

November 2012

STRENGTH IN NUMBERS

State Spending on K-12 Assessment Systems



Reuters

Matthew M. Chingos

The significant data collection effort undertaken for this project was skillfully coordinated by Mike Gallaher, with outstanding assistance by Christine Lai, Emily Russel, and Diana Stockwell. This report also benefited from helpful conversations with Linda Darling-Hammond, Scott Marion, Kim Moe, Jeff Nellhaus, Sheri Ranis, Russ Whitehurst, and Joe Willhoft.

EXECUTIVE SUMMARY



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

Standardized tests have become ubiquitous in U.S. elementary and secondary education over the last few decades. About ten years ago, states across the country expanded their assessment systems in response to the federal No Child Left Behind Act's requirement that students take annual math and reading tests in grades 3-8 and once in high school. In the coming years, states will need to make the most significant changes to their assessment systems since the passage of NCLB as they implement the Common Core State Standards, a common framework for what students are expected to know that will replace existing standards in 45 states and the District of Columbia.

The Common Core effort has prompted concerns about the cost of implementing the new standards and assessments, especially in states that have historically spent very little on their tests. Unfortunately, there is little comprehensive up-to-date information on the costs of assessment systems currently in place throughout the country. This report seeks to fill this void by providing the most current, comprehensive evidence on state-level costs of assessment systems, based on new data gathered from state contracts with testing vendors.

We find that the 45 states from which we obtained data spend a combined \$669 million per year on their primary assessment contracts, or \$27 per pupil in grades 3-9, with six testing vendors accounting for 89 percent of this total. Per-pupil spending varies significantly across states, with Oregon (\$13 per student), Georgia (\$14), and California (\$16) among the lowest-spending states, and Massachusetts (\$64), Delaware (\$73), and Hawaii (\$105) among the highest spending. We find that larger states tend to spend substantially less, per student, than smaller states, which is not surprising given that larger states save on fixed costs like test development by spreading them over more students and may have more bargaining power.

We estimate that states nationwide spend upwards of roughly \$1.7 billion on assessments each year, after adjusting the \$669 million figure to (1) account for the fact that six percent of students are located in states for which we were unable to obtain data, (2) reflect spending on assessments not included in states' primary assessment contracts, and (3) include state-level spending on assessment-related activities that are not contracted out. This seemingly large number amounts to only one-quarter of one percent of annual K-12 education spending. Were all statewide assessment activities to cease and the funding used to hire new teachers, the pupil-teacher ratio would only fall by 0.1 students. If instead the costs were devoted to an across-the-board pay increase for teachers, the average teacher would see her salary increase by one percent, or about \$550.

This relatively low level of spending on assessment, combined with concerns that the quality of tests in many states is not high enough to use them for high-stakes purposes such as teacher evaluation, strongly suggests that states should seek efficiencies in order to absorb budget cuts without compromising test quality or to free up resources that could be reinvested in upgrades to assessment systems. A clear strategy for cost savings suggested by our data is for states to collaborate on assessments so as to share the fixed costs of test and item development over larger numbers of students. Our cost model predicts substantial savings from collaborating on assessments. For example, a state with 100,000 students that joins a consortium of states containing one million students saves an estimated 37 percent, or \$1.4 million per year; a state of 500,000 students saves 25 percent, or \$3.9 million, by joining the same consortium.

Collaborating to form assessment consortia is not a new idea, and is in fact the strategy being pursued by nearly all of the states that have adopted the Common Core standards. Our model cannot be used to estimate the cost of the tests being developed by the Common Core consortia because they include innovative features not part of most existing systems and because they are substantially larger (in terms of students covered) than any existing state assessment system. But our model does suggest that these consortia will create opportunities to realize significant cost savings, all else equal, compared to the current model of most states going it alone.

The great advantage of cost savings achieved through an assessment consortium is that they do not necessarily have implications for test quality. But it is not yet clear whether larger consortia, like the two that are developing assessments for the Common Core, are a better choice than smaller ones formed more organically. It is clear, however, that in order to make informed decisions about assessment contractors, states need access to good information on pricing. In the current system, states solicit bids for their assessment contracts, but often receive information on pricing that is confusing or lacking in detail. Consortia of states (or even larger states alone) should use their market power to encourage test-makers to divulge more details of their pricing models.

More transparency in the pricing of assessments will be especially crucial when federal funding for the development of the Common Core assessments ends in September 2014, a full six or more months before the first operational tests are administered. States will presumably have to share in the costs of sustaining these assessments, and will need to be able to predict the cost of doing so in advance of deciding whether to participate. At the same time, states should use this time of transition to conduct parallel experiments with smaller-scale collaboration on assessments in other subject areas. Only this sort of experimentation will produce the kind of evidence that is needed for states to design and implement high-quality assessment systems at a cost they can afford.

Strength in Numbers: State Spending on K-12 Assessment Systems

Introduction

The most significant change to state assessment systems since the passage of NCLB is now on the horizon with the widespread adoption of the Common Core State Standards.

In the last few decades, standardized tests have become ubiquitous in U.S. elementary and secondary education. Some states have had assessment systems in place for decades, while other implemented them as part of state accountability laws passed in the 1990s. The remainder came on board as a result of the federal No Child Left Behind Act of 2001 (NCLB), the reauthorization of the Elementary and Secondary Education Act which required all states to have such systems. NCLB also expanded the number of students covered by existing systems in most states by requiring the administration of math and reading tests in grades 3-8 and once in high school. Less than a year before NCLB was passed, no more than 13 states met this requirement, according to data collected by the Pew Center on the States.¹

The most significant change to state assessment systems since the passage of NCLB is now on the horizon with the widespread adoption of the Common Core State Standards, a state-led effort coordinated by the Council of Chief State School Officers and the National Governors Association to develop and adopt common standards in core academic subjects, beginning with math and English language arts (ELA). The primary goal of this effort is “to provide a clear and consistent framework to prepare our children for college and the workforce.” The developers of these standards intend for them to be rigorous (i.e. more challenging than existing standards in most states) and aligned with the expectations of higher education institutions and employers, so that a student who meets the standards will be ready for college or a career.² Although adoption of the new set of standards is voluntary, the Obama administration threw federal support behind the effort by rewarding extra points in its “Race to the Top” competition to states that adopted the standards.

State assessments are designed to measure students’ mastery of the content specified by the state’s standards, so as states change their standards by adopting and implementing the Common Core they will need to revise or replace their tests. Forty-five states and the District of Columbia adopted the Common Core standards,³ and nearly all of them decided to collaborate on the assessment aspect of this effort by joining one or both of two consortia granted federal contracts under the “Race to the Top” program to develop new assessments aligned with the new standards.⁴ The two consortia, the Partnership for Assessment of Readiness for College and Careers (PARCC) and the Smarter Balanced Assessment Consortium (SBAC), are expected to develop assessments that are innovative and more sophisticated in terms of the types of test questions included and their administration and scoring.

The Common Core effort has prompted concerns about the cost of implementing the new standards and associated assessments.⁵ Estimates of these costs are based primarily on assumptions and guesswork, in large part because there is little comprehensive and current evidence on the costs of assessment systems currently in place throughout the country. The most comprehensive nationwide data were collected about a decade ago, in separate investigations by Caroline Hoxby and the Pew Center on the States. Hoxby reported that, according to the Association of American Publishers, test-makers' revenue in 2000 amounted to \$315 million, or less than \$7 per student (0.06 percent of per-pupil spending).⁶ (All dollar figures in this report are converted to 2012 dollars.) For the 25 states from which she collected expenditure data (reproduced in Table A1), Hoxby found that per-pupil costs of accountability systems (including assessments as well as other activities) ranged from as little as \$2 to as much as \$44.

The Pew Center on the States surveyed all 50 states about their spending on test development, administration, and scoring in 2001.⁷ States reported spending a total of \$552 million on testing that year, with per-pupil figures ranging from \$0 in Iowa (which had no statewide testing program at the time) to \$44 in Delaware. This total is significantly more than the revenue figure from the test-making industry, but still amounts to only 0.11 percent of per-pupil spending nationwide (about \$12 per student). The larger figure is likely the result of states including expenses related to assessments that are not paid to test-makers, and perhaps some amount of over-reporting by the state officials responding to the Pew survey.

The state-by-state data gathered by Pew are reproduced in Table A1. The differences between the Pew numbers and Hoxby's are likely explained by the fact that the Pew survey asked about costs of testing whereas Hoxby gathered data on all activities related to accountability (not just assessment). Consequently, Hoxby's figures should be larger, and they usually are. The exceptions to this general pattern highlight the difficulty of accurately measuring state spending on assessments, and may reflect errors in the responses to the Pew telephone survey or in the information provided to Hoxby.

It is unsurprising that testing costs increased in the decade following the passage of NCLB, as states expanded the use of their existing assessments to cover additional grades and developed new tests. The most current and comprehensive publicly available data on state spending on assessments were collected by the Common Core assessment consortia, SBAC and PARCC, through surveys of their member states in 2010.⁸ The SBAC states reported spending between \$7 and \$123 per student on their math and ELA assessments.⁹ For the vast majority of states, these numbers are substantially more than spending almost ten years earlier but still represent a small share of overall expenditures. The state that spent by far the most on testing on a per-student

Given rising pressures on schools to “do more with less,” all expenditures should be examined for their cost-effectiveness

basis (Hawaii) still devoted less than one percent of total spending to this purpose.¹⁰ PARCC reported spending per test, as compared to per student, but assuming that the average student takes two to four tests, the overall pattern of assessment costs is similar.

Costs that amount to less than one percent of per-pupil spending may seem trivial, but warrant careful examination for several reasons. First, given rising pressures on schools to “do more with less,” all expenditures should be examined for their cost-effectiveness, and even small amounts of spending or savings add up across our nation’s K-12 education system. Spending in U.S. public schools totaled \$658 billion in 2008-09 (the most recent year for which data are available), so even one-half of one percent would add up to more than \$3 billion each year.¹¹ And states can make changes to their assessment budgets with relative ease compared to some larger categories of expenditures, such as employee salaries, which are often constrained by collective bargaining agreements. For example, Georgia cancelled the upcoming spring 2013 administration of its state test to first- and second-grade students due to budget constraints.¹²

Second, there is the risk of multi-million-dollar assessment contracts contributing to a political backlash against testing among parents and taxpayers who oppose the use of standardized testing for accountability purposes or object to public dollars flowing to for-profit companies (as most of the testing contractors are). For example, the *Wall Street Journal* recently reported that “[a] national coalition of parents and civil-rights groups, including the NAACP Legal Defense and Educational Fund, signed a petition in April [2012] asking Congress to reduce federal testing mandates.”¹³ These anecdotes aside, Americans remain broadly supportive of testing. In a 2011 survey, only nine percent of all respondents (and 11 percent of parents) said they were opposed to the federal government requiring states to test students in math and reading in grades 3-8 and once in high school (as mandated under NCLB).¹⁴ But as education budgets continue to tighten, expenditures on testing may draw increased scrutiny.

Third, change is afoot in testing systems across the country. Cheating scandals have prompted concerns about test security, especially as more districts tie test scores to the evaluations of teachers—who often proctor their own students’ exams. Criticism of multiple-choice “bubble tests” has increased interest in moving toward exams with other types of items, such as performance tasks that are designed to assess students’ analytical reasoning skills more effectively than a question with a single correct answer. Some states have started to move towards computerized assessments, and both of the tests being developed by the Common Core consortia are computer-based. All of these proposed upgrades of assessments systems have implications for costs, and will be scrutinized in light of competing demands on state budgets.

