



Measuring More of What Matters Most for Future Student Success

Twenty years ago, laptops were the size of suitcases and cell phones the size of bricks. Google did not exist. Driverless cars were the stuff of science fiction books. Rapid changes in technology have transformed how we communicate, access information and build things. Now, robots work alongside employees in manufacturing plants and laboratories, doing automated tasks once reserved for high school graduates.

TODAY'S TESTS CAN HELP PREPARE TOMORROW'S CITIZENS

65% Percentage of all U.S. jobs requiring postsecondary education or training by 2020.¹

36,000,000 Job openings nationwide requiring postsecondary education or training between 2010-2020.²



IN ARIZONA

MOST NEW JOBS BETWEEN 2010-2020 IN OUR STATE REQUIRE MORE THAN A DIPLOMA. HIGH SCHOOL GRADUATES SHOULD BE WELL-PREPARED TO PURSUE OPPORTUNITIES IN OUR FASTEST GROWING JOB SECTORS.³

FAST GROWING JOB SECTORS	NUMBER OF NEW JOBS NEEDING MORE THAN HIGH SCHOOL DIPLOMA
HEALTHCARE SUPPORT	19,000
HEALTHCARE PROFESSIONAL & TECHNICAL	45,000
SOCIAL SCIENCES	5,000
MANAGERIAL AND PROFESSIONAL OFFICE	133,000

As these changes reshape the U.S. economy, they have implications for the skills and knowledge students need to learn to do well in today's (and tomorrow's) workforce. In addition to having stronger foundational skills than ever before in communications, mathematics and technology, well-prepared graduates need to be able to "problem find, problem frame and problem solve," according to experts.⁴

Tests are one important tool for state and education leaders to help schools reset their expectations and ensure students are learning what they need for success. Tests signal what is important, track how schools and students are progressing, and flag where more support or attention may be needed for student success.

States are upgrading their tests to prioritize the skills and knowledge that matter most today. These new tests take more time as students are required to do more writing and thinking. Independent reviews are showing some—especially the PARCC and Smarter Balanced tests—are better than others. Still, too many state tests haven't been seriously reviewed; they have promises for how they are better or measure what states expect but little evidence to back these claims. Why should policymakers care? A test that measures limited skills, relies primarily on multiple choice questions or doesn't match a state's expectations for student learning is a waste of money, time and effort.

State boards of education, state school superintendents and legislators should all play critical leadership roles in understanding what their state tests measure and how these tests themselves measure up. There are new tools and independent organizations to help.

STATES ARE UPGRADING THEIR TESTS TO PRIORITIZE THE SKILLS AND KNOWLEDGE THAT MATTER MOST TODAY:

- Deeper Understanding
- Innovating
- Knowledge Application
- Problem-Solving
- Synthesizing
- Writing

"The most important takeaway for parents, students and teachers is that these newer state tests can be useful tools to help us all understand how our children are thinking."

Maryann Woods-Murphy, New Jersey State Teacher of the Year & Reviewer in NNSTOY Study of Old vs. New State Tests

1. See https://cew.georgetown.edu/wp-content/uploads/2014/11/Recovery2020.ES_Web_.pdf.

2. See "Education Needs a Digital-Age Upgrade," *The New York Times*, August 7, 2011.

3. See https://cew.georgetown.edu/wp-content/uploads/StateProjections_6.1.15_agc_v2.pdf.

4. See <https://www.brookings.edu/blog/brown-center-chalkboard/2016/09/14/education-and-accelerated-change-the-imperative-for-design-learning/>.

WHERE DO STATES GET THEIR TESTS? 3 OPTIONS

For the past 15 years, federal law required states to track student learning in English language arts and mathematics with annual tests in grades 3-8 and once in high school (states also need to test in science, although not annually).

Many states are now using a collaboratively developed state assessment. Committed to creating next-generation tests to replace the multiple-choice tests widely used during the '00s, two different groups of states banded together in 2010 to create shared tests. By working together and pooling knowledge, state leaders expected to lower costs and exert more control over test quality. Today, nearly half of states participate in the PARCC (9 states, including DC) and Smarter Balanced (15 states) consortia—covering nearly 40% of all U.S. students. Participating states either use each consortium's complete testing package or customize it to add some state-specific questions. Member states also have access to classroom tools and teaching resources the consortia have created.

In addition to this option, some are using off-the-shelf tests designed by vendors for use by any state, most notably the ACT Aspire package for grades 3-10 (four states currently use some part of this package), or have opted to use ACT or SAT for their high school assessment (12 states).

Finally, a third option is for state leaders to work with vendors to design their own test. But buyer beware: Several states that tried to work with a vendor to create their own customized, next-generation tests saw these efforts implode and lead to delays, cost over-runs and unhappy parents and educators.⁵ More complicated tests are, well, complicated.

