to improve instruction.									
Comments:									

Endnotes

¹ Cassel, R.N., and Kolstad, R. (1998). The critical job-skills requirements for the 21st century: Living and working with people. *J. Instructional Psychology*, 25(3), 176-180; Creativity in Action (1990). Skills desired by Fortune 500 Companies (in order of importance). Buffalo, New York: Creative Education Foundation.

² Murnane, R. & Levy, F. (1996). *Teaching the new basic skills*. NY: The Free Press.

³ Darling-Hammond, L., & Adamson, F. (2010a). *Beyond Basic Skills: The Role of Performance Assessment in Achieving 21st Century Standards of Learning*. Stanford, CA: Stanford Center for Opportunity Policy in Education.

⁴ Yuan, K. & Le, V. (2012). *Estimating the Percentage of Students Who Were Tested on Cognitively Demanding Items Through the State Achievement Tests*. Santa Monica, CA: RAND Corporation.

⁵ Webb, N.L. (2002). *Depth-of-Knowledge Levels for Four Content Areas*. Retrieved August 24, 2011: <http://facstaff.wcer.wisc.edu/normw/All%20content%20areas%20%20DOK%20levels%2032802.doc>

⁶ See, for example, Polikoff, M.S., Porter, A.C., & Smithson, J. (2011). How Well Aligned Are State Assessments of Student Achievement With State Content Standards? *American Educational Research Journal*, 48 (4): pp. 965–995.

⁷ Polikoff, Porter, & Smithson (2011).

⁸ Amrein, A. L., & Berliner, D. C. (2002). High-Stakes Testing, Uncertainty, and Student Learning. *Education Policy Analysis Archives*, 10(18), 1-74. Accessed from <http://epaa.asu.edu/ojs/article/view/297>.

⁹ Darling-Hammond, L. & Wentworth, L. (2010). *Benchmarking Learning Systems: Student Performance Assessment in International Contexts*. Stanford, CA: Stanford Center for Opportunity Policy in Education.

¹⁰ Pellegrino, J.W., and Hilton, M.L., eds. *Education for Life and Work: Developing Transferable Knowledge and Skills in the 21st Century*. Washington, DC: National Academies Press, 2012.

¹¹ Herman, J.L. & Linn, R.L. (2013). *On the road to assessing deeper learning: The status of Smarter Balanced and PARCC assessment consortia*. (CRESST Report 823). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

¹² Ng, P. T. (2008). Educational reform in Singapore: From quantity to quality. *Education Research on Policy and Practice*, 7, 5-15.

¹³ For a summary, see Darling-Hammond, L. & Rustique Forrester, E. (2005). The Consequences of Student Testing for Teaching and Teacher Quality. In Joan Herman and Edward Haertel (eds.) *The Uses and Misuses of Data in Accountability Testing*. The 104th Yearbook of the National Society for the Study of Education, Part II, pp. 289-319. Malden, MA: Blackwell Publishing.

¹⁴ Picus, L., Adamson, F., Montague, W., & Owens, M. (2010). *A New Conceptual Framework Analyzing Costs Performance Assessment*. Stanford, CA: Stanford Center for Opportunity Policy in Education.

¹⁵ Assessment Solutions Group.

¹⁶ Chingos, M. (2012). *Strength in Numbers: State Spending on K-12 Assessment Systems*. Washington, DC: Brown Center on Education Policy at Brookings.

¹⁷ Topol, B., Olson, J., Roeber, E., & Hennon, P. (2012). *Getting to higher-quality assessments: Evaluating costs, benefits and investment strategies*. Stanford, CA: Stanford Center for Opportunity Policy in Education.

¹⁸ Heppen, J., Jones, W., Faria, A., Sawyer, K., Lewis, S., Horwitz, A., et al. (2012). *Using data to improve instruction in the Great City Schools: Documenting current practice*. Washington, DC, American Institutes for Research and The Council of Great City Schools.

¹⁹ Illinois Department of Education. (2011). *Action Report 0711-4209: Purchase of NWEA/MAP Assessments for Grades K-8*. Springfield, IL: Illinois Department of Education. Accessed from: [http://www.boarddocs.com/il/d365u/Board.nsf/files/8JHLSR57FE4F/\\$file/Measure%20of%20Academic%20Progress%20\(MAP\).pdf](http://www.boarddocs.com/il/d365u/Board.nsf/files/8JHLSR57FE4F/$file/Measure%20of%20Academic%20Progress%20(MAP).pdf)

²⁰ A popular preparation guide for the Texas state assessment, developed by Holt McDougal, charges \$7.15 per booklet; Practice Planet charges \$2 to \$8 per pupil; Study Island charges a minimum of \$3.70 per pupil; and MCAS Pass, aimed at the Massachusetts tests, charges \$4 to \$6 per pupil, depending on the number of students. Holt McDougal. (2010). *Holt McDougal: Get Ready for STAAR*. Accessed from http://www.hmh-education.com/tx/ws612/pdf/WS_STAAR.pdf; Practice Planet (2013). *Practice Planet: An online voyage to testing success!* <http://www.practiceplanet.com/docs/PPFlyer.pdf>; Study Island, (2013). Study Island. <http://info.edmentum.com/StudyIslandQuoteRequest.html>; MCAS Pass. (2013). *Basic Site License Pricing for Schools*. <http://johanglozman.com/testmastery/mcas/basicsite.html>.

²¹ Topol et al., 2012, p. 12.

²² For a summary, see Darling-Hammond, L. & Rustique Forrester, E. (2005).

²³ Darling-Hammond, L. (2010). *Performance Counts: Assessment Systems that Support High-Quality Learning*. Washington, DC: Council of Chief State School Officers and Stanford, CA: Stanford Center for Opportunity Policy in Education.

²⁴ Topol, B., Olson, J., & Roeber, E. (2010). *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.

²⁵ Topol et al. (2010).

²⁶ Resnick, L.B. (1987). *Education and learning to think*. Washington, DC: National Academy Press.

²⁷ Darling-Hammond, L., & Wentworth, L. (2010). *Benchmarking Learning Systems: Student Performance Assessment in International Contexts*. Stanford, CA: Stanford Center for Opportunity Policy in Education.

²⁸ Darling-Hammond, L. & Rustique Forrester, E. (2005).

²⁹ Pecheone, R. and Kahl, S. (2011). *Through a Looking Glass: Lessons Learned and Future Directions for Performance Assessment*. Stanford, CA: Stanford Center for Opportunity Policy in Education.

³⁰ Topol et al (2010).

³¹ Chingos (2012).



Stanford Center for Opportunity Policy in Education
Barnum Center, 505 Lasuen Mall
Stanford, California 94305
Phone: 650.725.8600
scope@stanford.edu
<http://edpolicy.stanford.edu>
@scope_stanford