And finally, though the federal government financed the initial development costs of the new assessments being created by the Common Core consortia, states will likely have to fund the maintenance and enhancement of these assessments after the federal grant ends in September of 2014, a full six or more months before the first operational tests are administered. It is not yet clear how participating states will share the cost of sustaining the consortia assessments, or whether the federal government will provide additional support for this effort. What is clear is that states that currently have inexpensive assessments will be under pressure to spend more to pay for the ongoing costs of the consortia assessments. For example, SBAC currently estimates that its summative assessment will cost about \$20 per student.¹⁵ This amount is less than many SBAC states currently report spending, but represents a cost increase for six states (see Table A1). These states will have to decide whether the benefits of their continued participation are worth the increase in costs.

This report provides the most current, comprehensive information on state-level costs of assessment systems, based on data gathered from state contracts with testing vendors. In addition to calculating the overall and per-pupil costs of each state's main assessment contract (for the 45 states from which we obtained data), we also describe key elements of the math and reading tests, including the number and types of items included. Our descriptive model of costs indicates that enrollment is the most consistent predictor of per-pupil costs, with larger states receiving substantially lower prices, on average, than smaller states. After discussing evidence on additional aspects of testing systems based on a handful of case studies, we conclude with a discussion of the predicted cost savings from joining assessment consortia.

Measuring Assessment Costs

The costs of assessment systems are incurred at multiple levels. Teachers in individual schools spend time administering and scoring tests.¹⁶ Schools and districts are responsible for coordinating certain aspects of test administration and, in some states, are responsible for scoring exams. States have assessment offices that employ staff who oversee statewide testing efforts and contract with testing vendors. Under these contracts, the test-making entities are responsible for a variety of activities related to test development, administration, and scoring.

This report focuses on the costs of contracts between states and test-making vendors because they constitute the lion's share of state-level expenditures on testing. According to assessment cost data gathered by PARCC from its member states, of 21 states that provided both total assessment cost and contract cost data, 18 states reported contract costs making up more than 85% of total costs. Other state-level costs are surely important, such as the salaries paid to state assessment officials who play a vital role in selecting contractors and overseeing

the vendors through test development, administration, and scoring. But such costs are difficult to track consistently across states, and usually represent a small fraction of the testing budget.

The roles played by school and district employees who aid in test administration and scoring are important as well, but the cost of this work is challenging to measure. Calculating such costs requires information on which employees have these responsibilities, their compensation levels, how much time they devote to test-related activities, and what work they would be doing if they weren't involved in testing. Future research should attempt to measure how significant these costs are, how they vary across different types of tests, and whether there are efficiencies to be gained by outsourcing more of the responsibilities currently delegated to teachers and administrators.

By focusing on contract costs, this report is able to include data for most states in the nation and generate findings relevant to state policymakers seeking to economize on assessment costs in difficult economic times. Staff in the Brookings Institution's Brown Center on Education Policy obtained contracts between states and testing vendors for several recent years. These contracts were obtained through a combination of direct requests to state assessment offices and Freedom of Information Law (FOIL) requests. A handful of states provided contract budget data in lieu of copies of the actual contracts.

For each state that provided contracts (or contract budget data), we identified the contract(s) associated with the main state assessments in math and reading. We recorded the total contract amount for the main contract for each of the years 2007 through 2012 (referring to school years using the calendar year of the spring semester, which is when tests are nearly always administered). The total contract amount was updated with revised contract amounts (or amounts actually paid) using data from contract amendments when they were available. These yearly amounts were all adjusted for inflation (to 2012 dollars), and then averaged to create a single yearly amount for each state.

Contract data were relatively straightforward to obtain from states, but cannot be precisely compared across states. Our focus on data from the main assessment contract(s) was meant to make the data as comparable as possible. However, some contracts include additional assessments (beyond math and reading), such as tests in science, social studies, and writing; high school end-of-course and graduation exams; and alternate assessments for students with disabilities. In some cases, we were able to subtract the costs of large line-item expenditures such as alternate assessments. In other cases, however, the budget data were not broken down in a way that allowed us to extract the cost of the math and reading assessments.

Several states in our data switched contractors at some point during the period covered by the contracts provided to us (in many cases to hire a new contractor to develop a new test). For such states, we use data on the more

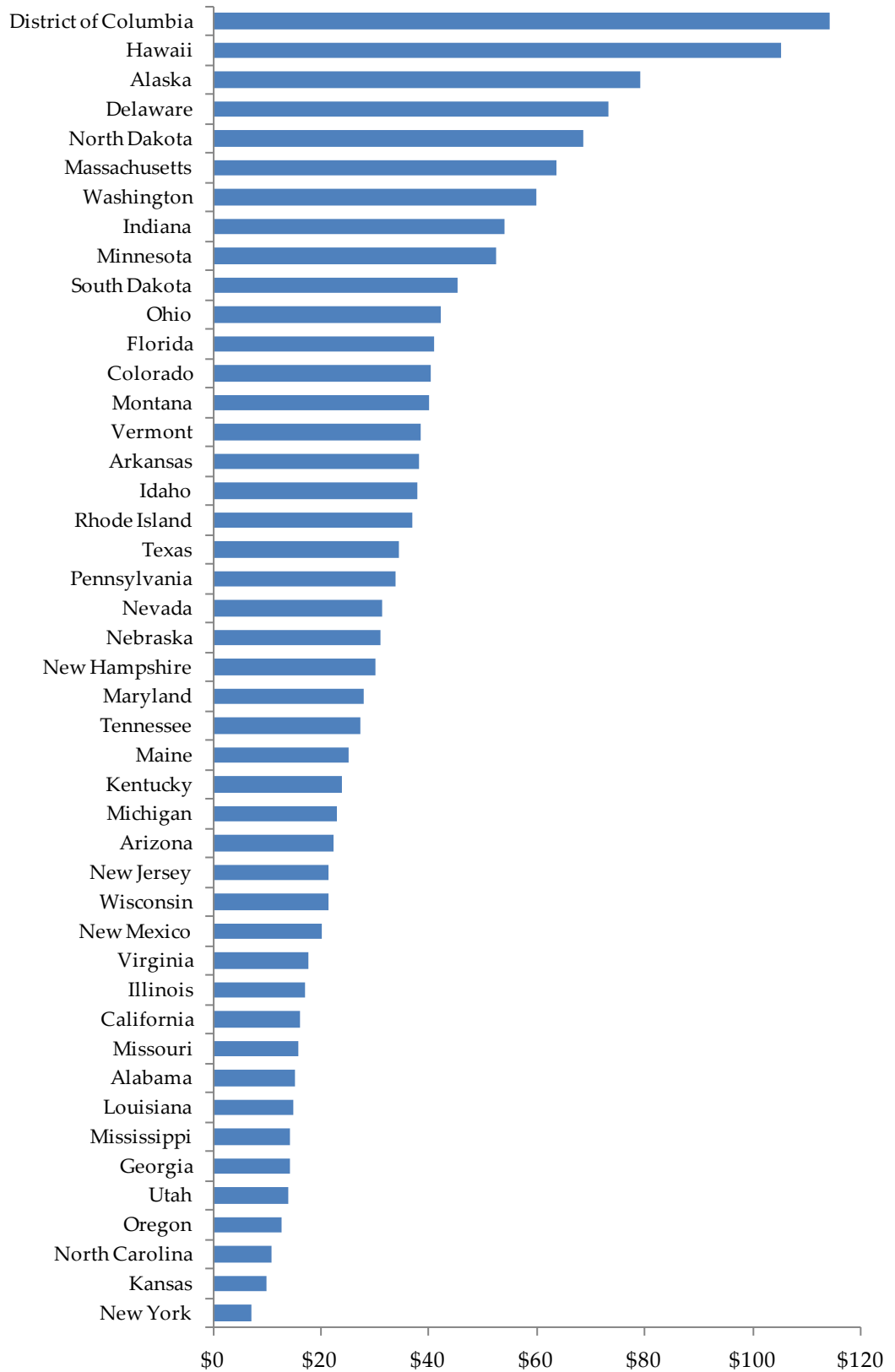
recent test/contract when the data from the new contract are sufficiently complete. However, we generally did not use the data from the new contract if it largely covered future years and thus could not have been amended to reflect the adjustments to contract costs that are often made after the original contract is signed (which we use to revise the contract budget amounts in our data whenever possible).

Most contracts cover multi-year periods, and for new assessments the period often includes years when states pay costs related to the development of a new test that is not yet operational. For these contracts, we either ignore the development costs (instead focusing on the contract costs during operational test years) or divide the development costs equally over the operational years. Which approach we took depended on the nature of the contract budget data. For example, in the case of a multi-year contract that did not contain clear allocations by year, we divided the contract evenly over the years for which the vendor was responsible for operational tests.

These data represent the most comprehensive and up-to-date information on state contract costs related to their primary assessments, but they are far from perfect. Consequently, readers should take care when comparing data for different states, especially given that different states include different numbers of assessments (covering different numbers of students) in their main assessment contract(s). For example, a state may appear to spend more on a per-student basis on its assessment system compared to another state when in fact the higher-spending state's contract simply includes science and social studies tests whereas the other state's contract only includes the core math and reading assessments. We investigate below whether contracts that include more tests systematically have higher per-student costs.

We obtained usable contract data for 44 states and the District of Columbia. These 45 jurisdictions contain 94 percent of U.S. students in grades 3-9. (We were not able to obtain the main assessment contracts for Iowa, Oklahoma, South Carolina, West Virginia, or Wyoming.) Table A2 lists, for each state, the years of data included, the average yearly amount (in current 2012 dollars), the number of students enrolled in grades 3-9 (averaged over the years for which we have contract data), the cost per student, the primary contractor, and the number of additional test types included (up to five, which include writing, science, social studies, high school graduation or end-of-course exams, and alternate assessments).¹⁷

Figure 1. Per-Pupil Costs of Main Assessment Contracts, by State

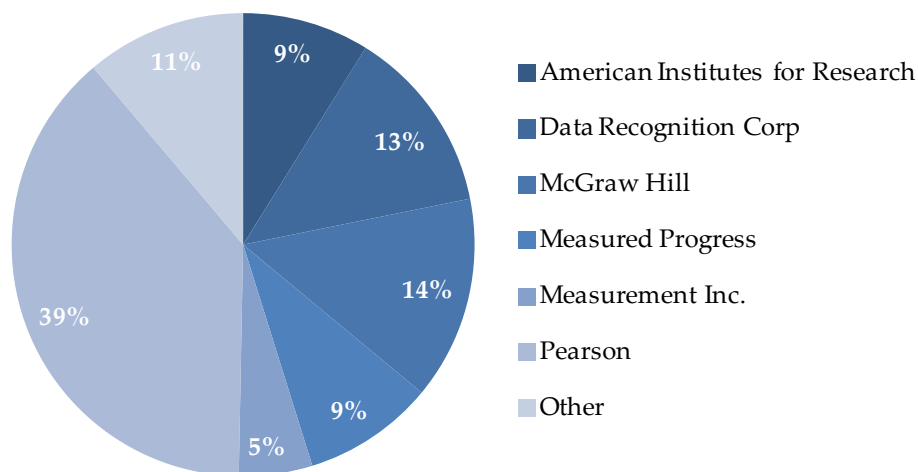


Source: Table A2
 State Spending on K-12 Assessment Systems

The 45 jurisdictions in our data account for about \$669 million in average yearly spending, or \$27 per pupil in grades 3-9 (we use enrollment in these grades because NCLB requires states to test students in grades 3-8 and once in high school). As Figure 1 shows, per-pupil costs range from \$7 in New York to \$114 in the District of Columbia. These are both unique jurisdictions because DC is a single city and because assessment scoring is a local responsibility in New York (and therefore not included in the contracted costs). The next two lowest-spending states are Kansas and North Carolina, both of which are unique because their contractors are state universities. Ignoring New York, DC, Kansas, and North Carolina yields a range from \$13 in Oregon to \$105 in Hawaii. These figures are roughly consistent with SBAC's data from 2010, which found that Hawaii spent the most per student and Oregon was among the lowest-spending states.

