BROWN CENTER on Education Policy at BROOKINGS

October 2013

STANDARDIZED TESTING AND THE COMMON CORE STANDARDS

You Get What You Pay For?

Reuters

Matthew M. Chingos

This report greatly benefited from helpful conversations with Beth Akers, Tony Alpert, Bryan Bleil, Scott Marion, Jeff Nellhaus, Morgan Polikoff, Candy Shyer, Paul Weeks, Russ Whitehurst, and Joe Willhoft. I also thank Ben Ehrlich, Laila Heid, Ellie Klein, and Amy Yeung for research assistance.

EXECUTIVE SUMMARY

Matthew M. Chingos
is a fellow in the Brown Center on Education Policy at the Brookings Institution.

Over the next two years, the biggest changes to standardized testing in the U.S. in more than a decade will be put in place as 45 states implement the Common Core standards in math and English language arts (ELA). These states, which comprise 85 percent of American students, will adopt new assessment systems aimed at measuring whether students are learning the material specified by the new standards, which are designed to ensure student readiness for success in college and the labor force. A potentially important benefit of common standards is that they will allow states to adopt better tests at a lower price through collaboration on common tests. This would be particularly consequential in small states, which pay more for their assessment systems because they lack the economies of scale of large states.

The leading options for Common Core assessments are the tests being developed by two consortia of states that are funded by federal grants: the Partnership for Assessment of Readiness for College and Careers (PARCC) and the SMARTER Balanced Assessment Consortium (SBAC). Thirty-three states currently belong to at least one of these consortia, both of which are developing math and ELA assessments for widespread use beginning in 2014-15. PARCC and SBAC have estimated that their computer-based end-of year assessments will cost $29.50 and $22.50 per student, respectively. These costs are not far from the nationwide average of what states currently pay for their existing tests, but many states have expressed concerns about these costs, especially states that currently spend well below average.

Assessment costs have always been a political football given the controversy surrounding standardized testing, and that has only intensified with the debate over Common Core. Concerns about the cost of the consortia's tests likely stem in part from a sense of uncertainty because the consortia have announced estimates, but not firm prices. States may be concerned that the price will go up, especially if states leave the consortia, and that they will be left without an affordable alternative. Opponents of the Common Core may be hoping that the withdrawal of a few states from the common assessments will lead to the unraveling of the consortia.

This report tackles this question through an empirical analysis of the consortia's cost estimates. By dividing per-student costs into the component that is the same regardless of the number of students (e.g., developing test questions) and the part that depends on the number of students (e.g., scoring essays), it is possible to estimate how prices will change if the consortia continue to lose members. This analysis shows that the departure of a few states will have a minimal impact on cost. For example, the possible departure of Florida, PARCC's second-largest member, only means a per-student price increase of about 60 cents for the remaining states.

This analysis also shows that rising costs from defections are unlikely to pose a serious threat to the existence of the consortia. If all of the states where political debate over the Common Core is most intense were to drop out of the consortia, costs would increase by no more than $2 or $3. In other words, for either PARCC or SBAC to face any real cost-based threat from states dropping out, the political opponents of the Common Core would have to be successful in generating withdrawals from all of the states where they have been most active and in several additional states.

States are right to pay careful attention to how they spend taxpayer dollars. States can use the consortium model to generate savings on assessments, but they must be careful not to be penny wise and pound foolish. Results of standardized tests are used for high-stakes decisions, ranging from which schools to close to which students to hold back a grade to which teachers to fire. All of the Common Core assessments under consideration cost less than a single textbook, and represent a drop in the $10,500 bucket of annual per-pupil spending on education. Consequently, quality should be at least as important as cost as states select their new assessment systems. This report boils down test quality into a set of design principles that are organized around a simple idea: tests should support and drive instruction in desirable ways.

State policymakers who support high-quality tests should form consortia and stick together so that high quality is affordable and sustainable, and gather good information on test quality to justify the spending required. Congress should support these efforts in the reauthorization of the federal No Child Left Behind law by restricting a small amount of federal education funding, such as $30 or $40 per student, to be spent only on assessments. This can be accomplished without an increase in total spending, and would force low-spending states to either upgrade their assessment systems or leave money on the table.

As implementation of the Common Core continues, states should consider the quality and cost of all available options, not just the two consortia's exams. ACT is planning to offer a suite of tests for $20 per student, and two states have worked with Pearson to develop state-specific Common Core tests for $30-$34 per student. States will not have all of the information they need to compare the growing array of Common Core assessments for some time, as more becomes known about the existing options and other vendors announce new alternatives. It is too early to tell which path will be the best choice for students, but two facts are clear: taxpayers get more bang for their buck when states collaborate, and students cannot afford for policymakers to compromise on assessment quality.

Standardized Testing and the Common Core Standards: You Get What You Pay For?

Introduction

A state with 100,000 students would save 37 percent on testing costs by joining a consortium containing one million students.

The Common Core State Standards represent the most significant change to the standards-based assessment movement since the No Child Left Behind (NCLB) Act required all states to adopt standards and administer standardized tests over a decade ago. But unlike the federal mandates of NCLB, Common Core is a voluntary, state-led effort—albeit one that has received support from the federal government. The new standards in mathematics and English language arts (ELA) have been adopted by 45 states and the District of Columbia, representing 85 percent of American students. Eight states claim to have already fully implemented Common Core, 20 are in the midst of implementation this year (2013-14), and most of the remaining states expect to implement the standards next year (2014-15).[1]

These new academic standards are touted by their creators as "designed to be robust and relevant to the real world, reflecting the knowledge and skills that our young people need for success in college and careers" and "the first step in providing our young people with a high-quality education."[2] It is important to emphasize that high-quality standards are only a "first step" in light of research questioning whether the quality of state standards has any effect on student achievement. For example, Tom Loveless has found that there is no association between measures of the content quality of state standards and student performance on the National Assessment of Educational Progress (NAEP).[3]

If new and improved standards are the "first step," what are the next steps? One potentially important one is the adoption of better tests at a lower price, something that the common standards achieve by creating opportunities for states to collaborate on common tests. This would be particularly consequential in small states, which currently pay more per student for their assessment systems, on average, than larger states. An analysis of current state spending on assessments suggests that a state with 100,000 students would save 37 percent on testing costs by joining a consortium containing one million students.[4] A state could realize those savings in the form of reduced expenditures on testing, or reinvest them in improvements to test quality.

It is still too early to tell whether these kinds of collaborations will be successful on a broad scale, but an ambitious experiment with student assessments has been under way since shortly after most states adopted the new standards. In September 2010, the U.S. Department of Education awarded a total of $330 million in grants to two groups of states: the Partnership for Assessment of Readiness for College and Careers (PARCC) and the SMARTER Balanced Assessment Consortium (SBAC). At the time the grants were announced,

In tough economic times, cost must be a factor, but so should quality—especially when almost all of the available testing options cost less than a single textbook.