Confirming whether a test is well-designed and truly measures skills and knowledge critical to student success becomes especially important when a state chooses create its own test with a private vendor.

“When I had the chance to sit down and compare content side-by-side, it became evident the substance of the [PARCC and Smarter Balanced] tests outshined the material from the old tests in several ways.”

Josh Parker, Maryland State Teacher of the Year
& Reviewer in NNSTOY Study of Old vs. New State Tests

A GOOD TEST IN 1987

- ✓ Measured students' basic skills
- ✓ Not aligned to standards for what students should know and be able to do
- ✓ Compared students against an average
- ✓ Consisted solely or mostly of multiple choice questions
- ✓ Used manufactured articles and text passages—often poorly written—for students to read and analyze
- ✓ Only accessible to students who thrived in a traditional classroom setting; special populations were not accommodated
- ✓ Often irrelevant to the student's learning experience

A GOOD TEST IN 2017

- ✓ Probes students' writing, problem solving and critical thinking skills
- ✓ Aligns fully to grade-level expectations for what students should know and be able to do and expectations for what students need for success after high school
- ✓ Measures a range of performance, including moderate and high-performing students, not just the average
- ✓ Employs a variety of test question types (such as multiple choice, multi-step problems, short answer where students explain/justify their conclusion, extensive writing) well-matched to the skills and knowledge being measured
- ✓ Uses excerpts from actual research journals, famous essays and literature for students to read and analyze
- ✓ Allows nearly all students to demonstrate what they know and can do regardless of whether they have special learning needs or language barriers (using innovations such as pop-up word glossaries, options for simplified formats, and questions in multiple languages)
- ✓ Reflects great teaching and strong instructional practices that should be used in classrooms

5. See for example this analysis: <http://forstudentsuccess.org/the-fallacy-of-going-it-alone-on-student-assessments-memo/>

DOES OUR STATE ASSESSMENT MEASURE WHAT MATTERS? |||||

In 2012, RAND researchers examined state, national and international tests to gauge how well they measured students' writing, problem solving and critical thinking skills. The big headline: Most state tests even just a few years ago were not measuring many of these skills well; the average share of questions measuring these skills across these tests was 22% in reading and writing and, worse, just 2% in math.⁶ More recently, organizations as diverse as the Thomas B. Fordham Institute, HumRRO and the National Network of State Teachers of the Year have examined some of the latest generation of state tests—and found large variations between tests too.^{7,8}

THE AVERAGE SHARE OF QUESTIONS MEASURING PROBLEM-SOLVING AND WRITING SKILLS ACROSS STATE TESTS WAS 22% IN READING AND WRITING AND, WORSE, JUST 2% IN MATH.⁶

RAND's study included six well-known tests (such as Advanced Placement, NAEP and PISA) as well as tests in 17 states. The Fordham/HumRRO research focused on ACT Aspire, Massachusetts' old state test, PARCC or Smarter Balanced. All these reviews considered the tests' levels of "cognitive challenge" using a four-point scale.⁹ Test questions at the lower end of the scale rely on rote recall and procedures, while those at the higher end of the spectrum require reasoning, multi-step problem solving and extended thinking. High-quality tests include more questions that measure student skills at the higher end of this scale, especially as students get older and have mastered more basic skills.