A relatively small number of test-makers account for most of the contracted costs in the U.S. Figure 2 shows that six vendors account for 89 percent of the main assessment contract dollars, with a single vendor (Pearson) making up 39 percent of the market. A small number of states, including North Carolina and Kansas, contract with a test-making center in one of their public universities, but that is the exception. Test-makers' market share calculated based on student enrollment rather than contract dollars follows a similar pattern, with a smaller number of vendors covering the vast majority of enrolled students (Table A3).

Figure 2. Share of Main Assessment Contract Dollars by Vendor



Source: Table A3

The 45 jurisdictions in our data account for about \$669 million in average yearly spending.

Per-pupil contract costs vary noticeably by vendor, from a low of \$10-\$11 at the state universities in North Carolina and Kansas to a range of \$22-\$42 at the six main contractors that serve at least two states. There are two primary reasons that average costs vary by vendor. First, these vendors serve states with wide-ranging assessment systems, for example in terms of the amount of customization that is required to develop a test specific to an individual state. Second, prices can vary widely by vendor for the same exact test. This is seen most clearly when different vendors submit bids for the same contract. For example, in 2008 the Florida Department of Education considered bids from two bidders for a four-year contract for the development and administration of its assessment system. Pearson's winning bid received higher scores from reviewers, and came in at about \$200 million (37 percent) cheaper, than CTB McGraw-Hill's bid.¹⁸

In addition to characterizing assessment contract costs and the types of additional tests included in those contracts, we also gathered data from state education agencies on the number and types of items included in state tests.¹⁹ We differentiated between multiple-choice and non-multiple-choice items because the latter are expected to be more expensive to develop and score than the former. Examples of non-multiple choice items include short answer questions and essays.²⁰ Specifically, we obtained the number of each type of items included in the fourth- and eighth-grade reading and math tests. We chose these grades so as to have one grade from elementary school and one from middle school.

These data, which were obtained for all states except three (Maryland, Missouri, and Utah), are reported in Table A4 and indicate that states vary widely in the number of items included on state tests.²¹ For example, the average state has 52 multiple-choice items on its fourth-grade reading test, but this number ranges from 24 in Connecticut to 117 in Arkansas and 185 in Iowa. In most states, the large majority of questions are multiple-choice. However, counting items likely understates the importance of non-multiple-choice items given that these items tend to be more involved and count for more points (per item) than multiple-choice questions.

State assessment systems vary significantly in terms of their characteristics and their costs, but what is the relationship between characteristics and costs? Are there any measurable features of assessment programs, or the states themselves, that are systematically associated with costs? We address this set of questions by developing a descriptive model of contracted assessment costs.

Model of Assessment Costs

Assessment contract costs are in some ways like the price of airline tickets. A prospective passenger expects to pay more for travelling a greater distance, buying a ticket at the last minute, or traveling during a popular time of year. But the exact formula that translates these factors and others into the ticket price is opaque. When shopping for a ticket the would-be traveler simply sees a final price for a given trip.

Likewise, states expect to pay more for assessments that cover more grades and subjects, for developing more items and forms, and for replacing multiple-choice questions with items that cannot be scored by a computer. Like the budget airline ticket seeker on Kayak.com, the state assessment office can put out a request for proposals and seek bids from prospective vendors. The bids (and contracts with the winning bidders) often contain information on pricing, including detailed breakdowns in some cases, but they do not follow a formula that translates assessment system features into a price. Of course state assessment offices can negotiate with potential vendors, so two states may pay different amounts for roughly the same assessment contract.

The types of costs included in contracts with test-makers can be roughly categorized based on whether they are variable (i.e. increase as the number of students tested increases) or fixed (i.e. remain the same regardless of the number of students tested) and whether they are one-time or ongoing. A handful of examples are shown in Table 1.

Table 1: Examples of Assessment-Related Activities

	Fixed	Variable
One-time	Design of new test, item development Creation of new data system to store/analyze test-score data Determination of scores that indicate student “proficiency”	Statewide training sessions and communications about a new test in the initial year
Ongoing	Development of new items Development of multiple forms of same test to mitigate cheating Upgrades to assessment system	Printing/mailing test booklets Scoring tests Reporting results

The existence of both fixed and variable costs implies that, all else equal, larger states will pay less per student than smaller states because larger states spread their fixed costs over a greater number of students.²² For example, a state with 100,000 students that pays \$500,000 for item development and \$10 per test for administration and scoring will have total per-student costs of \$15. A state

The key finding of our model is that the number of students is the only factor that is consistently related to per-student expenditures.

with 1,000,000 students that faces the same prices will pay \$10.50 per student because the fixed costs amount to only \$0.50 per student. The larger state may in reality face even lower prices if it has greater bargaining power vis-à-vis the testing vendors because of the total contract being much larger than that of a smaller state.

Contract costs are also expected to be related to the tests covered by the contract, both in terms of their number and their characteristics. The addition of science or writing tests should add to the total cost, as should the use of items that are more expensive to develop and score. The division of responsibilities between the state and vendor will also affect the contract price. For example, contract costs will be lower if scoring is done by local schools rather than by the vendor.

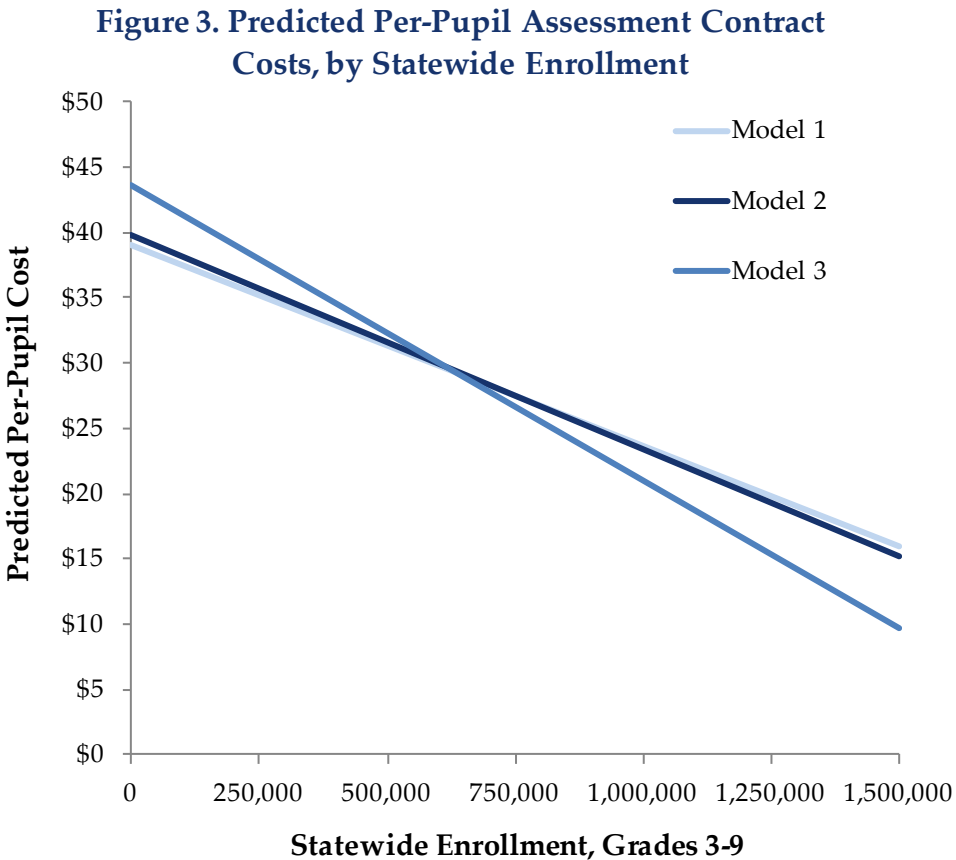
With detailed data on contract costs broken down by subject, grade, number of students, number and types of items, etc., one could build a model of these costs “from the ground up.” Ideally the data would include not just how much was charged for a given category of expenditures but how the price was computed. For example, the contract would indicate that the cost of \$80,000 to develop 50 multiple-choice items and 10 short-answer questions for a third-grade reading test was calculated as \$1,000 per multiple-choice item and \$3,000 per constructed-response question. This kind of detailed information could be used to build a model that estimates total assessment costs (perhaps averaging over the practices of different vendors) as a function of the key drivers of those costs as recorded in contracts.

In reality, assessment contract budgets vary widely in terms of the amount of detail they include and very few provide line-item amounts much less the underlying formulas used to calculate them. Some state contracts include little more than total amounts, usually broken down by year, and in some cases broken into large categories such as subject area and activity type (development, scoring, etc.). In order to take advantage of data from as many states as possible, we used the procedures described above to calculate basic information about each state’s primary assessment contract, most importantly average yearly spending and the number of additional assessments included.

We use these data to fit a regression model of per-student assessment costs based on a relatively small number of factors that theoretically should be associated with costs. We focus on per-student costs in order to make it easier to interpret differences between larger and smaller states, as the number of students is the largest driver of overall spending given the significant amount of variable costs in assessment contracts. This descriptive model allows us to measure whether per-student costs are systematically related to each of the following factors: the number of students in the state in grades 3-9, the number and types of additional tests covered by the contract, average teacher salary in the state (a proxy for labor costs and the cost of living in the state), the average number of

items on the four tests we gathered data on, and the share of those items that are multiple choice. The specifics of this model are discussed in greater detail in Appendix B.

The key finding of our model is that the number of students is the only factor (among those we examined) that is consistently related to per-student expenditures. Figure 3 shows the estimated relationship using three different variants of our model. Model 1 examines the unadjusted relationship between enrollment and per-pupil costs, whereas Model 2 also takes into account the number of additional assessment types included, and Model 3 adds controls for average teacher salary, the average number of items, and the share of multiple choice items.



Source: Based on coefficients in Table B1

All three variants of this model indicate that per-pupil costs are significantly lower in larger states than in smaller states. The range of enrollments included in Figure 3 covers all states except for California (3.3 million students) and Texas (2.5 million). The most conservative estimates (Model 1) indicate that moving from a state with about 100,000 students in grades 3-9 (such as Maine or Hawaii)

Moving from a state with about 100,000 students in grades 3-9 to one with about 1,000,000 students is associated with a decrease in per-pupil assessment costs from \$37 to \$24, a savings of \$13 or about 35 percent.

to one with about 1,000,000 students (such as Illinois) is associated with a decrease in per-pupil assessment costs from \$37 to \$24, a savings of \$13 or about 35 percent. In other words, if the smaller state in this example could get the price paid by the larger state, the smaller state would save \$1.3 million per year. Conversely, if the larger state faced the per-pupil cost of the smaller state, it would pay \$13 million more per year than it currently does.