PARCC and SBAC represented 26 and 31 states, respectively.[5] These grants cover the costs of developing new tests in math and ELA through September 2014, with member states responsible for ongoing costs beginning with the 2014-15 school year, when the new tests will be administered to all students for the first time.

The new tests being developed by PARCC and SBAC are often billed as part of "next-generation" assessment systems that will better assess student learning and use the consortium model to drive down costs through economies of scale. PARCC estimates that its tests will cost $29.50 per student, and SBAC anticipates a cost of $22.50 per student. These costs are not far from the nationwide average of state's contracted assessment costs (for math and ELA) of $27 per student.[6] But many states have expressed concerns about costs, especially states that currently spend well below average. States may also be concerned about the uncertainty around the consortia's price estimates, which could increase if states continue to withdraw. Four states have completely withdrawn from the consortia so far: Alabama, Georgia, Oklahoma, and Utah.

Costs can also be used as a political tool among those that oppose the Common Core standards or assessments for a variety of reasons, which often center on how the federal government's involvement in the Common Core standards and assessments might threaten state and local control of public schools. Georgia, an original PARCC member that has historically spent about $14 per student,[7] cited costs as a primary reason for dropping out of the consortium in 2013.[8] However, there is a much broader debate about Common Core underway in Georgia.[9] In neighboring Florida, the leaders of the state legislature have called for the state to leave PARCC because of cost concerns,[10] although existing data indicate that Florida's tests are more expensive than PARCC's.[11]

Political opponents of the Common Core may be pursuing a strategy of convincing a handful of states to withdraw from the consortia with the hope of creating a snowball effect in which the withdrawals lead to higher costs, the higher costs lead to more withdrawals, and the consortia eventually collapse. This study treats this question as an empirical one: how is cost expected to change if states leave each of the consortia? Does the departure of Florida signal the beginning of the end for PARCC, or just a small change in price? Would the departure of several states where political battles over the Common Core standards are raging have a large impact on the price of the consortia's tests?

In the next year, all states that have adopted the Common Core standards will have to decide whether to use tests from PARCC, SBAC, or somewhere else. In tough economic times, cost must be a factor, but so should quality—especially when almost all of the available testing options cost less than a single textbook. Assessment costs cannot be accurately evaluated without also carefully

considering the quality of those assessments. States need to know both what they're paying and what they're getting for it.

Costs of Common Core Assessments

Both of the consortia developing Common Core assessments have released cost estimates based on analyses conducted by Assessment Solutions Group (ASG), a group of consultants with extensive experience in the testing industry.[12] The publicly released estimates indicate that computer-based summative assessment in math and ELA will cost a total of $29.50 and $22.50 per student for PARCC and SBAC, respectively.[13] ASG claims that its cost model is accurate to within 5 percent if vendors' assumptions can be factored into the model (and within 10 percent if not), but it is not possible to verify this claim without being able to examine ASG's proprietary model.[14] However, through information obtained via discussions with consortia staff and freedom of information requests, it is possible to dig deeper into the assumptions underlying these cost estimates and make some rough predictions of how costs might change in response to changes in assumptions—such as the number of states in the consortium.

In this report, I focus on the summative assessments being developed by PARCC, SBAC, and other entities. However, it is important to note that many of these organizations are also developing formative assessments that can be used as part of an integrated system with the summative assessments. SBAC estimates that its full system of tests, which includes interim and formative tests in addition to the summative tests, will cost about $5 per student more than the summative assessment alone.[15] PARCC is developing diagnostic, mid-year, and K-2 assessments, but has not yet released cost estimates for those tests.[16]

PARCC

PARCC is developing summative math and ELA assessments that will consist of a performance-based test administered in the early spring and an end-of-year test administered in the late spring. Both tests are designed to measure critical-thinking and problem-solving skills, and the performance-based test will ask students to demonstrate their ability to formulate solutions to rich tasks such as real-world math problems and essays in which students analyze texts.[17] The tests will cover grades 3-11 with grade-based tests except for high school math, which will use course-based tests.[18] The ELA test will assess writing at every grade level with three extended writing exercises.

The cost estimate of $29.50 per student was calculated when PARCC had 22 member states containing almost 16 million students in tested grades (3-11). This cost estimate is for the computer-administered version of the test; PARCC estimates that the paper and pencil test will cost an additional $3-4 per student.

If Florida ultimately withdraws from PARCC, the per-student test cost would increase by 63 cents.

PARCC emphasizes that the computer-based version of their tests will allow for cost savings as well as greater security, quicker reporting of results, and the ability to use technology-enhanced test items.[19] However, the cost comparison does not factor in the costs of any technology upgrades that are needed in order for schools to be ready to deliver computer-based tests, a factor that will vary by school.

Hand-scoring of constructed response items accounts for nearly 75 percent of the total cost of delivering the PARCC assessment.[20] Because scoring is the biggest driver of PARCC's estimated costs, significant cost savings are possible if PARCC is able to use automated scoring technology to replace some of the scoring that is done by hand.[21] Below, I construct some rough estimates of how costs might change as the balance of human and computer scoring is altered using data from SBAC.

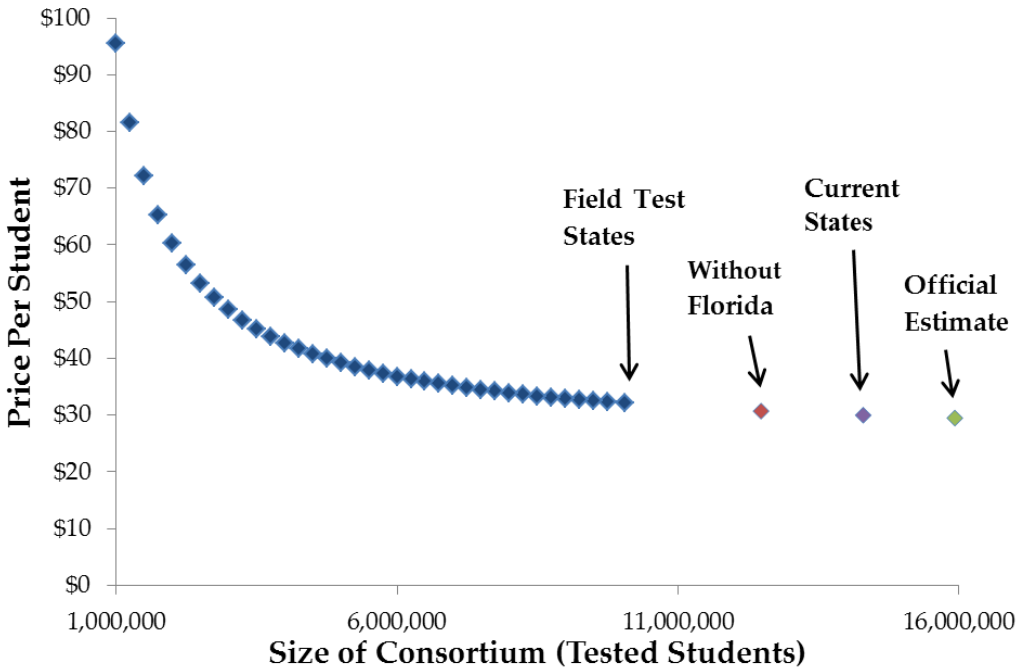
For a given number of member states, scoring costs are the biggest driver of per-student costs, as hand-scoring an individual student's essay costs roughly the same regardless of whether 10,000 or 10 million essays need to be scored. But the size of the consortium is also a potential cost driver, as the significant fixed costs of developing and maintaining the test are spread over member states. The fixed costs of content development, labor support, and travel accounted for roughly 15 percent of the total costs when the consortium had 22 member states. The remaining 85 percent include variable costs such as online assessment delivery, production and distribution, scoring, and reporting. In other words, of the $29.50 per-student cost, about $25 is variable costs that are the same (per student) regardless of how many students are tested and the other $4 is the per-student share of the total fixed costs of about $70 million (a cost that would vary on a per-student basis depending on the number of students tested).

