Testing Overload in America’s Schools

By Melissa Lazarín  October 2014
Background on testing

Federally driven testing

Although tests have long played a role in K-12 schools, the No Child Left Behind Act, or NCLB, elevated their role by requiring states to annually assess students in reading and math in grades 3-8 and once in grades 10-12. It also requires states to test students in science once in grades 3-5, 6-8, and 10-12. In addition, NCLB requires states to annually assess English language proficiency among students who are learning English.

To meet the federal law’s requirements, 23 states expanded their assessment programs to test students in grades they had not been previously assessing. In the shift to the Common Core standards, two groups of states—known as the Smarter Balanced Assessment Consortium, or Smarter Balanced for short, and the Partnership for Assessment of Readiness for College and Careers, or PARCC—have joined together to develop Common Core-aligned assessments that can be shared across multiple states. These tests are expected to take the place of the assessments that states have instituted to comply with NCLB, though a few states are developing their own tests or are considering alternative assessments such as ACT’s Aspire Common Core test.

A number of states successfully piloted the Common Core tests last school year, 2013-14, and will implement them in earnest this academic year. Computer-based, the tests are expected to be of higher quality compared with current state assessments, featuring open-response test items and problems that will encourage and test higher-order thinking skills. Because these tests will be shared across multiple states, they are expected to support greater transparency of student achievement across states.
State-driven testing

While NCLB triggered state development of statewide assessments more broadly, it is clear that states have also opted to supplement the federally required tests with other assessments. High school exit and end-of-course exams, statewide diagnostic policies, and efforts to boost college-going participation make up many of these supplemental exams.

Twenty-five states administered high school exit exams in the 2011-12 school year. Some states are also shifting toward the use of end-of-course exams, or EOC tests. Fifteen states required the class of 2012 to take EOC exams, with nine of these states requiring successful passage of the exams as a prerequisite for graduation.

In addition to summative assessments such as these, some states require districts to administer assessments to identify students for programs and services. Ohio, for example, requires its districts to administer a state-approved assessment to screen students for gifted-education programming, and many districts comply by administering a screening test to all students in several grades.

Partially due to federal initiatives such as the Race to the Top – Early Learning Challenge grant, more states are also administering kindergarten-readiness assessments. In 2012, half of all states required an assessment in kindergarten.

A handful of states also require all students in a grade cohort to take college-entrance or preparation exams, such as the PSAT, SAT, and ACT. Twelve states, for example, foot the bill and require their high school juniors to take the ACT. Such policies are often instituted for lofty reasons, including fostering a college-going culture and setting high expectations for all students.

District-driven testing

Districts, too, have added additional assessments over the past decade. During the 2011-12 academic year, education-technology companies generated $2.2 billion in revenue from district investments in digital testing and assessments. Many districts increased their use of benchmark assessments, or interim assessments, which are largely intended to help teachers improve their instructional practice. Some districts also use these tests to regularly track student academic
achievement prior to the state summative exams. In a 2005 Education Week survey, 7 out of 10 superintendents reported administering district-wide tests on a regular basis, and that number was expected to grow by 10 percent the following year. Even during the economic recession, as states faced tight budget constraints, many continued to invest in interim tests.  

Some argue that the Common Core standards are driving the current growth in the testing market. As part of this report, we spoke with several district officials about why they administer district-level assessments. Most of the districts we spoke with reported that they would likely be moving in this direction anyway. But some of the shortcomings of states’ old tests have also encouraged them to rely on other measures.

At least two districts remarked that while the old state exams serve an important purpose, they are taken at the end of the school year, and school districts do not receive the results right away. “We are still waiting for our results from the spring,” said Marco Muñoz, an evaluation specialist with Jefferson County Public Schools in Kentucky, when interviewed in mid-September. Peter Weber, chief of data and strategy for the District of Columbia Public Schools, agreed. The interim tests that many districts administer across their schools provide timely information, which enables them to take action before the end of the year. “It’s useful to have more real-time information so that you can know on a given day how a student is doing,” said Weber.

The benchmark proficiency exams also give district officials insight into how their students will perform on the state’s end-of-the-year tests. “I need to know a little about the ongoing progress of my players before they go into the final game of the season,” continued Muñoz. Kentucky’s Jefferson County has four cycles of proficiency assessments throughout the year.