The other factors we included in our analysis were not consistently related to per-student assessment costs. We expected the number of additional assessment types included in the contract, average teacher salary, and the average number of test items to be positively associated, and the share of multiple choice items to be negatively associated, with costs. These estimated relationships are reported in Appendix B, but the bottom line is that they are measured with a great deal of imprecision and sometimes point in the opposite direction from what was expected.

It is important to emphasize that these results do not mean that these additional factors are unimportant in the determination of contract costs. The results only mean that we could not find a significant relationship between these factors and costs in our data. This likely results from contract costs being measured with some error in terms of consistency across states (given the different activities and tests included in different states' contracts) and the limited number of observations inherent in state-level datasets (even complete data would only have 51 data points).

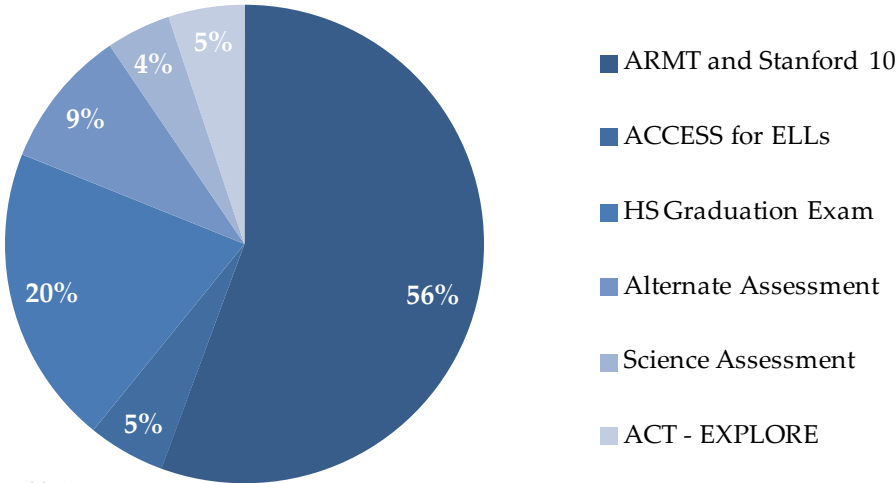
Analyzing data that cover only a limited number of factors and are measured imprecisely is necessary in order to include as many states as possible and thus for the results to be relevant nationally. However, some states provided more detailed contract budget data that enable us to use a "case study" approach to examine a set of questions that we could not tackle using the national dataset. Specifically, we present evidence from a handful of states on the cost of assessments other than the main math and reading tests, costs by subject area, costs of different types of assessment-related activities (e.g., development, administration, and scoring), and the costs of developing new assessment systems.

Costs of Different Tests and Testing Activities

Most states administer tests in addition to the reading, math, and science tests required by NCLB.²³ These include tests in subject areas such as social studies and writing that are administered in elementary and middle school; end-of-course tests administered in high school; high school graduation exams; college entrance exams such as the ACT; modified assessments for students with disabilities; and others. Our focus on the main assessment contract enabled us to make the data more comparable across states, but a downside of doing so is ignoring the costs of states' other assessment activities.

To address the cost of assessments other than the ones covered by the primary contract, we examined all assessment-related contracts from five states for years in which the same assessments were given.²⁴ On average, the main assessment accounted for about half of the total contract costs, ranging from 37 percent in Maryland (which spends a significant amount on high school assessments) to 62 percent in New Hampshire. The data for all five states are presented in Table A5, and Figure 4 highlights Alabama. In 2011, the main assessments in Alabama (the ARMT and Stanford 10 in math and reading) accounted for 56 percent of spending, with the high school graduation exam accounting for another 20 percent. The remaining quarter of spending was divided among a test for English language learners, a science test, the alternate assessment, and the ACT.

Figure 4. Spending on Assessment Contracts in Alabama, 2011

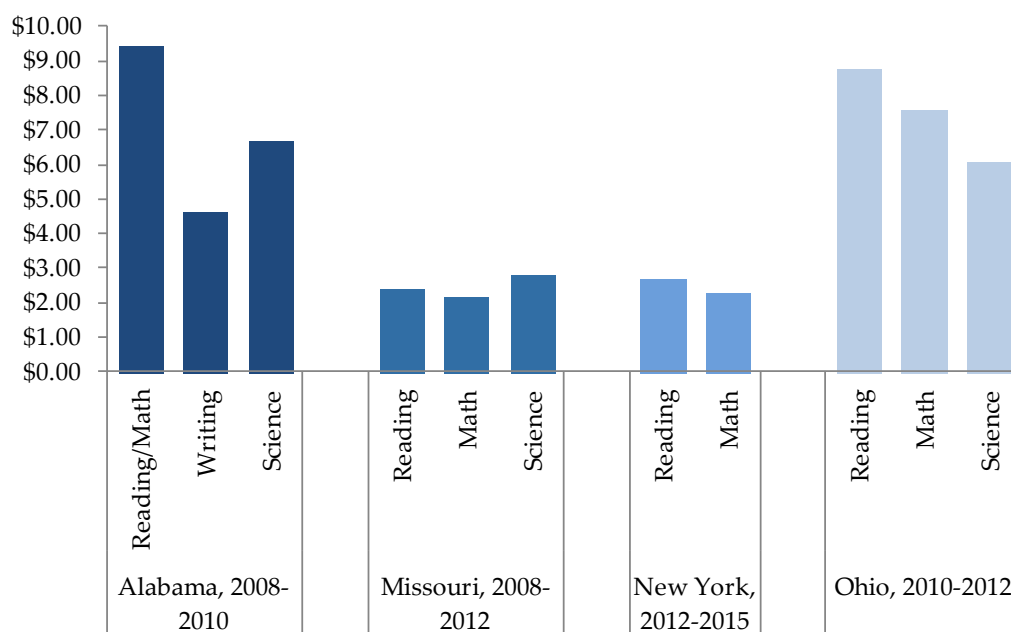


Source: Table A5

The total amounts that states spend on different assessment contracts provide a sense of the range of assessments administered by states, but do not provide any information about whether certain types of activities are more expensive than others on a per-student basis. For example, the Alabama data presented in Figure 4 (depicting total spending by assessment contract) do not tell us whether the science assessment is more expensive on a per-student basis than the reading and math tests because although the science test is cheaper overall it is administered to fewer grades. However, we are able to calculate per-student assessment costs by subject in four states for which we have both total spending by subject and can estimate the number of students tested in each subject.

Figure 5 presents per-student contract costs by subject area for these four states: Alabama, Missouri, New York, and Ohio. The data do not reveal any consistent patterns of spending by subject. One might expect reading tests to be more expensive than math tests because reading tests tend to include more non-multiple-choice items that cost more to score. But the differences in reading and math costs in the three states where they can be separated are modest, with reading costs only \$0.20 per student more in Missouri, \$0.40 more in New York, and \$1.01 more in Ohio. Science tests are slightly more expensive than reading and math tests in Missouri, but noticeably less expensive in Alabama and Ohio. The Alabama writing test costs less per student than both reading/math and science.

Figure 5. Per-Pupil Assessment Contract Costs, by Subject

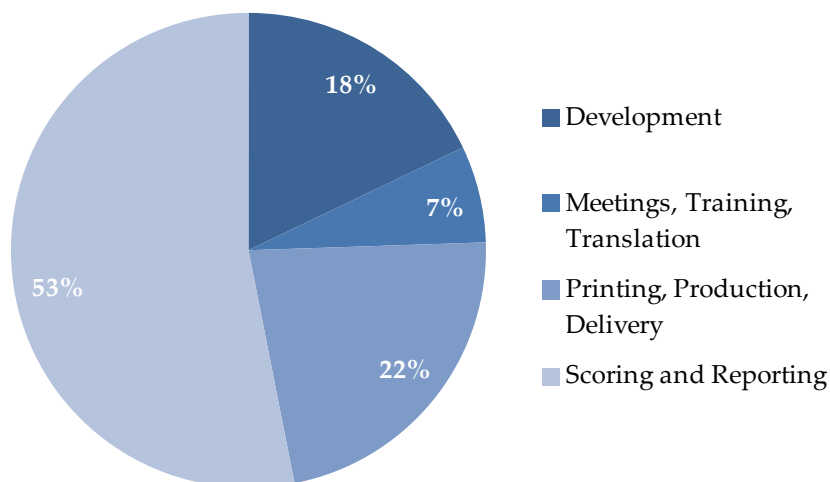


Source: State Assessment Contracts

Costs vary somewhat, albeit inconsistently, by subject area in the states that we examined, but the division of costs by broad activity categories yields a more consistent pattern of results. Figure 6 shows assessment contract costs for different categories of expenditures in New Jersey. A majority of costs are devoted to scoring and reporting the exams, with about a quarter of costs going to printing, production, and delivery, and the balance going to development, meetings, training, and translation. Condensing these categories further indicates that about three-quarters of costs are going to test administration

(scoring, reporting, printing, producing, and delivering) with the other quarter covering development costs (meetings, training, and translation).

Figure 6: Assessment Contract Costs by Activity Type, New Jersey, 2008-2012

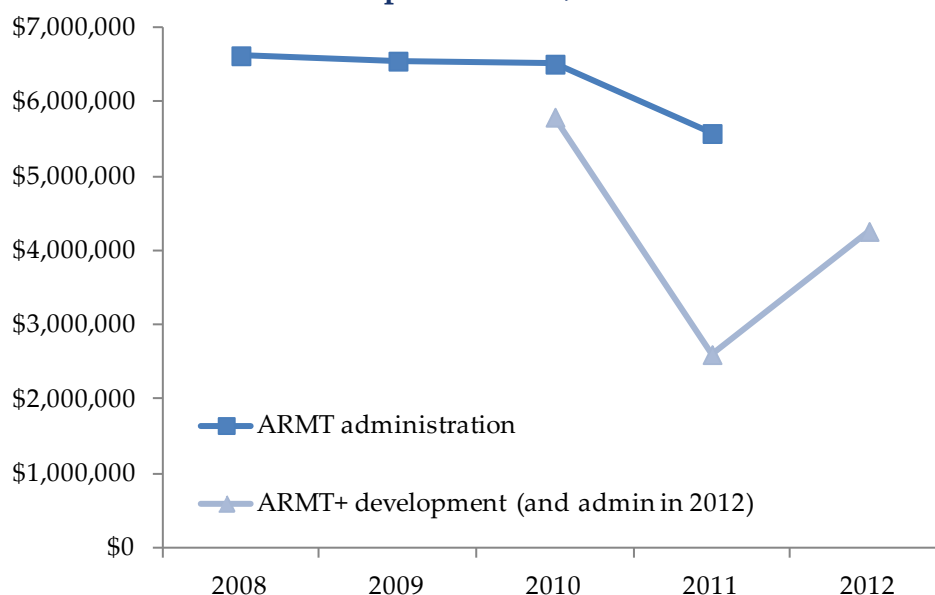


Source: Table A6

This breakdown is similar to that reported in the contracts of two other states, as shown in Table A6. The split between development and administration in Missouri’s contract was 28/72 percent. In New Mexico, the split is roughly the same if one counts the materials and scoring costs that are billed directly to districts (and thus not included in our main analysis). The New York data do not follow this pattern, but that is likely because test scoring is a local responsibility in that state and thus is not included in the contract with the test-maker (except for the cost of training scorers and scoring a sample of exams from the state).