Since the cost estimates were completed earlier this year, three states containing about 1.6 million students in tested grades have left PARCC. I estimate that this change will add about 50 cents to the per-student cost. More recently, Florida's governor announced his desire for Florida to end its role as PARCC's fiscal agent and to consider all available assessment options.[22] I estimate that losing Florida would add another 63 cents to the per-student test cost, bringing it to $30.72. A more pessimistic scenario is to estimate the cost if only the 15 states that are field-testing the PARCC tests end up adopting them in 2014-15, in which case the per-student cost increases to $32.08.

Figure 1 shows my rough estimate of the per-student PARCC costs for a range of consortium sizes. As expected, the test is more expensive if the fixed costs are spread over fewer states. But the price of the tests is not predicted to change drastically unless the consortium becomes quite small. PARCC can lose half the membership (based on student enrollment) of its field-testing group and still keep the cost under $40 per student, according to these calculations. In fact, if PARCC only included the five states that currently spend the most per-student

on testing (according to my analysis of contract data) those states would still all save money or break even. However, it is important to note that the largest of these five states is Florida—a state where political considerations may trump considerations of test cost and quality.

Figure 1. Estimated Cost of PARCC Assessment, by Size of Consortium



Could additional politically motivated withdrawals from consortia cause the whole enterprise to unravel due to rising costs? A May 2013 Associated Press report identified nine states where there is significant resistance to the new standards, and a September 2013 post from the Heritage Foundation identified three more that are considering withdrawing from the consortia.[23] Of these 12 states, four are PARCC members—Indiana, Florida, Louisiana, and Pennsylvania—but only one (Louisiana) is planning to field test the PARCC exams. My estimates indicate that losing these states would push the PARCC price up to $31, hardly a catastrophe for the remaining states.

Smarter Balanced

SBAC's basic assessment system will consist of summative tests in math and ELA in grades 3-8 and 11 administered during the last 12 weeks of the school year.[24] SBAC and PARCC share the goal of measuring students' college and career readiness through the use of test items that gauge students' critical-thinking and problem-solving skills, but there are some key differences between the consortia and the tests they are developing. Both tests are computer-based,

SBAC's decentralized model, in the consortium's view, allows states to select services from the open market to meet their specific needs.

but SBAC tests are also computer-adaptive, meaning that students receive harder or easier questions based on their answers to previous questions in order to more precisely estimate their true performance. SBAC will also offer a non-adaptive, paper-and-pencil version of its test for three years in order to include schools that are not ready to administer computer-based tests.[25]

PARCC and SBAC also differ significantly in terms of which assessment services are provided by the consortium and which are provided by the states. PARCC's model is centralized, with PARCC (through its vendors) providing all assessment-related services, including test administration and scoring. SBAC is much more decentralized, with the consortium responsible for item development but states in charge of test delivery, scoring, and reporting. Because scoring is the largest cost driver, the majority of the costs of the SBAC tests will be paid by states directly to vendors rather than through SBAC. The $22.50 per-student total cost estimate SBAC has released reflects $6.20 in consortium services and $16.30 in state-managed services.[26]

ASG's cost analysis for SBAC, which I was able to obtain through a freedom of information request, estimates consortia costs of $2.43 per student.[27] According to SBAC, this figure only includes the consortia costs related to the operational aspects of the test that were estimated by ASG. The difference between $2.43 and the official estimate of $6.20 reflects SBAC's estimates of other costs not modeled by ASG, including indirect costs charged by UCLA (SBAC's future home), consortium staff, and information technology maintenance.[28] The ASG cost estimates assume that states will not go it alone on state-managed services, but rather will join mini-consortia of five to six states containing a total of roughly two million students. According to ASG's estimates, single-state costs are projected to be about $10 higher per student, although larger states will see smaller increases due to their existing economies of scale. Or to put it in ASG's words: "Everyone should become California's best friend."[29]

According to SBAC, the decision to have states manage key assessment services reflects the consortium's identity as a state-led organization. States currently manage their assessment systems in different ways. Two SBAC states use state universities for key assessment responsibilities. Some states have long-standing relationships with a contractor, and in some cases want to associate their math and ELA assessments with other assessments that are not provided by either consortia (e.g., in science or social studies). SBAC's decentralized model, in the consortium's view, allows states to select services from the open market that meet their specific needs. For example, some states may want more advanced score reporting systems than others. States will thus have considerable flexibility in how they administer the test, but SBAC will certify whether each state meets the consortium's implementation standards.[30] The potential downside of the decentralized model is that it will lead to higher costs for states that do not collaborate with other states as part of the mini-consortia model

envisioned by SBAC. At this point, it is too early in the procurement process to tell whether mini-consortia will successfully form.

ASG's detailed cost estimates for SBAC, summarized in Table 1, shed a great deal of light on the main drivers of the assessment costs, both for the consortium and for member states. At the consortium level, the main cost drivers are content development (primarily item development) and labor supply (program management, psychometrics, IT staff, etc.), which together account for about 80 percent of consortia costs. But if the costs not modeled by ASG were factored in, those costs would account for just over 60 percent of the $6.20 per student charge to states for consortium services and reduce the content development and labor supply share to 31 percent of overall cost.