The poor quality of the old state tests also compelled some districts to use additional assessments. Muñoz finds that his district’s benchmark tests are of better quality than some of the current state assessments on the books because the district’s assessments include a greater number of open-response test items, encouraging students to demonstrate their writing skills. However, he asserted that the new Common Core-aligned assessments will be more performance- and project-based, starting with science and following with other subject areas.
Finally, some districts use additional assessments to assess different skills than the state tests. As Weber described:

*The CAS [the District of Columbia’s Comprehensive Assessment System] is pretty good about having a student read a passage and telling us whether the student can interact with the text in a meaningful way. I think this is what the PARCC will do as well. They’re good barometers for college readiness. But for some students, we might need more refined information. For example, we use the DIBELS [Dynamic Indicators of Basic Early Literacy Skills] test to help determine whether a student can sound out all the words—phonemic awareness. And we also have a test that evaluates whether they can understand the meaning of the passage.*

Rochester City School District Superintendent Bolgen Vargas agreed. “The state assessments tell you a lot about what the child achieved in a given year, but when you have a school-wide assessment like the NWEA [the Northwest Evaluation Association’s Measures of Academic Progress assessment], you get a sense of where children are at a specific point in time. It tells us what kind of instructional changes we should make.”

There is some evidence that the new Common Core-aligned assessments may alleviate the need for some district-level testing. For example, the PARCC and Smarter Balanced assessments are expected to provide more timely results than current state assessments, which is at least one reason why districts have come to rely heavily on interim assessments.

The new Common Core tests are also widely expected to be of higher quality than current state assessments. They include a greater proportion of open-response questions and problems that will encourage and test higher-order thinking skills. Indeed, many students who piloted the new tests last spring found the tests more rigorous and demanding, with less of an emphasis on “regurgitating facts.” Therefore, some districts—such as Kentucky’s Jefferson County—may find more value in relying on these assessments in the future.

As some districts indicated, interim tests that are taken prior to the end-of-the-year summative tests can be useful because they provide information about their students throughout the year that can be addressed before gaps widen or become more problematic. While districts may continue to rely on interim tests for this reason, the quality and benefits of some of the district interim tests currently in place
are questionable. Research suggests that “interim assessments are useful but not sufficient to inform instructional improvement” and should be linked to districts’ curricula and state standards, professional learning for teachers in how best to analyze and use test data, and a clear sense of purpose for using the test.39

Smarter Balanced and PARCC—the groups of states developing the Common Core-aligned summative tests—are also developing interim tests and midyear assessments for districts to use at their discretion.40 The PARCC and Smarter Balanced summative and interim assessments may offset the need for districts to use some of the interim assessments they currently have in place, given that these tests are aligned to states’ standards and are of highe quality.

The role of teacher evaluation in the use of district-level assessments

In recent years, efforts to reform the manner in which educators are evaluated have gained traction. Spurred by federal initiatives such as Race to the Top and flexibility from certain NCLB requirements, or Elementary and Secondary Education Act waivers, states and districts have adopted policies that require educators to be evaluated in part by data on student academic growth. For grades that are tested in reading, math, and science as required by the Elementary and Secondary Education Act states must ensure that districts’ teacher-evaluation systems include students’ scores on state assessments as a basis for student academic growth.41 States must use alternative measures of student learning for subjects and grades that are not tested under federal law. States and districts have the flexibility to choose these alternative measures, which could include performance on end-of-course exams or other “objective performance-based assessments.”42

In response, some states and districts have incorporated additional tests to help measure teacher performance. New York State Education Department officials believe efforts to reform teacher evaluation have unintentionally led to the proliferation of district assessments in their state.43 The state’s guidelines require 20 percent of a teacher’s evaluation to be based on student growth

### Key test types

**Formative, interim, and summative**

The National Center for the Improvement of Educational Assessment offers a useful framework for characterizing the various assessments that schools administer to students.

**Summative assessments:** These tests are administered once at the end of an academic semester or school year. Offered by states and districts, summative tests help determine student performance relative to a set of content standards. Most federally required statewide assessments fall into this category.

**Interim assessments:** Often referred to as benchmark assessments, these tests are generally administered school or district wide. While teachers might use these tests to inform their instruction, “a crucial distinction is that these results can be meaningfully aggregated and reported at a broader level”—at the school or district level. Education leaders use these tests for a variety of reasons, including predicting student performance on summative assessments, evaluating a program’s effectiveness, or diagnosing student needs and learning gaps.

**Formative assessments:** Administered by teachers, formative tests are embedded in classroom instruction and can be as short as five seconds. An example of a formative assessment would be to ask students to explain the main point of a lesson in one to two sentences. These results are primarily useful to teachers, who use them to evaluate students’ grasp of the lesson at hand and modify instruction if necessary.

on the state test, 20 percent on student growth on “local assessments or a set of predetermined learning objectives,” and 60 percent on class observations. Many districts, as a result, have instituted pre- and post-tests for the sole purpose of calculating a local measure of student academic growth.

But implementing more robust teacher-evaluation systems does not necessitate additional testing. The following section highlights how state and district officials in New York have modified their response to the call for more meaningful evaluation systems with minimal additional testing.
Findings

CAP examined the available data on standardized tests by the following grade spans: K-2, 3-5, 6-8, and 9-12. Below are our major findings based on available information for the 2013-14 school year.