Finally, we examine the costs of developing new tests (which are distinct from ongoing item development costs). As discussed earlier, our nationwide analysis either ignores test development costs because they were incurred in earlier years or divides them evenly over the operational years of multi-year contracts. Figure 7 presents contract cost data from Alabama for the administration of its previous test (ARMT) and the development (and one year of administration) of its new test (ARMT+). The ARMT+ follows the same standards as the ARMT and has the same scoring, format, and item types. Key differences between the tests are the addition of a science assessment for grades 5 and 7 and increased rigor overall. For example, on the math tests the problems were rewritten at a higher level of difficulty, and on the ELA tests the reading passages were lengthened.²⁵

Figure 7. Assessment Administration and Development Costs, Alabama

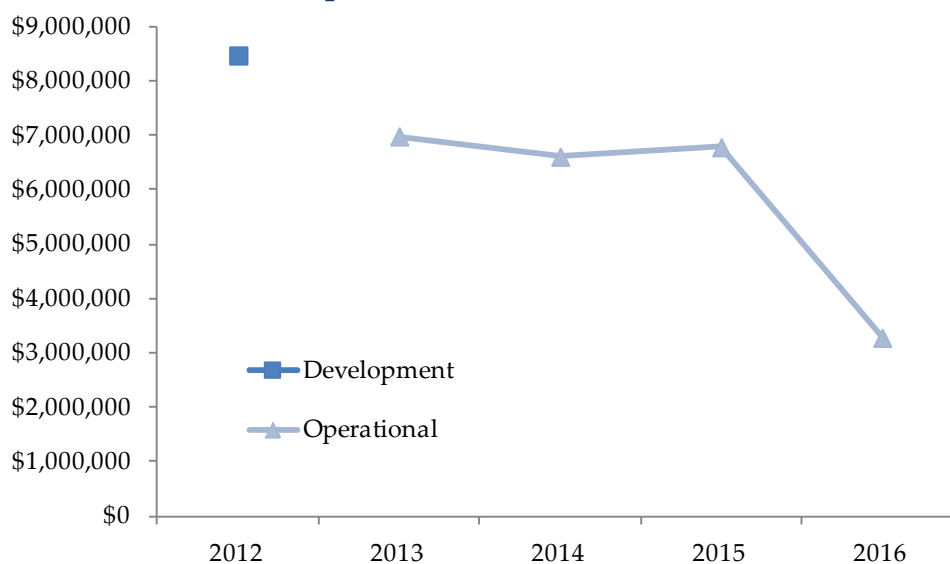


Source: Author's calculations from state assessment contract data.

The up-front costs of developing the new test were substantial, with first-year development costs almost doubling the state's total spending on its main assessments (both old and new) in 2010. Development costs were lower in the second year, and then increased in the third year when the new test was implemented statewide for the first time. This example highlights how the handling of development costs affects the calculation of overall costs. For example, the 2012 costs of the ARMT+ were \$4.3 million. If instead all three years are counted for this one operational year, the total is \$12.7 million. But if the fourth year were to cost \$4.3 million, then the two-operational-year average of the four years of costs would drop to \$8.5 million.²⁶

Whereas the Alabama test was developed over two years, New York had a new test developed in one year, but the one-year development cost of \$8.5 million was similar to the two-year total in Alabama of \$8.4 million (the tests were not developed by the same vendor). Figure 8 shows New York's contracted annual costs in the operational years of \$6.5-7 million in 2013-2015 and \$3.3 million in the last year.²⁷ Dividing the total five-year contract over the four operational years yields a yearly cost of \$8 million, which of course would decrease if the test were kept in place for additional years beyond the initial contract. The Alabama and New York examples highlight how decisions about how to allocate costs across years make it difficult to compare costs across states.

Figure 8. Assessment Costs During Development and Operational Years, New York



Source: Author's calculations from state assessment contract data.

These case studies are not meant to offer comprehensive information on the questions they are used to address, but rather to offer preliminary evidence on the kinds of questions that cannot be addressed with the national data. They also show the kind of detailed information that is available in contracts from some states, which is fertile ground for future research on the costs of assessment systems.

Saving by Collaborating

The new data assembled for this report indicate that total contracted costs of states' primary assessments amount to \$669 million annually. Adjusting this estimate to account for the fact that six percent of students are located in states for which we were unable to obtain data increases the estimated total to \$723 million.²⁸ Continuing this back-of-the-envelope calculation, we can also account for contracted spending on other assessments, which the discussion above indicates may be as large as the main assessments, as well as the state-level spending on assessment-related activities that are not contracted out (most importantly the personnel in state assessment offices).²⁹ The resulting rough estimate of nationwide state-level spending on assessments is \$1.7 billion per year.

Is the United States over- or under-investing in student assessment? Testing critics would likely point to a figure like \$1.7 billion and charge that such a huge sum of money would be better spent on other uses, especially at a time when

Were all assessment activities to cease and the state-level funding to be used to hire new teachers, the pupil-teacher ratio would only fall by 0.1 students, from 16.0 to 15.9.

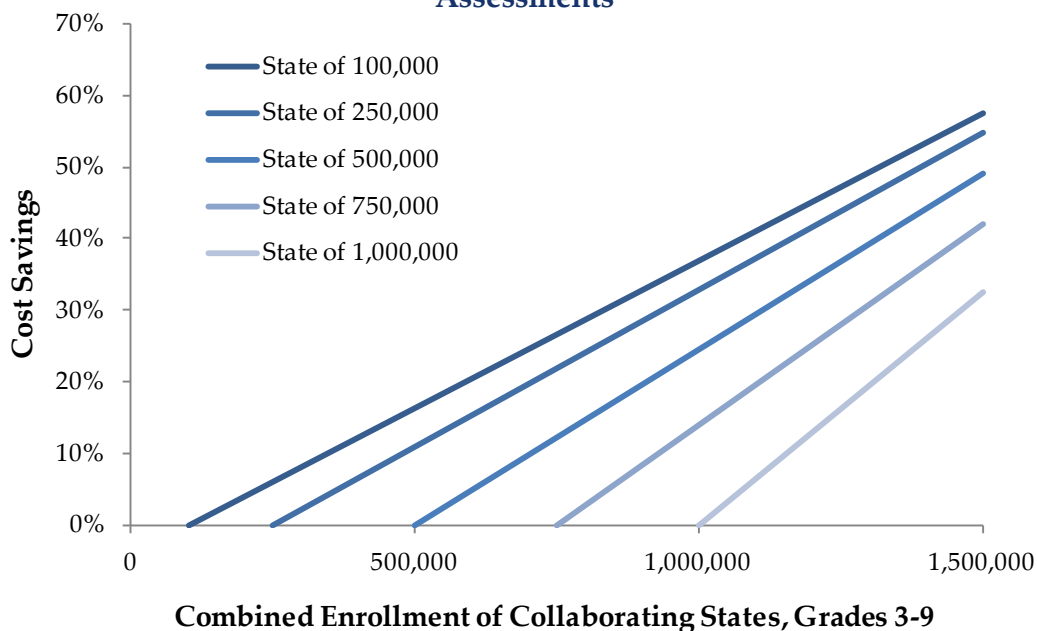
budget cuts are being made in many states. But \$1.7 billion amounts to one-quarter of one percent of annual K-12 education spending in the U.S., or about \$34 per student (\$65 per student in grades 3-9).³⁰ Were all assessment activities to cease and the state-level funding to be used to hire new teachers, the pupil-teacher ratio would only fall by 0.1 students, from 16.0 to 15.9. If instead the costs were devoted to an across-the-board increase in teacher salaries, the average teacher would see her pay increase by one percent, or about \$550.³¹

Another way to look at these figures is that the U.S. is spending too little on assessment systems given the increasingly prominent role played by standardized testing in education policy. Test scores are used to hold all schools accountable as mandated by NCLB; many states require students to pass tests in order to graduate from high school; and districts across the country are starting to evaluate teachers based in part on their students' performance on assessments. Concerns about multiple-choice tests leading to low-quality, "drill and kill" instruction are not concerns about testing per se but about test quality.³² Upgrades to assessment systems that mitigate their unintended consequences are likely to increase costs, but such investments may be warranted given the importance of the tests and the relatively low level of per-pupil spending currently devoted to them.

In times of fiscal exigency, additional resources for upgrades to assessment systems are unlikely to be forthcoming. A clear strategy for cost savings is for states to collaborate on assessments so as to share the fixed costs of test and item development over larger numbers of students.³³ This strategy is not a novel one, and is already being pursued by the 45 states that are members of PARCC or SBAC (or both). But this report provides new evidence on the cost savings likely to be realized through such collaborations.³⁴ These cost savings could be used to obtain higher-quality assessments, reallocate funds to another purpose, or absorb budget cuts.

Figure 9 shows the cost savings (as a percentage of currently predicted costs) for hypothetical states of different sizes, based on the total number of students enrolled in grades 3-9 in the states participating in the collaboration. The savings are based on the most conservative estimates shown in Figure 3. Smaller states have the most to gain in percentage terms because their tests are the most expensive to begin with, but larger states often save more in absolute dollar amounts. For example, a state with 100,000 students that joins a consortium of states containing 1 million students saves 37 percent, as compared to 25 percent savings for a state with 500,000 students that joins the same consortium. But the smaller state saves \$1.4 million, as compared to \$3.9 million for the larger state.

Figure 9. Cost Savings from State Collaboration on Assessments



Source: Author's calculations based on coefficients in Table B1

The great advantage of cost savings achieved through an assessment consortium is that they do not necessarily have implications for test quality. Other methods of cost savings, such as relying more on multiple-choice items or reducing the number of items developed and forms created, may make tests easier to game or cheat. But collaborating with other states achieves cost savings simply by spreading fixed costs over more students, although of course these collaborations do entail the non-monetary cost of using an assessment that is not designed specifically for a single state.

It is important to emphasize that although our estimates are certainly relevant to the work of the Common Core consortia, they cannot be used on their own to estimate the cost savings from joining one of the consortia for at least two reasons. First, the consortia are developing assessments with features such as computer-based testing that were not part of most of the assessments covered by our cost data. Second, we only estimate cost savings for assessment consortia of up to 1.5 million students because that is the range of enrollment for the vast majority of states in our data (only two states, California and Texas, are larger). There surely could be additional savings to forming even larger consortia of states, but we prefer to be cautious by not extrapolating beyond the data underlying our model. PARCC members include 23 states containing 12.9 million students in grades 3-9, and SBAC states number 25 with 10.5 million students.³⁵ Presumably there will be significant savings, all else equal, from spreading the fixed costs of testing over such large numbers of students, but we

There are good reasons why states, especially small ones, should not go it alone when it comes to assessments.

cannot put a dollar figure on those savings because there simply aren't data on which to base such a projection.

This report is also agnostic regarding the Common Core consortia because it is not clear whether it is better for states to have a choice of two large consortia or a larger number of smaller consortia. Larger consortia can spread out fixed costs over larger numbers of students, but without much competition they may not have much of an incentive to operate efficiently. Likewise, a smaller number of larger consortia means that states will have fewer choices and thus may be less likely to find a consortium that is a good match for their education system.