Table 1. Breakdown of Estimated Costs, SBAC Summative Assessment

Type of Cost	Total Amount	Per Student	Percentage
Consortium Costs			
Content Development	$9,575,357	$1.04	43%
Labor Support	$8,069,306	$0.88	36%
Online Delivery	$403,900	$0.04	2%
Scoring	$2,526,081	$0.27	11%
Travel	$1,821,162	$0.20	8%
Total	$22,395,806	$2.43	
State Costs			
Labor Support	$23,035,921	$2.50	15%
Online Delivery	$23,509,091	$2.55	16%
Production and Distribution	$2,882,972	$0.31	2%
Receiving/Scanning/Machine Scoring	$585,273	$0.06	0%
Reporting	$1,065,642	$0.12	1%
Scoring	$97,911,953	$10.63	65%
Travel	$1,151,851	$0.13	1%
Total	$150,142,703	$16.29	
Total of Consortium and State Costs	$172,538,509	$18.73	

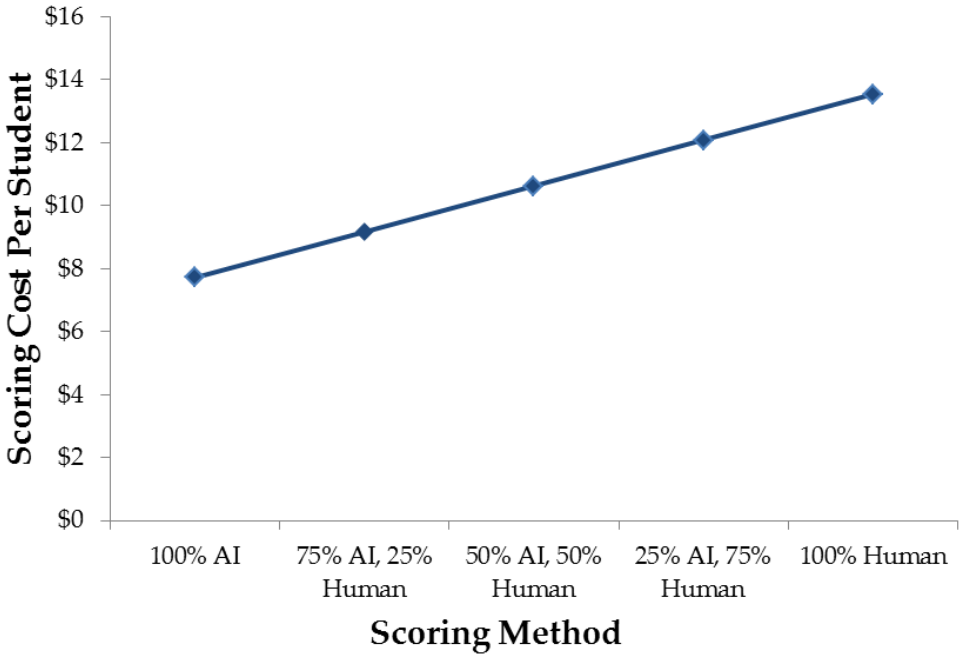
Source: Assessment Solutions Group cost estimates for SBAC, March 2013.

Scoring is the main driver of costs at the state level, at 65 percent of costs. At the time of ASG's cost estimates, SBAC was planning for states to score roughly half of the constructed-response and performance-task items using artificial intelligence (AI) software. AI is a cost-saver, as it allows computers to score

items that would otherwise have to be done by hand. According to ASG's estimates, 36 percent of scoring costs are for AI scoring even though roughly 50 percent of relevant items are scored in this way. I use these estimates to calculate how per-student costs would change if the mix of AI and human scoring were changed.[31] Figure 2 shows that scoring costs are predicted to increase from $10.63 with a 50/50 mix to $12.08 if AI were only used for 25 percent of items and to $13.53 if AI were not used at all.

Since ASG's cost estimates were completed, SBAC has decided to remove AI as a key component of its scoring approach for at least the first operational year of its assessment (2015). In order to allow states to stay within the original cost targets, the consortium has reduced the number of hand-scored items by about eight percent.[32]

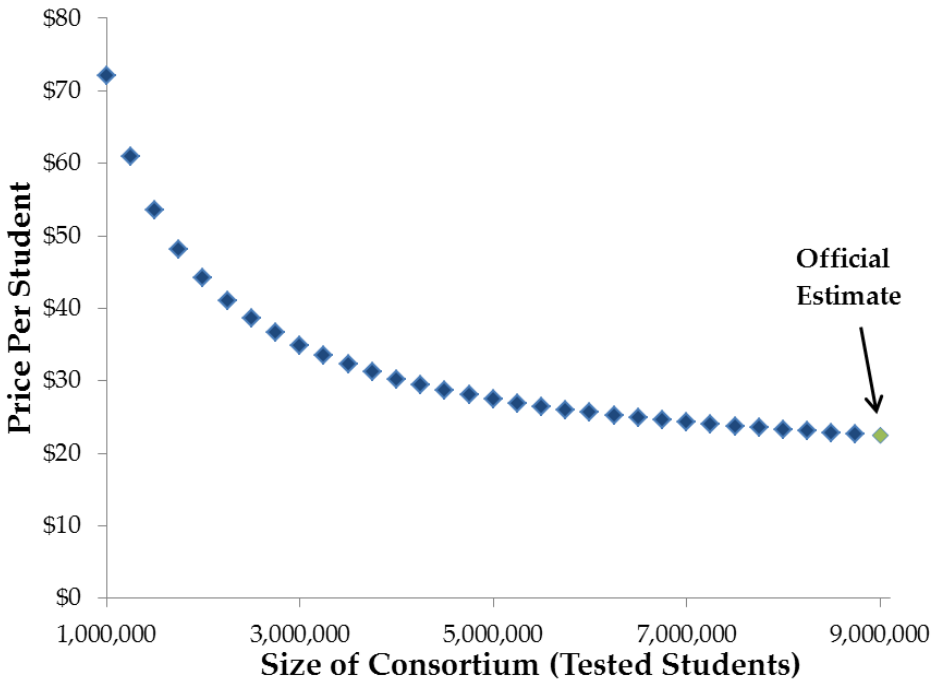
Figure 2. SBAC Per-Student Scoring Costs, by Scoring Method



I also calculate how SBAC costs might change if the consortium loses members. For the purposes of this rough simulation, I assume that the $6.20 per student charge for the consortium is a fixed cost and the $16.30 cost states are predicted to pay is a variable cost.[33] The official estimate of $22.50 is based on the 22 governing states in the consortium, which include approximately 9 million students in tested grades. As illustrated in Figure 3, SBAC can lose half of its members without the cost rising above $30, and more than two-thirds of its members before the cost rises to about $40.

What would the effect on the SBAC price be if politics leads to more withdrawals from the consortium? The group of states identified above as having significant political debate over Common Core includes six members of SBAC: Kansas, Missouri, Michigan, Pennsylvania, South Carolina, and Wisconsin. My simulations suggest that the withdrawal of these six states would increase the SBAC price to about $25 per student—only a few dollars more than the current estimates. In sum, for either PARCC or SBAC to face any real cost-based threat from states dropping out, the political opponents of the Common Core would have to be successful in all of the states where they have been most active and in several additional states.