Despite the perception that federally mandated state testing is the root of the issue, districts require more tests than states. State tests alone are not to blame for testing fatigue. District-level tests play a role too. Students across all grade spans take more district-required exams than state tests. Students in K-2 are tested three times as much on district exams as state exams, and high school students are tested twice as much on district exams. But even students in grades that must be assessed per No Child Left Behind took between 1.6 and 1.7 times more district-level exams than state exams.

Most of the district-level tests in use were interim benchmark exams that are taken two to four times throughout the year. Other district-wide exams included diagnostic tests and end-of-course exams for students taking certain required courses.

Students are tested as frequently as twice per month and an average of once per month. Testing can occur very frequently for some students. Students in grades in which federal law requires annual testing—grades 3-8—take the most tests. This means about 10 tests, on average, throughout the year. But in the Jefferson County school district in Kentucky, which includes Louisville, students in grades 6-8 were tested approximately 20 times throughout the year. Sixteen of these tests were district-level assessments. In the Sarasota County, Florida, school district, middle school students were tested 14 times on state and district tests throughout the year. These interruptions in instruction may likely be contributing to public sentiment regarding students being overtested.

Students in grades K-2 and 9-12, who do not take or are less frequently tested using federally required state exams, take the fewest number of tests—approximately six tests in a year.
Actual test administration takes up a small fraction of learning time. Students spend, on average, 1.6 percent or less of instructional time taking tests. This corresponds to findings from other similar examinations of testing time.81

On average, students in grades 3-5 and 6-8 spend 15 and 16 hours, respectively, on district and state exams. In contrast to the average total hours of instructional time, the amount of time spent on test-taking is comparatively small.82 These students did spend more time on state tests than district tests—nearly three more hours, on average.

Students in grades K-2 and 9-12, who take the fewest number of tests—approximately six tests in a year—spent the least amount of time taking tests in the year at approximately four and nine hours, respectively. The fact that these students do not take or are less frequently tested using federally required state exams is a contributing factor.

There is a culture of testing and test preparation in schools that does not put students first.

Based on our analysis, test-taking time does not appear in itself to be problematic. But the culture of testing, particularly in urban districts, may play a more prominent role in the schooling experience. The frequency at which testing interrupts the school calendar in some districts and the fact that the testing burden seems to disproportionately affect urban schools are important aspects of today’s testing culture that warrant additional consideration. Our research also finds that some districts and states may be administering tests that are duplicative or unnecessary, including the use of practice tests.

More difficult to capture but an integral part of the testing conundrum nonetheless is how much time schools spend on test-preparation activities, training teachers on how to administer assessments, and analysis of test results. District-level staff in the Rochester City School District, for example, acknowledged that the time it took to administer and analyze pre- and post-tests, which they were using for teacher-evaluation purposes, was a motivating reason to consider alternatives strategies and eliminate some district-wide tests.
District-level testing occurs more frequently and takes up more learning time in urban districts than in suburban districts.

We examined how district-level testing compared across urban and suburban districts and found some substantial differences. Urban high school students, in particular, spend more time taking district-level exams than suburban high school students. Urban high school students take three times as many district-level tests and spend up to 266 percent more time taking them compared with suburban high school students.

A few districts, such as the suburban South-Western City School District in Ohio, do not have any required assessments for high school students. At the other end of the spectrum, high school students in the urban district of Jefferson County, Kentucky, are tested approximately, on average, 13 times throughout the year, and high school students in Denver Public Schools spend an average of nearly 17 hours taking district-level tests.

In grades 3-5 and 6-8, urban-district students spend approximately 80 percent and 73 percent more time, respectively, taking district-level exams than their suburban peers. In grades K-2, urban students spend about 52 percent more time on district tests.

Districts are not transparent about testing practices or purposes.

Chicago Public Schools stood out among all the districts we studied for its transparency regarding district assessments. The district publicly posts its assessment calendar on its website, and it identifies which tests are state or district required. In addition to the testing administration dates, the Chicago district’s assessment calendar also includes the estimated time duration for each test, descriptions of the assessments, and which subgroups of students are assessed.

Most districts post their assessment calendars on their websites, but those calendars lack information that might be useful for parents and other stakeholders. For example, we found only one district—Knox County—in addition to Chicago Public Schools that includes information regarding test administration time.

One district—Denver Public Schools—informed us that additional information about the tests it provides is available to parents behind a secure firewall. This could be the case with others.
While parents likely know when their children take a test, we believe that many parents might not fully understand the purpose of all the tests or when the tests are state or district required. Opt-out efforts and opposition to standardized tests have largely focused on statewide assessments. But as is the case with Lee County, Florida, state assessments alone are not to blame in many cases. This type of information can further support parent knowledge and information regarding testing.