Smaller consortia may not have the same cost advantages as larger ones, but may find it easier to balance the competing needs of member states. The New England Common Assessment Program (NECAP), which New Hampshire, Rhode Island, and Vermont created in the wake of NCLB's requirement to test students in additional grades, is an example of this kind of collaboration. The NECAP was first administered in 2004, and Maine joined the consortium in 2007. One commentator described the choice the New England states faced as "multiple choice or collaboration" because retaining non-multiple-choice items would have been costly had the states chosen to comply with NCLB's requirements on their own.³⁶

Our contract cost data provide some suggestive evidence that participating in NECAP has led to significant cost savings for the participating states. The four NECAP states had average per-student spending of \$33, roughly half of the \$62 spent by the ten other states with total enrollments of less than 200,000. Dropping the two jurisdictions with the highest spending (DC and Hawaii) reduces the difference from \$29 to \$17 per-pupil. Differences between NECAP and non-NECAP states in enrollment, the number of additional tests covered by the contract, and average teacher salary in the state do not explain the apparent cost savings enjoyed by the NECAP states.

There are good reasons why states, especially small ones, should not go it alone when it comes to assessments. Whether larger consortia like PARCC and SBAC are a better choice than smaller ones like NECAP may well vary by state, and it may make sense for states to join or form different consortia for different elements of their assessment systems. For example, even though Maine is a NECAP member it uses its own high school graduation exam. One could imagine consortia forming for different kinds of tests, ranging from elementary science tests to high school graduation exams. Assuming cost-sharing arrangements can be agreed upon, states might choose to join SBAC for math and PARCC for reading, NECAP for science, and form a new consortium for their high school end-of-course exams.³⁷

In order to make informed decisions about assessment contracts, states need access to good information on pricing. In the current system, states usually put out requests for proposals on which contractors then bid. The resulting

Larger states, or consortia of states, should use their market power to encourage test-makers to divulge more details of their pricing models.

agreements between states and contractors often contain little detail on how the bottom-line price was determined, so the only way states can learn the price of alternative versions of the assessment is to ask the contractor for the price of specific configurations. Furthermore, assessment officials are often frustrated by practices such as the frequent asymmetry between the price of adding an item to a contract and the savings from deleting the same item from the contract (the latter is often substantially smaller than the former).³⁸

Larger states, or consortia of states, should use their market power to encourage test-makers to divulge more details of their pricing models. Ideally, states would be provided with the formula used to calculate the price of their assessment contracts and thus be able to manage the trade-offs between cost and quality. But at a minimum, states should have access to breakdowns of assessment costs by item type, know what the (per-student) savings are to testing more students (perhaps by collaborating with another state), etc. Increased transparency in the market place would sharpen competition among the small group of vendors that develop and administer a large majority of the tests in the U.S., perhaps enhancing efficiencies in ways that would yield savings to states and taxpayers.

It is clear that the market for assessments is changing rapidly with the shift toward the Common Core assessments that most states have committed to adopting. But at present the assessments being developed by PARCC and SBAC only cover math and reading, leaving the balance of assessment selection decisions to individual states. As implementation of the Common Core standards proceeds, states should conduct a parallel experiment with smaller-scale collaboration on assessments (and perhaps standards as well) in other subject areas. Only this sort of experimentation will produce the kind of evidence that is needed for states to design and implement high-quality assessment systems at a cost they can afford.

Appendix A: Tables

Table A1. Previous Estimates of Spending on State Assessment Systems

State	Per Pupil			Per Test
	2001 (Hoxby)	2001 (Pew)	2010 (SBAC)	2010 (PARCC)
Alabama	-	\$7	\$29	\$13
Alaska	-	\$33	-	-
Arizona	\$11	\$7	-	\$8
Arkansas	-	\$10	-	\$24
California	\$26	\$9	-	-
Colorado	\$21	\$20	\$36	\$16
Connecticut	\$21	\$5	\$51	-
Delaware	\$44	\$44	\$67	\$5
Florida	-	\$12	-	\$14
Georgia	\$6	\$13	\$17	\$6
Hawaii	-	\$10	\$123	-
Idaho	\$21	\$4	\$35	-
Illinois	-	\$11	-	\$17
Indiana	\$32	\$25	-	\$16
Iowa	-	\$0	\$24	-
Kansas	-	\$3	\$11	-
Kentucky	\$24	\$17	\$25	\$15
Louisiana	-	\$17	-	\$13
Maine	-	\$20	\$73	-
Maryland	\$32	\$26	-	\$32
Massachusetts	\$27	\$27	-	\$25
Michigan	\$9	\$12	\$22	-
Minnesota	\$17	\$8	-	-
Mississippi	-	\$20	-	-
Missouri	\$20	\$20	\$12	-
Montana	-	\$2	\$37	-
Nebraska	-	\$7	-	-
Nevada	-	\$13	\$30	-
New Hampshire	\$13	\$16	\$31	-
New Jersey	\$17	\$17	\$45	\$19
New Mexico	-	\$3	\$77	\$6

Table A1 continued

State	Per Pupil		Per Test	
	2001 (Hoxby)	2001 (Pew)	2010 (SBAC)	2010 (PARCC)
New York	\$6	\$6	-	-
North Carolina	-	\$11	\$9	-
North Dakota	-	\$2	\$72	-
Ohio	\$11	\$9	\$52	\$23
Oklahoma	-	\$5	\$40	\$13
Oregon	-	\$17	\$16	-
Pennsylvania	\$11	\$11	\$31	\$14
Rhode Island	-	\$19	-	\$16
South Carolina	\$2	\$16	\$42	\$12
South Dakota	-	\$7	\$97	-
Tennessee	-	\$22	-	\$12
Texas	\$27	\$9	-	-
Utah	-	\$4	\$7	-
Vermont	-	\$6	\$36	-
Virginia	\$22	\$21	-	-
Washington	\$19	\$10	\$46	-
West Virginia	\$17	\$2	\$60	-
Wisconsin	\$8	\$3	\$25	-
Wyoming	-	\$24	-	-

Notes: Data for the District of Columbia are not included in any of these data sources. Amounts are in 2012 dollars. Hoxby numbers include all accountability costs (not just assessment). Pew numbers are based on telephone survey to states (which include "only the cost of developing, administering and correcting the state test") and total pre-K-12 enrollments provided to Pew by Education Week. SBAC numbers includes only mathematics and ELA assessment costs, and only students in grades 3-8 and 11 were counted in cost per student calculations. PARCC numbers reflect cost per student per test.

Sources: Caroline M. Hoxby, "The Cost of Accountability," in *School Accountability*, Williamson M. Evers & Herbert J. Walberg, eds. Stanford, CA: Hoover Institution Press (2002); Tiffany Danitz. Special Report: States Pay \$400 Million For Tests in 2001. *Stateline*, Pew Center on the States. February 27, 2001. Online at <http://www.pewstates.org/projects/stateline/headlines/special-report-states-pay-400-million-for-tests-in-2001-85899393054>; SBAC Race to the Top Application (appendices, p. 189); and PARCC.

Table A2. Main Assessment Contract Costs

State	Years Included	Yearly Amount	Enrollment, Grades 3-9	Cost per		Primary Contractor	Additional test types
				Student	Student		
Alabama	2008-2011	\$6,317,838	414,099	\$15	Pearson	0	
Alaska	2010-2012	\$5,447,060	68,905	\$79	Data Recognition Corp	2	
Arizona	2010-2012	\$12,888,858	576,351	\$22	Pearson	2	
Arkansas	2011-2012	\$9,838,200	257,066	\$38	Questar	2	
California	2010-2012	\$53,566,116	3,312,499	\$16	ETS, subcontract to Pearson	4	
Colorado	2008-2012	\$17,236,354	428,854	\$40	McGraw Hill	1	
Delaware	2012-2012	\$5,124,262	69,975	\$73	American Institutes for Research	3	
District of Columbia	2010-2011	\$3,855,778	33,759	\$114	McGraw Hill	2	
Florida	2010-2012	\$58,340,052	1,426,357	\$41	Pearson	3	
Georgia	2008-2011	\$12,792,058	904,341	\$14	McGraw Hill	2	
Hawaii	2010-2012	\$10,109,334	95,975	\$105	American Institutes for Research	2	
Idaho	2011-2012	\$5,683,999	149,799	\$38	Data Recognition Corp	0	
Illinois	2007-2011	\$19,011,148	1,110,683	\$17	Pearson	1	
Indiana	2008-2008	\$30,350,390	562,660	\$54	McGraw Hill	0	
Kansas	2010-2011	\$2,451,278	248,770	\$10	University of Kansas	2	
Kentucky	2012-2012	\$8,458,018	354,262	\$24	Pearson	3	
Louisiana	2008-2012	\$5,590,463	372,113	\$15	Data Recognition Corp	2	
Maine	2010-2010	\$2,490,420	99,595	\$25	Measured Progress	1	
Maryland	2007-2011	\$12,435,754	446,598	\$28	Pearson	0	
Massachusetts	2007-2012	\$32,469,904	509,312	\$64	Measured Progress	0	
Michigan	2012-2012	\$19,210,832	832,600	\$23	Measurement Inc.	2	
Minnesota	2008-2012	\$22,637,414	430,650	\$53	Pearson	3	
Mississippi	2007-2012	\$3,836,879	267,086	\$14	Pearson	0	
Missouri	2008-2012	\$7,639,387	482,396	\$16	McGraw Hill	0	
Montana	2010-2012	\$3,028,362	75,718	\$40	Measured Progress	1	
Nebraska	2011-2012	\$4,702,177	151,268	\$31	Data Recognition Corp	1	
Nevada	2008-2012	\$7,475,247	239,349	\$31	Measured Progress	2	

Table A2 continued

State	Years		Yearly Amount	Enrollment, Grades 3-9		Student Cost per	Primary Contractor	Additional test types
	Included	Amount		Grades 3-9	Student			
New Hampshire	2010-2012	\$3,168,565	105,719	\$30	Measured Progress	2		
New Jersey	2008-2012	\$15,143,734	705,393	\$21	Measurement Inc.	1		
New Mexico	2010-2012	\$3,614,594	180,485	\$20	Measured Progress	3		
New York	2007-2011	\$10,324,483	1,422,768	\$7	McGraw Hill	0		
North Carolina	2010-2010	\$8,969,794	814,456	\$11	North Carolina State University	4		
North Dakota	2011-2012	\$3,424,688	49,967	\$69	McGraw Hill	1		
Ohio	2011-2012	\$40,208,376	951,932	\$42	American Institutes for Research	4		
Oregon	2008-2012	\$3,829,838	302,588	\$13	American Institutes for Research	3		
Pennsylvania	2009-2012	\$32,224,978	952,014	\$34	Data Recognition Corp	3		
Rhode Island	2007-2012	\$2,917,997	78,774	\$37	Measured Progress	1		
South Dakota	2009-2012	\$2,979,006	65,846	\$45	Pearson	2		
Tennessee	2009-2012	\$14,238,129	519,822	\$27	Pearson	3		
Texas	2007-2010	\$85,059,944	2,482,169	\$34	Pearson	1		
Utah	2010-2012	\$4,274,512	308,614	\$14	Measured Progress	0		
Vermont	2007-2012	\$1,798,897	46,926	\$38	Measured Progress	0		
Virginia	2007-2011	\$11,602,747	657,369	\$18	Pearson	3		
Washington	2009-2012	\$33,093,724	553,756	\$60	Data Recognition Corp	2		
Wisconsin	2008-2012	\$9,358,807	436,290	\$21	McGraw Hill	1		
Total (45 states)		\$669,220,393	24,555,926	\$27				

Notes: All amounts are in 2012 dollars.