Figure 3. Estimated Cost of SBAC Assessment, by Size of Consortium



Can the SBAC and PARCC cost estimates be compared to determine whether one test is more expensive than the other? The published cost estimates suggest that SBAC's tests are $7, or about 24 percent, less expensive per student than PARCC's. However, uncertainty about the assumptions underlying these estimates renders such a comparison nearly meaningless. A close examination of these assumptions suggests that there is more uncertainty about SBAC's estimates than about PARCC's. The primary reason for this is the decentralized structure of SBAC, with states facing a range of costs depending on the nature of the assessment administration and scoring services they procure and whether they are able to collaborate with other states.[34] An SBAC state that goes it alone

would face a test price that is $3, or about 10 percent, higher per student than the PARCC cost.

In sum, the available evidence suggests that PARCC's estimates are more conservative than SBAC's, so the final difference in price may be smaller than the current estimates indicate.

The available evidence suggests that PARCC's estimates are more conservative than SBAC's, so the final difference in price may be smaller than the current estimates indicate.

Other Options

PARCC and SBAC exams have received the most attention and been viewed as the quasi-official Common Core assessments, given that their start-up costs were completely funded by the federal government. But as implementation of the Common Core standards gets into full swing and political battles heat up over the new standards and tests, states are considering other options as partial or full alternatives to the PARCC and SBAC exams. These include new tests developed for individual states by the same vendors that have developed tests for years, as well as a new testing system being developed by a national organization.

At this point, the only direct competitor with PARCC and SBAC is ACT, most widely known for the college admissions test of the same name, which is launching its Aspire testing system in the spring of 2014. Aspire will include a suite of yearly tests in English, math, reading, science, and writing in grades 3-10 that is being billed as linked to the ACT exam (and ACT's college-readiness standards) and aligned to the Common Core.[35] One state, Alabama, has committed to using the test at launch and has received an early adopter price of $11.70 per student. ACT expects to eventually charge $20 for the computer-based version of the test and $26 for the paper-and-pencil version. Versions of the test with only a subset of the five subjects will also be made available, but pricing has not yet been announced.[36]

ACT is the only test-maker that has announced a Common Core aligned test that could be used by multiple states, but states also have the option of contracting with vendors for the development of state-specific tests. Kentucky and New York have gone this route in order to transition to Common Core tests before the consortia's tests are available in 2015. Both states are members of PARCC (although only New York is participating in the field test), but it is possible that they will use their current Common Core tests beyond the current transition period. Kentucky has a contract with Pearson at a cost of about $30 per student in 2012-13 and 2013-14, but that contract was amended to include science and social studies tests as well as additional item development in math and reading, raising the cost to $37 per student in 2013-14.[37]

New York has a five-year contract with Pearson (through 2015), at a cost of about $5 per student. However, the contract does not include printing, shipping, or scoring costs. The state assessment office estimates a cost of about $13 per

student once printing and shipping costs are included.[38] However, scoring is a local responsibility in New York paid for by individual school districts, so the $13 figure cannot be compared to the estimated costs of the SBAC, PARCC, ACT, or Kentucky tests, which include scoring costs. New York City charter schools can participate in a scoring consortium for $10.32 per student per test, implying a total per-student testing cost in math and ELA of about $34 (including both state and local costs).[39]

It seems short sighted to accept a significant decrease in test quality in order to save $10 or $20 per student in the context of an education system that spends more than $10,000 per student.

As implementation of the Common Core standards proceeds, there are likely to be additional options that emerge. The College Board might decide to expand its high school testing program into the elementary and middle grades, perhaps in collaboration with ETS and Pearson, as it suggested in a 2010 report.[40] Commercial vendors might market off-the-shelf tests that any Common Core state can use. There is even the possibility that another state consortium might form in order to pool their resources to develop a Common Core assessment. But it is not yet clear whether any of these options will be available by the spring of 2015, when the PARCC and SBAC are expected to go live.

Gauging the Quality of Assessments

If states wanted to spend as little as possible on their assessment systems, they would buy inexpensive off-the-shelf assessments from vendors that can be scored automatically, with the only constraint on quality being any standards set by the U.S. Department of Education under No Child Left Behind. But states at least claim to want high-quality assessments that can be used as a factor in high-stakes decisions while hopefully supporting student learning. And they ought to care about the quality of the measures used for such high-stakes decisions such as whether to close a school or whether to fire a teacher. Consequently, it is critical that any discussion of test cost be balanced by a discussion of quality. There may well be ways to save money without sacrificing quality, but there is no doubt that states need to consider both cost and quality when choosing an assessment system. This is especially true given the small financial stakes involved, with all of the tests discussed so far coming in well below the cost of a textbook. It seems short sighted to accept a significant decrease in test quality in order to save $10 or $20 per student in the context of an education system that spends more than $10,000 per student.

A detailed exploration of this set of issues is beyond the scope of this report, but there are at least four key design principles that policymakers should take into account when selecting or commissioning the development of a high-quality assessment system.[41] These principles are organized around a simple idea: tests should support and drive instruction in desirable ways. In other words, if policymakers are going to use tests to hold students, teachers, and schools accountable, the tests should provide honest and credible measurement of the standards and they should be worth teaching to. Tests send clear signals about

The design of tests should take into account the likelihood of "teaching to the test" rather than viewing it as an unintended negative consequence.

what—and, to some degree, how—teachers should teach, so the design of tests should take into account the likelihood of "teaching to the test" rather than viewing it as an unintended negative consequence.

First, tests should include the kinds of tasks that we want students to learn in school. An exclusively multiple-choice test may be psychometrically valid, at least in some subjects, but can create unintended consequences when it has high stakes attached to it.[42] The most-cited example is extensive "test prep" in which students practice multiple-choice questions over and over rather than engaging with the content underlying the test in a variety of ways. A high-quality test mitigates these kinds of consequences by including a mix of items that represent the kinds of tasks that students should be learning in school, so instruction oriented at these tasks is desirable rather than undesirable. Currently, many states do not include such a mix of items. One-third of states give completely multiple-choice exams in math and reading to their 4th- and 8th-grade students.[43]

Writing is the clearest example of a task that should be included in assessments because it is important for students to learn. Many states assess ELA through separate reading and writing tests, with writing tests only administered in certain grades. Consequently, tests do not create incentives for schools and teachers to focus on the development of students' writing ability across all grade levels. This is likely a cost-saving measure, as writing tests entail additional costs to develop, administer, and score. In recent years, North Carolina spent about $14 per student on its 10th-grade writing test, Utah spent $8 per student testing 5th- and 8th-grade students, and Alabama spent about $4 on its writing test in grades 5, 7, and 10.[44]

Second, tests should cover the full range of content included in the standards, probe the depths of student thinking and levels of knowledge expected by the relevant standards, and accurately measure the performance of all students. Existing state assessments often do not adequately cover the content in states' standards. A 2011 study found that only half of state standards were covered on the corresponding tests, and that 17-27 percent of test content was not mentioned in the standards.[45] Presumably many of these tests were customized for the state's standards, yet fell significantly short on this design principle. In the coming years, many tests will be marketed as "Common Core tests," but states need to be careful to select one that assesses student performance on the full range of standards (and excludes extraneous material) if the standards and assessments are to send consistent signals to educators.