Sources: Author's calculations from state assessment contract data (spending and test data) and the NCES Common Core of Data (enrollment in grades 3-9).

Table A3. Contractor Market Shares

Contractor	States	Students	Student Share	Total Yearly Contracts	Dollar Share	Per-Pupil Cost
American Institutes for Research	4	1,420,470	6%	\$59,271,810	9%	\$42
Data Recognition Corp	6	2,247,855	9%	\$86,742,401	13%	\$39
ETS (subcontract to Pearson)	1	3,312,499	13%	\$53,566,116	8%	\$16
McGraw Hill	8	4,321,035	18%	\$94,981,945	14%	\$22
Measured Progress	9	1,644,490	7%	\$61,238,497	9%	\$37
Measurement Inc.	2	1,537,993	6%	\$34,354,566	5%	\$22
North Carolina State University	1	814,456	3%	\$8,969,794	1%	\$11
Pearson	12	8,751,292	36%	\$257,805,786	39%	\$29
Questar	1	257,066	1%	\$9,838,200	1%	\$38
University of Kansas	1	248,770	1%	\$2,451,278	0%	\$10
Total	45	24,555,926		\$669,220,393		\$27

Notes: Student counts include all students in grades 3-9. Amounts are in 2012 dollars. Each state's primary assessment contract(s) are associated with the primary contractor for the state, so in the small number of states that have multiple contracts for the primary assessment the costs of the secondary contract will be associated with the primary contractor.

Source: Author's calculations from state assessment contract data.

Table A4. Number of Test Items, by Subject, Grade, Type, and State

State	Reading				Math			
	Grade 4		Grade 8		Grade 4		Grade 8	
	MC	Non-MC	MC	Non-MC	MC	Non-MC	MC	Non-MC
Alabama	60	4	54	4	60	4	55	5
Alaska	52	3	52	3	56	3	56	3
Arizona	54	0	54	0	85	0	85	0
Arkansas	117	4	136	4	60	6	64	6
California	65	0	75	0	65	0	65	0
Colorado	56	14	56	14	54	15	45	15
Connecticut	24	8	20	10	80	16	81	36
Delaware	48	2	48	2	50	0	50	0
District of Columbia	45	3	45	3	51	3	51	3
Florida	45-50	5-7	45-50	5-7	45-50	0	40-50	5-8
Georgia	90	0	90	0	60	0	60	0
Hawaii	37-38	2-3	37-38	2-3	40-41	4-5	38-39	4-5
Idaho	108	0	119	0	55	0	60	0
Illinois	50	1	50	1	65	5	65	5
Indiana	45	5	45	5	46	4	46	4
Iowa	185	0	229	0	82	0	107	0
Kansas	58	0	84	0	73	0	86	0
Kentucky	30	5	30	5	30	7	30	7
Louisiana	33	10	33	11	60	3	60	4
Maine	42	9	42	9	41	24	38	21
Massachusetts	36	4	36	4	32	10	32	12
Michigan	30	1	30	1	40	13	40	11
Minnesota	46	0	57	0	42	0	42	0
Mississippi	50	0	70	0	45	0	50	0
Montana	52	2	52	2	55	5	55	5
Nebraska	45	0	50	0	55	0	60	0
Nevada	57	2	57	2	57	2	57	2
New Hampshire	42	9	42	9	41	24	38	21
New Jersey	24	3	36	4	35	9	32	11
New Mexico	33	8	33	8	47	12	53	14
New York	51	8	49	8	45	12	42	12
North Carolina	56	0	56	0	50	0	60	0
North Dakota	52	3	52	3	52	2	52	2
Ohio	29	7-8	32	6	32	8	32	6
Oklahoma	50	0	50	0	50	0	50	0

Table A4 continued

State	Reading				Math			
	Grade 4		Grade 8		Grade 4		Grade 8	
	MC	Non-MC	MC	Non-MC	MC	Non-MC	MC	Non-MC
Oregon	45	0	50	0	40	0	40	0
Pennsylvania	40	4	40	4	60	3	60	3
Rhode Island	42	9	42	9	41	24	38	21
South Carolina	36	0	50	0	56	0	63	0
South Dakota	56	0	56	0	84	0	84	0
Tennessee	75	0	83	0	69	0	69	0
Texas	40	0	48	0	41	1	49	1
Vermont	42	9	42	9	41	24	38	21
Virginia	40	7	47	8	60	0	60	0
Washington	26	4	30	5	20	10	25	10
West Virginia	44	0	44	0	45	0	45	0
Wisconsin	54	2	54	2	46	4	51	4
Wyoming	44	6	42	14	50	5	60	5
Average	52.2	3.3	56.3	3.7	52.3	5.5	53.9	5.9

Notes: "MC" refers to the number of multiple-choice items and "Non-MC" refers to the number of non-multiple-choice items. Math multiple choice items included gridded items. Field test items are excluded whenever possible.

Source: State education agencies (websites and direct inquiries).

Table A5. All Assessment Contracts, Selected States

State and Test	Amount	Share of total
Alabama, 2011		
ARMT and Stanford 10	\$5,579,058	56%
ACCESS for ELLs	\$524,459	5%
HS Graduation Exam	\$2,031,243	20%
Alternate Assessment	\$942,394	9%
Science Assessment	\$437,899	4%
ACT - EXPLORE	\$514,175	5%
Maryland, 2008-2011		
Math/ELA	\$12,439,879	37%
Science Assessment	\$4,984,366	15%
High School and Alternate HS Assessments	\$16,401,619	48%
Michigan, 2012		
MEAP (reading, writing, math, science, social studies)	\$19,210,832	49%
Michigan Merit Examination	\$14,370,770	37%
English Language Proficiency Assessment	\$2,042,679	5%
Alternate Assessment Program	\$3,617,488	9%
New Hampshire, 2010-2011		
NECAP	\$3,203,120	62%
Science	\$804,668	16%
Alternative Learning Progressions	\$1,135,680	22%
Utah, 2010-2011		
ELA/Math and Computer Based Testing	\$4,087,646	51%
Direct Writing Assessment	\$753,615	9%
Science	\$1,518,832	19%
ACT	\$914,436	11%
Academic Language Proficiency Assessment	\$777,417	10%

Notes: Amounts are in 2012 dollars

Source: Author's calculations from state assessment contract data.

Table A6. Assessment Contract Costs by Activity

State and Activity	Share
Missouri, 2008-2012	
Develop	28%
Administer	72%
New Jersey, 2008-2012	
Development	18%
Meetings	5%
Print/Produce/Deliver	22%
Score and Report	53%
Training	1%
Translate	1%
New Mexico, 2010-2012	
Item Development	18%
Meetings/workshops/studies	9%
PD	1%
Materials/scoring (billed to districts)	72%
New York, 2004-2009	
Development	37%
Research (mainly field test)	17%
Admin, printing/dissemination	33%
Scoring (training and sample)	12%

Source: Author's calculations from state assessment contract data.

Appendix B: Modeling Assessment Costs

Our descriptive model of assessment costs relates per-student contract costs (measured as described above) to various features of the state assessment system using ordinary least squares (OLS) regression. The factors examined, and their expected relationships with per-student assessment costs, are:

- 1) Number of students in grades 3-9; we expect larger states to have lower per-student costs by virtue of spreading fixed costs (e.g., development costs) over a larger number of students and by having more negotiating power with testing vendors.
- 2) Additional tests included in the main assessment contract (science, social studies, writing, high school, and alternate); we expect more tests included to increase per-student costs.
- 3) Average teacher salary in the state, which we expect to be positively correlated with assessment costs because it is a proxy for labor costs and the overall cost of living in the state (although the expected magnitude of the relationship is unclear given that much of the contracted work is often performed in a different state).
- 4) The average number of items on the fourth- and eight-grade reading and math tests; we expect additional items to result in increased costs because of item development costs.
- 5) The share of items that are multiple choice, which we expect to be negatively correlated with per-student costs because multiple choice items can be scored automatically, which is expected to be less expensive than scoring non-multiple-choice items (which is often done by trained scorers).

The results of these models, along with heteroskedasticity-robust standard errors, are reported in Table B1. We initially specified a quadratic function between enrollment and per-pupil costs to allow for the possibility of a non-linear relationship, specifically the possibility that per-pupil costs would decline as state sized increased but would eventually level out. This theory is initially supported by a statistically significant coefficient on enrollment squared (column [1] of Table B1). However, these estimates imply that per-pupil costs decline as statewide enrollment (in grades 3-9) increases to about two million, but then increase above that. The enrollment and cost data are plotted in Figure B1, and show that only two states enroll more than 1.5 million students in these grades (California and Texas). Below this threshold, the estimated relationship is approximately linear.

Consequently, we re-estimate the enrollment-cost relationship dropping the two states that are outliers in terms of enrollment, as well as two small jurisdictions (the District of Columbia and Hawaii) that are outliers in terms of per-pupil contract costs (both above \$100). In the sample of states that excludes these four outliers, the coefficient on enrollment squared is no longer statistically

significant from zero, so we exclude it in all future model (column [2] of Table B1). The estimated linear relationship, which is reported in column (3) and displayed in Figure B2, indicates that per-pupil assessment costs are predicted to fall by \$0.15 for every 10,000-student increase in enrollment.

Column (4) adds dummy variables indicating which additional tests are included in the main assessment contract. The coefficients on all of these dummies are imprecisely estimated and not statistically significant from zero, and we also cannot reject the null hypothesis that this set of dummies is jointly equal to zero. In light of this finding and in order to preserve degrees of freedom given the limited sample size, in the rest of the models we replace the set of dummies with a single variable that is the sum of the dummies (i.e. the number of additional assessment types included). Column (5) indicates that the coefficient on this variable has the expected positive sign, but also has a large standard error and is not statistically distinguishable from zero. Neither form of controlling for the additional tests included has more than a trivial impact on the enrollment coefficient.

Columns (6) through (8) progressively add to the model the average teacher salary, the average number of items on the fourth- and eighth-grade reading and math tests, and the share of items on these tests that are multiple-choice. In these models, the coefficient on the number of additional test types remains positive but imprecisely estimated, with point estimates in the range of \$0.70 to \$2.57 per additional test type included. Average teacher salary is positively correlated with per-student costs, but those estimates are also not precise enough to be statistically significant from zero. The coefficient on the average number of items has the opposite sign from what theory would predict, but once again we can not confidently rule out a null relationship. Finally, the coefficient on the share of items that are multiple-choice is very imprecisely estimated.

In addition to modeling per-student contract costs, we also explored models of total contract costs (results not shown) in order to test the theory that variables that are more closely associated with fixed costs (as opposed to variable costs), such as the number and type of items, might be more strongly associated with total costs than with per-student costs. As expected, the number of students was, by far, the strongest predictor of these costs. However, we did not find any consistent evidence that the number or type of items was significantly related to total costs.

As discussed in the text, it is important to bear in mind that a statistically insignificant relationship that is imprecisely estimated does not mean that the true correlation is zero but rather that the available data are not particularly informative as to the direction and magnitude of the relationship. In our data, the imprecision in the coefficient estimates likely results from the small sample size and from measurement error in contract costs (due to the difficulty of measuring these costs consistently across states).