Good tests include items that probe student thinking at different depths and levels, both within and across standards. Some standards are more challenging than others, and for a given standard, items of differing difficulty gauge how well a student has mastered that standard. One way that test designers can ensure the inclusion of a variety of items along this dimension is to write items

that assess different depths of student thinking. One formal framework for this process is Norman Webb's Depth of Knowledge (DOK) four-category classification: 1) recall, 2) simple application, 3) application requiring abstract thinking, and 4) extended analysis or investigation.[46] For example, a level-2 item might ask a student to solve a simple math problem, whereas a level-3 item might ask the student to figure out what math equation needs to be solved. A recent RAND study applied the DOK classification to math and ELA tests in 17 states and found a low overall level of rigor.[47]

Tests need to include not just a general variety of items along these dimensions, but a sufficient number to accurately measure the full range of student performance. Tests are usually more accurate at measuring the performance of students who perform about average, and are less accurate for the highest and lowest performers. In the extreme, tests can have ceiling or floor effects in which a non-trivial number of students answer all of the questions correctly or incorrectly. In order to accurately measure the performance of all students, tests either need to be long enough to include a sufficient number of items at all levels of difficulty, or have adaptive features in which students receive items based on their performance on previous test items.

Third, as assessments that purport to measure college and career readiness are put in place, policymakers should demand evidence that they are indeed predictive of success in college and careers. This principle is most salient at the high school level, where existing college-admissions tests are at best weak predictors of success in college for most students.[48] The next generation of assessments should reliably identify the chances that a student will persist in and graduate from college, or whether non-college-bound students will find success in the labor market. At the same time, policymakers should recognize that the academic knowledge and skills measured by standards and assessments represent only a portion of college and career readiness.

Finally, states need to be concerned not just with the quality of the tests themselves but also with the reporting system used to provide feedback to students, teachers, and schools. Feedback should be as timely and informative as possible so that, for example, teachers can identify the strengths and weaknesses of their students and incorporate that information into instruction. This is possible in a system of summative assessments, where teachers use information from the prior year's assessment to plan instruction for their students. But it has much greater potential in a system of both formative and summative assessments, in which students take low-stakes tests throughout the year that provide timely and detailed information on their progression toward the standards that will ultimately be assessed on the end-of-year test.

For too long, states have used tests that were added on top of the existing education system rather than designed to be part of it. These tests met the basic technical standards of the psychometric profession in terms of reliability,

validity, and bias, but in many cases were being used for high-stakes purposes for which they were not designed. Next-generation tests need to be systems of assessments that are designed to be an integrated part of the instructional process so that "teaching to the test" can be a fact of life rather than a derisive criticism of standards-based testing.

For too long, states have used tests that were added on top of the existing education system rather than designed to be part of it.

How well do the Common Core tests now in development by PARCC and SBAC fare when evaluated against these design principles? It is too early to tell at this stage in the process, with field testing still months away and only a limited number of sample items publicly available. The quality of the score reporting systems will not be known until closer to the operational administration of the tests—and, in the case of SBAC, will vary by state. And it will be several years before students who have taken the new tests are in college or the labor market, enabling researchers to measure the ability of the new tests to predict success in college and careers.

But there are certainly promising signs that PARCC and SBAC are designing tests with many of these principles in mind. The consortia's assessments will include a mix of multiple-choice and constructed-response items, as well as technology-enhanced items on the computer-based version of the assessments. The new tests will assess writing at every grade with multiple constructed-response items (a mix of essays and shorter written responses), eliminating the need for participating states to administer separate writing tests. And both PARCC and SBAC are designing tests that they expect to accurately measure students at all levels, which SBAC plans to accomplish through the adaptive nature of its test and PARCC will accomplish by administering a longer test.

Conclusions and Policy Recommendations

It is easy to understand why assessment costs receive a great deal of attention given the controversy surrounding standardized testing and the budgetary pressures that states have been facing for several years. States and school districts should do everything in their power to operate as efficiently as possible. Collaborating on assessments in order to spread the fixed costs of test development over a larger number of students is an obvious strategy to reduce costs, especially for smaller states. For example, in previous work, I estimated that a state with 100,000 students would save 37 percent on testing costs by joining a consortium containing 1 million students.[49]

At the same time, states must not be penny wise and pound foolish. Results of standardized tests are being used for high-stakes decisions, ranging from which schools to close to which students to hold back a grade to which teachers to fire. All of the Common Core assessments discussed above cost less than the price of a textbook, as do most of the assessments currently in use through the country. The most recent data indicate that U.S. schools spend about $10,500 per

student each year. In that context, it seems almost silly to worry about whether it costs $15 or $30 or $45 to properly measure how much students are learning.

The main reason these costs must be taken seriously is a political one. Much of the opposition to the Common Core standards has little to do with quality or cost, although the latter is sometimes raised as a concern. My analysis clearly shows that, even if efforts to withdraw from Common Core are successful in several states, the price of the PARCC and SBAC assessments will not change by more than a few dollars. Consequently, even successful efforts to convince states to withdraw from the consortia are unlikely to have repercussions in other states based on the price of the tests.

Making the cost of state assessments into a political issue certainly pre-dates the Common Core effort, and will continue to be a point of contention even in states that are not questioning their adoption of the new standards, especially those that currently spend very little on their tests. A difference of $10 per student is a drop in the bucket of per-pupil spending, but in a state with 1 million students it adds up to $10 million dollars. Multi-year contracts with for-profit testing vendors in large states, such as Texas's five-year, $468 million contract with Pearson, are easy targets for opponents of standardized testing. The political dynamics do not favor high-quality assessments. Those most concerned about the low quality of many existing tests, who should in theory be supportive of spending a little more to increase test quality, often oppose testing in any form. On the other side, many supporters of testing and accountability either do not recognize differences in item and test quality or do not want to concede the weaknesses of these tests.

In order to overcome these political barriers, state policymakers who support high-quality tests should adopt two approaches. First, they should support state-led efforts to form consortia and stick together so that high quality is affordable and sustainable. For states which pay very little for low-quality tests, this will still mean increasing their testing budget. Second, in order to justify their spending on assessments—especially if an increase is required—states need to carefully gather good information on test quality. High-quality assessment systems will support and drive instruction in desirable ways by sending clear signals about what students should learn in school and provide timely feedback on their progress.

One potential advantage of the changes to the market for assessments created by common standards is that every state does not have to reinvent the wheel in evaluating the relative quality of all of the available options. Surely states will want to examine how various test aspects meet their specific needs, but there is clear room for an independent organization to carry out this analysis and provide the results as a public service to states as they select a Common Core assessment. For example, a nationally recognized group of testing experts could carefully evaluate how well each test includes the kinds of tasks that students

should learn in school and the extent to which the items challenge students to use different kinds of critical-thinking skills while covering the full range of content in the standards.

Congress should amend NCLB so that part of federal education funding must be spent on assessments.

In the current political environment, in which the Obama administration's support for the Common Core standards has caused a backlash against the standards among some conservative groups, supporters of the new standards and assessments will find it in their interest to ensure that this type of effort not be undertaken by the federal government. However, the U.S. Department of Education has an existing review process for standards and assessments that is currently undergoing revision. The revised process should make it more difficult for states to use low-quality tests as long as the process is transparent and clearly aimed at test quality in general, and does not appear to be designed to favor particular tests.[50]

Federal testing mandates may also undergo revision whenever Congress finally reauthorizes the No Child Left Behind Act that required states to administer yearly tests in most grades beginning more than a decade ago. One proposal for absorbing the increased costs of higher quality tests is to relax the federal mandate so that states must only test students once in each of the elementary, middle, and high school grades.[51] This would represent a huge setback to accountability efforts, as it would render it impossible to track student growth from one year to the next, to provide timely feedback to students and teachers, and to fairly compare the performance of students in different classrooms and schools.

Instead, Congress should amend NCLB so that part of federal education funding must be spent on assessments. Without any increase in total spending, Congress could set a minimum amount per student that must be spent on assessments, with the minimum chosen to reflect policymakers' view as to the cost of high-quality assessments. For example, if Congress set the amount at $30 per student, states would lose $10 per student in federal funding if they chose to spend only $20 on their assessment. Pushing states to spend modestly more on testing without an increase in overall spending would have only a trivial impact on other categories of spending. For example, for a state that currently spends $15 per student on testing, the additional $15 the state would be encouraged to spend represents less than two percent of the education funding received from the federal government (and about 0.1 percent of total education spending).

States will not have all of the information they need to compare the growing array of Common Core assessments for some time. PARCC and SBAC will presumably release more items and make final decisions about costs once field testing is complete in 2014. Over the next year, other vendors may well announce new alternatives, ACT will continue to market its test, and states' existing vendors may try to convince them to use a customized test rather than one of the multi-state options. It is too early to tell which path will be the best

choice for students, but two facts are clear: taxpayers get more bang for their buck when states collaborate, and students cannot afford for policymakers to compromise on assessment quality.

Endnotes

[1] Common Core State Standards Initiative, "In the States," http://www.corestandards.org/in-the-states, accessed September 26, 2013. Two states are planning full implementation for 2015-16 and two additional states did not provide an implementation date.
[2] Common Core State Standards Initiative, http://www.corestandards.org/.
[3] Tom Loveless, "The 2012 Brown Center Report on American Education: How Well Are American Students Learning?" Brown Center on Education Policy, Brookings Institution, February 2012. For research coming to a different conclusion on the relationship between standards and student achievement, see, e.g., William H. Schmidt and Richard T. Houang, "Curricular Coherence and the Common Core State Standards for Mathematics," *Educational Researcher*, 41(8): 294-308.
[4] Matthew M. Chingos, "Strength in Numbers: State Spending on K-12 Assessment Systems," Brown Center on Education Policy, Brookings Institution, 2012.
[5] U.S. Department of Education, "U.S. Secretary of Education Duncan Announces Winners of Competition to Improve Student Assessments," September 2, 2010, available at http://www.ed.gov/news/press-releases/us-secretary-education-duncan-announces-winners-competition-improve-student-asse.
[6] Matthew M. Chingos, "Strength in Numbers: State Spending on K-12 Assessment Systems," Brown Center on Education Policy, Brookings Institution, 2012.
[7] Ibid.
[8] Maureen Downey, "Georgia will develop its own tests for Common Core. So, how will we know where we stand nationally?" *Atlanta Journal-Constitution* Get Schooled blog, July 22, 2013, http://www.ajc.com/weblogs/get-schooled/2013/jul/22/breaking-news-georgia-will-develop-its-own-tests-c/.
[9] Greg Bluestein and Wayne Washington, "Deal Orders Review of Common Core," *Atlanta Journal-Constitution*, August 21, 2013, http://www.ajc.com/news/news/state-regional-govt-politics/deal-orders-review-of-common-core/nZYbc/.
[10] Andrew Ujifusa, "Florida Gov. Rick Scott: State Will Curtail Role in Common-Core Testing Consortium," September 23, 2013, http://blogs.edweek.org/edweek/state_edwatch/2013/09/florida_leaves_common-core_testing_group_as_gov_scott_pulls_plug.html.
[11] Matthew M. Chingos, "Strength in Numbers: State Spending on K-12 Assessment Systems," Brown Center on Education Policy, Brookings Institution, 2012.
[12] Assessment Solution Group, "who we are," http://www.assessmentgroup.org/aboutus.html.
[13] PARCC, "Assessment Cost Estimates," http://parcconline.org/assessment-cost-estimates; Smarter Balanced, "Frequently Asked Questions," http://www.smarterbalanced.org/resources-events/faqs/.
[14] Assessment Solutions Group, "ASG Response to SBAC RFP #24," January 2013.
[15] Smarter Balanced, "Frequently Asked Questions," http://www.smarterbalanced.org/resources-events/faqs/.
[16] PARCC, "Assessment Cost Estimates," http://parcconline.org/assessment-cost-estimates.
[17] PARCC, "PARCC Tests: An Investment in Learning," July 2013, http://ca539dfd55636c55e922-fd4c048d1c793e15a27f954b34a49d25.r49.cf1.rackcdn.com/PARCCCostEstimates_07-22-13.pptx.
[18] PARCC, "Assessment Blueprints and Test Specifications," http://parcconline.org/assessment-blueprints-test-specs.