Table B1. Predictors of Per-Pupil Assessment Contract Costs

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Grade 3-9 Enrollment (10,000s of students)	-0.373 (0.149)*	-0.456 (0.253)+	-0.154 (0.077)+	-0.156 (0.075)*	-0.164 (0.077)*	-0.232 (0.117)+	-0.220 (0.123)+	-0.226 (0.131)+
Enrollment squared	0.001 (0.000)*	0.002 (0.002)						
Science included				5.889 (6.690)				
Social studies included				-6.057 (7.313)				
Writing included				2.197 (5.795)				
High school tests included				3.609 (11.345)				
Alternate assessments included				-0.586 (7.982)				
Number of additional types included (0-5)					1.629 (2.455)	2.565 (2.413)	0.941 (2.946)	0.696 (2.966)
Average teacher salary (1,000s of dollars)						0.680 (0.491)	0.475 (0.620)	0.514 (0.654)
Average number of items							-0.207 (0.228)	-0.246 (0.253)
Share items that are multiple choice								15.455 (34.314)
Constant	48.993 (7.464)**	44.888 (6.732)**	39.026 (4.844)**	35.049 (5.947)**	36.876 (5.348)**	1.648 (24.461)	27.478 (40.897)	14.320 (54.677)
Exclude Outliers?	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	45	41	41	41	41	41	38	38
R-squared	0.160	0.133	0.092	0.136	0.104	0.172	0.177	0.182

Notes: ** p<0.01, * p<0.05, + p<0.1; robust standard errors in parentheses. "Exclude outliers" indicates that the two jurisdictions with per-pupil spending of more than \$100 (DC and Hawaii) and the two states with enrollment of more than 1.5 million (California and Texas) are excluded.

Source: Author's calculations from state assessment contract data, state test item data, and NCES Common Core of Data (enrollment).

Endnotes

¹ Tiffany Danitz, "Special Report: States Pay \$400 Million for Tests in 2001," *Stateline*, Pew Center on the States, February 27, 2001. The true number may be less than 13 because the Pew data indicate only the grades that were tested, not the subjects (NCLB requires testing in both math and reading in grades 3-8 and once in high school). Additionally, NCLB required schools to test students in science in at least three grades beginning in 2007-08, whereas the math and reading test requirement went into effect beginning in 2005-06 (U.S. Department of Education, "Testing: Frequently Asked Questions," <http://www2.ed.gov/nclb/accountability/ayp/testing-faq.html>).

² Common Core State Standards Initiative, "About the Standards," <http://www.corestandards.org/about-the-standards>.

³ Common Core State Standards Initiative, "In the States," <http://www.corestandards.org/in-the-states>.

⁴ In August 2012, Utah withdrew from the consortium it joined (Lisa Schencker, "Utah Drops out of Consortium Developing Common Core Tests," *Salt Lake Tribune*, August 4, 2012).

⁵ See, e.g., AccountabilityWorks, "National Cost of Aligning States and Localities to the Common Core Standards," Boston, MA: Pioneer Institute, February 2012; and Patrick Murphy and Elliot Regenstein, "Putting a Price Tag on the Common Core: How Much Will Smart Implementation Cost?" Washington, DC: Thomas B. Fordham Institute, May 2012.

⁶ Caroline M. Hoxby, "The Cost of Accountability," in *School Accountability*, Williamson M. Evers & Herbert J. Walberg, eds., Stanford, CA: Hoover Institution Press, 2002.

⁷ Tiffany Danitz, "Special Report: States Pay \$400 Million for Tests in 2001," *Stateline*, Pew Center on the States, February 27, 2001.

⁸ The SBAC data are reported on p. X-189 of the appendices to the consortium's Race to the Top application (available at <http://www.smarterbalanced.org/wordpress/wp-content/uploads/2011/12/Smarter-Balanced-RttT-Application-Appendices.pdf>). The PARCC data were provided to the author by PARCC.

⁹ The SBAC data are reported as per student in grades 3-8 and 11 (because NCLB requires testing in grades 3-8 and once in high school), whereas the Pew and Hoxby data are reported as per student in the state.

¹⁰ In a recent report written for the National Research Council, Douglas Harris and Lori Taylor estimate that the costs of standards, assessments, and accountability systems totaled (as of 2007) about \$167 per pupil, which is about 1.8 percent of per-pupil spending and totals about \$8.1 billion per year (in 2012 dollars). However, these estimates are based on data collected from only three states and reflect the costs of standards and accountability systems in addition to the assessment costs (Douglas N. Harris and Lori L. Taylor, "The Resource Costs of Standards, Assessments, and Accountability," Final Report to the National Research Council, March 2008).

¹¹ Spending data are converted to 2012 dollars and are from Table 186 of: Digest of Education Statistics: 2011, National Center for Education Statistics, U.S. Department of Education, May 2012.

¹² Georgia Department of Education, “Criterion-Referenced Competency Tests (CRCT)”, <http://www.doe.k12.ga.us/Curriculum-Instruction-and-Assessment/Assessment/Pages/CRCT.aspx>.

¹³ Stephanie Banchero, “School-Test Backlash Grows,” *Wall Street Journal*, May 16, 2012.

¹⁴ *Education Next*-PEPG Survey 2011 Survey Results, available at http://educationnext.org/files/EN-PEPG_Complete_Polling_Results_2011.pdf.

¹⁵ Smarter Balanced Assessment Consortium, “Frequently Asked Questions,” <http://www.smarterbalanced.org/resources-events/faqs/>.

¹⁶ Time spent preparing for end-of-year tests may also be considered a “cost,” but it is one that is nearly impossible to measure given the difficulty of separating instructional time that is geared specifically towards preparation for the test as compared to for some other purpose (presumably all instruction aimed at student learning in a given subject area is at least tangentially related to performance on a test of knowledge in that subject area).

¹⁷ Enrollment data are from the NCES Common Core of Data. The most recent data are from 2010-11; because state-level enrollments do not change much over short periods of time, we also use the 2010-11 data for 2011-12.

¹⁸ The bid amounts are not adjusted for inflation. Including the two-year contract renewal period increases the difference between the two bids to more than \$300 million (Florida Department of Education, “FCAT Contract Resources,” <http://www.fldoe.org/fcat.asp>). The Pearson contract generated controversy when the release of statement assessment results was significantly delayed in 2010 (Cara Fitzpatrick, “FCAT Errors Don’t Faze State: Florida Plans to Give More Work to a Firm with a List of Blunders,” *Palm Beach Post*, June 26, 2010).

¹⁹ These data were usually available on state education agency websites, but in some cases were obtained by contacting the agency’s office.

²⁰ We counted as multiple-choice items those that were not technically multiple choice but could easily be graded by machine, such as “gridded” items on math tests.

²¹ Field test items are excluded from the counts whenever possible.

²² This issue has been discussed in earlier studies of assessment costs dating back at least to a 1993 GAO report (*Student Testing: Current Extent and Expenditures, With Cost Estimates for a National Examination*, U.S. General Accounting Office, January 1993). See also Lawrence O. Picus, Frank Adamson, William Montague, and Margaret Owens, “A New Conceptual Framework for Analyzing the Costs of Performance Assessment,” Stanford, CA: Stanford University, Stanford Center for Opportunity Policy in Education, 2010.

²³ NCLB also requires states to administer tests of English proficiency to all students who are limited English proficient (U.S. Department of Education, “Testing: Frequently Asked Questions,” <http://www2.ed.gov/nclb/accountability/ayp/testing-faq.html>).

²⁴ Examining data from states that provided multiple contracts likely causes us to overstate the importance of assessments other than those covered by the main contract because, for the purpose of this case study, we did not examine states that only had one contract (the main one, which would account for 100 percent of all assessment contract costs). In these states, it was not clear whether additional contracts did not exist or did exist but were not provided to us by the state.

²⁵ Interview with Alabama State Department of Education staff, November 2012.

²⁶ Because we only have one year of contract data on ARMT+ administration, which also coincided with the third year of test development, in our national analysis we use budget data from the previous test (ARMT) only.

²⁷ The New York contract amounts are not adjusted for inflation because they are specified for future years.

²⁸ This adjustment is made by predicting per-pupil costs in states for which we did not obtain data, using the relationship between costs and enrollment in grades 3-9 (estimated using the states for which we do have data). These states are smaller, on average, than the states for which we do have data, so it is unsurprising that their predicted per-pupil assessment contract costs are modestly higher (\$34, as compared to \$27).

²⁹ Specifically, we assume that contract costs for the main assessment constitute 50 percent of total contracted costs (based on the case studies discussed above), and that total contracted costs account for 85 percent of total state spending on assessments (based on the PARCC data discussed above).

³⁰ Total per-pupil spending (used to calculate assessment spending as a percentage of the total) is from the 2008-09 Common Core of Data. Student enrollment counts are from fall 2010, as reported in the 2012 Digest of Education Statistics.

³¹ These calculations are based on student enrollment, teacher employment, and teacher salary data from 2010-11 reported in the 2012 Digest of Education Statistics. The simulated change in the pupil-teacher ratio is estimated by calculating how many additional teachers could be hired using \$1.7 billion dollar at the average teacher salary of \$57,253, assuming that teachers receive benefits equal to 40 percent of their salary. The simulated change in average teacher salary is simply \$1.7 billion divided by the number of teachers in the country.

³² See, e.g., Sarah D. Sparks, "Today's Tests Seen as Bar to Better Assessment," *Education Week*, November 13, 2012.

³³ The existence of assessment consortia could also produce cost savings by enabling states to switch assessments without incurring the substantial costs of hiring a contractor to develop an entirely new assessment for the state.

³⁴ A recent report that estimated the cost savings from states forming assessment consortia (as well as the cost changes associated without other assessment characteristics) is Barry Topol, John Olson, and Ed Roeber, "The Cost of New Higher Quality Assessments: A Comprehensive Analysis of the Potential Costs for Future State Assessments," Stanford, CA: Stanford University, Stanford Center for Opportunity Policy in Education, 2010. But this report, which is based on a proprietary cost model and dataset, does not provide sufficient information to allow readers to assess the validity of its cost model, such as measures of how well actual data on assessment costs fit the predictions of the model.

³⁵ The number of states is taken from the lists published on the PARCC and SBAC websites, as of November 2012. The number of students in grades 3-9 in these states is from the 2010-11 Common Core of Data.

³⁶ Kathryn A. McDermott, "Interstate Governance of Standards and Testing," in *Education Governance for the Twenty-First Century*, Paul Manna and Patrick McGuinn, eds., Washington, DC: Brookings Institution Press, 2013.

³⁷ As of April 2012, 32 states had joined one of two consortia developing alternate assessments and 29 states were members of a consortium developing an English language proficiency assessment. These three consortia were awarded start-up grants by the U.S. Department of Education (“Coming Together to Raise Achievement: New Assessments for the Common Core State Standards,” Center for K-12 Assessment & Performance Management at ETS, Education Testing Service, April 2012).

³⁸ Interview with Scott Marion, Associate Director, Center for Assessment, November 2012.

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>

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

Ashley Inman

Email your comments to browncenter@brookings.edu

The Brookings Institution is a private non-profit organization. Its mission is to conduct high-quality, independent research and, based on that research, to provide innovative, practical recommendations for policymakers and the public. The conclusions and recommendations of any Brookings publication are solely those of its author(s), and do not reflect the views of the Institution, its management, or its other scholars.

Support for this publication was generously provided by the 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.