[19] PARCC, "PARCC Tests: An Investment in Learning," July 2013, http://ca539dfd55636c55e922-fd4c048d1c793e15a27f954b34a49d25.r49.cf1.rackcdn.com/PARCCCostEstimates_07-22-13.pptx.
[20] Email correspondence with Jeff Nellhaus, PARCC, August 2013.
[21] PARCC, "PARCC Tests: An Investment in Learning," July 2013, http://ca539dfd55636c55e922-fd4c048d1c793e15a27f954b34a49d25.r49.cf1.rackcdn.com/PARCCCostEstimates_07-22-13.pptx.
[22] Andrew Ujifusa, "Florida Gov. Rick Scott: State Will Curtail Role in Common-Core Testing Consortium," September 23, 2013, http://blogs.edweek.org/edweek/state_edwatch/2013/09/florida_leaves_common-core_testing_group_as_gov_scott_pulls_plug.html.
[23] The Associated Press, "Some States Push Back on Common Core Standards," May 28, 2013, available at http://www.huffingtonpost.com/2013/05/28/states-common-core-push-back-standards_n_3346210.html; and Lindsey Burke, "Florida to Withdraw from Common Core Assessments; Other States May Soon Follow," The Foundry, Heritage Foundation, September 30, 2013, available at http://blog.heritage.org/2013/09/30/florida-to-withdraw-from-common-core-assessments-other-states-may-soon-follow/.
[24] SBAC is also developing interim tests as well as secure summative tests for grades 9, 10, and 12. The costs of these tests are not considered in this report.
[25] Smarter Balanced, "Frequently Asked Questions," http://www.smarterbalanced.org/resources-events/faqs/.
[26] Joe Willhoft, "The Future is (Almost) Now: Implementing Smarter Balanced Assessments in 2014-15," June 2013, https://ccsso.confex.com/ccsso/2013/webprogram/Presentation/Session3700/3%20Willhoft_2013-06-21_NCSA_SmarterBalanced%20Sustainability%5B2%5D.pdf.
[27] Barry Topol, "RRP requirement 2f, 2g," March 17, 2013.
[28] Phone interview with Tony Alpert, SBAC, October 2, 2013 and email correspondence with Tony Alpert, SBAC, October 24, 2013.
[29] Assessment Solutions Group, "SBAC Assessment Cost Update/Sustainability Planning, RFP Requirement 2e, 2f, 2g," March 13, 2013.
[30] Phone interview with Joe Willhoft, SBAC, September 25, 2013.
[31] My linear model of scoring costs is likely to be least accurate for estimating the costs of fully AI or fully hand-scoring given the fixed costs of using each of these scoring methods.
[32] Email correspondence with Tony Alpert, SBAC, October 24, 2013.
[33] In other words, I assume that the mini-consortia model will work and that states pay the same per-student cost regardless of the size of the entire Smarter Balanced consortium.
[34] Another difference in assumptions is that PARCC assumes a 30 percent read-behind rate (the percent of open-ended tasks that are scored by a second person, usually a second scorer for measuring inter-rater reliability or an expert rater for measuring rater accuracy), whereas SBAC assumes a 10 percent rate. This difference could also explain a small part of the difference in cost estimates.
[35] ACT Aspire, http://www.discoveractaspire.org/pages/home.
[36] Phone interview with Paul Weeks, ACT, 9/23/13.
[37] Author's calculations from Kentucky state contracts.

[38] Phone interview with Candy Shyer, New York State Department of Education, September 25, 2013.
[39] New York City Charter School Center, Test Scoring Consortium, http://www.nyccharterschools.org/scoring.
[40] Stephen Lazer, John Mazzeo, Jon S. Twing, Walter D. Way, Wayne Camara, and Kevin Sweeney, "Thoughts on an Assessment of Common Core Standards," ETS, Pearson, and College Board, 2010.
[41] Two other recent discussions of design principles for high-quality assessments include Council of Chief State School Officers, "States' Commitment to High-Quality Assessments Aligned to College- and Career-Readiness," 2013, available at http://blogs.edweek.org/edweek/curriculum/CCSSO%20Assessment%20Quality%20Principles%2010-1-13%20FINAL.pdf; and Linda Darling-Hammond and others, "Criteria for High-Quality Assessment," Stanford Center for Opportunity Policy in Education, 2013, available at https://edpolicy.stanford.edu/sites/default/files/publications/criteria-higher-quality-assessment_2.pdf.
[42] For evidence on the effect of standards-based testing on instruction, see, e.g., William A. Firestone, David Mayrowetz, and Janet Fairman, "Performance-Based Assessment and Instructional Change: The Effects of Testing in Maine and Maryland," *Educational Evaluation and Policy Analysis,* 20(2): 95-113
[43] Table A5 of Matthew M. Chingos, "Strength in Numbers: State Spending on K-12 Assessment Systems," Brown Center on Education Policy, Brookings Institution, 2012.
[44] Author's calculations from state contracts with assessment vendors in North Carolina (2011), Utah (2011), and Alabama (2008-2010).
[45] Morgan S. Polikoff, Andrew C. Porter, and John Smithson, "How Well Aligned Are State Assessments of Student Achievement With State Content Standards?" *American Educational Research Journal,* 48(4): 965-995.
[46] Joan Herman and Robert Linn, "On the Road to Assessing Deeper Learning: The Status of Smarter Balanced and PARCC Assessment Consortia," CRESST Report 823, National Center for Research on Evaluation, Standards, & Student Testing, UCLA, 2013.
[47] Kun Yuan & Vi-Nhuan Le, "Estimating the Percent of Students Who Were Tested on Cognitively Demanding Items Through the State Achievement Tests," RAND Education, 2012.
[48] William G. Bowen, Matthew M. Chingos, and Michael S. McPherson, *Crossing the Finish Line: Completing College at America's Public Universities,* Princeton, NJ: Princeton University Press, 2009.
[49] Matthew M. Chingos, "Strength in Numbers: State Spending on K-12 Assessment Systems," Brown Center on Education Policy, Brookings Institution, 2012.
[50] The current Standards and Assessments Peer Review Guidance is available at http://www2.ed.gov/policy/elsec/guid/saaprguidance.pdf.
[51] Marc Tucker, Linda Darling-Hammond, and John Jackson, "Fewer, Better Tests Can Boost Student Achievement," *Education Week,* October 7, 2013.

About the Brown Center on Education Policy

Established in 1992, the Brown Center on Education Policy conducts research and provides policy recommendations on topics in American education. The Brown Center is part of The Brookings Institution, a private nonprofit organization devoted to independent research and innovative policy solutions. For more than 90 years, Brookings has analyzed current and emerging issues and produced new ideas that matter - for the nation and the world.

Brown Center on Education Policy

The Brookings Institution
1775 Massachusetts Ave. NW
Washington DC, 20036
202.797.6090
202.797.6144 (f)
http://www.brookings.edu/brown.aspx

Support for this publication was generously provided by Lumina Foundation.

Brookings recognizes that the value it provides is in its absolute commitment to quality, independence and impact. Activities supported by its donors reflect this commitment and the analysis and recommendations are not determined or influenced by any donation.

Governance Studies
The Brookings Institution
1775 Massachusetts Ave., NW
Washington, DC 20036
Tel: 202.797.6090
Fax: 202.797.6144
www.brookings.edu/governance.aspx

Editing, Production, & Layout
Liz Sablich

Email your comments to browncenter@brookings.edu

This paper is distributed in the expectation that it may elicit useful comments and is subject to subsequent revision. The views expressed in this piece are those of the authors and should not be attributed to the staff, officers or trustees of the Brookings Institution.