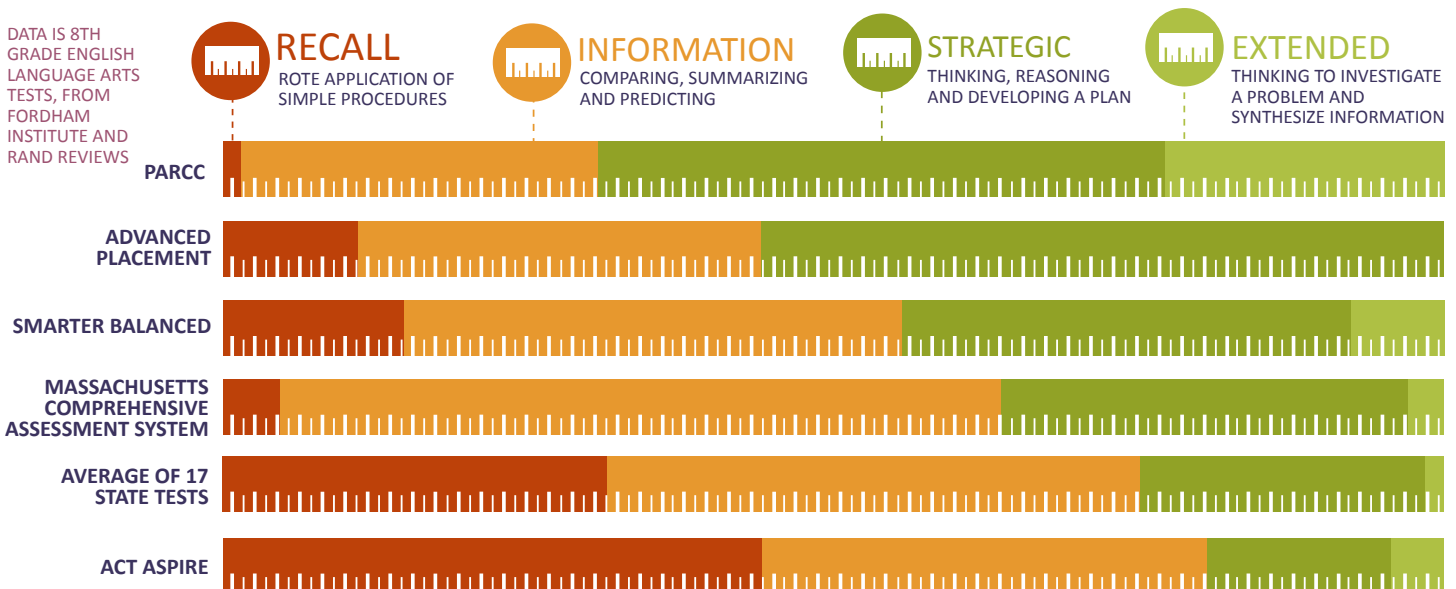
Only a handful—including the newer PARCC and Smarter Balanced tests—prioritized the "problem finding, problem framing and problem solving" skills students should be learning. Indeed, in the Fordham study, ACT Aspire and Massachusetts' test had twice as many multiple-choice questions in mathematics as PARCC and Smarter Balanced (76% and 77% vs. 25% and 32%, respectively), for example.

Across reviews, PARCC and Smarter Balanced win marks for their alignment with college readiness expectations and states' standards, for innovative approaches to accommodating students with special needs or learning English, and for their alignment with strong instructional practices. Meanwhile, the Massachusetts test—considered one of the best state tests from the past decade—was found to fall short of measuring the skills students need for success after high school, especially in English language arts. With this information, Massachusetts' state leaders are now working with PARCC on a newly designed test.

Few of the other, newer state tests now being used have been reviewed independently with publicly released results; policymakers don't know their quality. *How confident are we that a new test in our state would fare well in similar studies?*

NOT ALL STATE TESTS ARE CREATED EQUAL |||||

One study found only a handful of English language arts tests—just AP, PARCC and Smarter Balanced—measured synthesizing, thinking and writing in a significant way. Most state tests traditionally have emphasized more simplistic language and comprehension skills.



6. See https://www.rand.org/content/dam/rand/pubs/working_papers/2012/RAND_WR967.sum.pdf and https://www.rand.org/content/dam/rand/pubs/research_reports/RR400/RR483/RAND_RR483.pdf.
 7. See http://blogs.edweek.org/edweek/curriculum/2016/02/parcc_smarter_balanced_assessments_score_well.html. Fordham Institute review (grades 3 and 8 tests) at <https://edexcellence.net/publications/evaluating-the-content-and-quality-of-next-generation-assessments>, and HumRRO review (high school tests) at <https://humro.org/corpsite/press-release/next-generation-high-school-assessments>. Both reviews used the same methodology and were done in parallel.
 8. NNSTOY convened award-winning teachers in two panels to compare old state tests in seven states (DE, IL, NH, NV, NJ, OR and WA) with the newer tests these states are now using from PARCC and Smarter Balanced. But studies found the new tests superior to the old tests they replaced. See www.nnstoy.org.
 9. Mathematician Norman Webb's "Depth of Knowledge" scale categorizes learning tasks—such as test questions or classroom assignments—into four levels, according to the complexity of thinking required by students to successfully complete them.

THE QUICKEST ANSWER OR THE BEST SOLUTION?

Ultimately, it is not enough to assume that a state's test is measuring, reinforcing and reporting how well students are learning to write, read critically and problem solve. As both researchers and teachers have found, test quality still varies widely.

As state education agencies try different approaches and testing vendors pitch different products, don't just trust—verify. Every vendor and test creator claims its test measures state standards, accommodates all learners, contains the right mix of test questions, yields accurate results and provides helpful information to teachers, all using the least amount of students' and teachers' time. But asserting these things doesn't make them true—as independent, careful reviews of state tests are showing.

States invest millions of dollars in testing each year. The four steps below can ensure a state spends these funds in smart and efficient way, buying a test that measures what matters most.

1 Don't start from scratch; use existing tools.

In 2014, the Council of Chief State School Officers published criteria for measuring tests' quality; many states use these ideas to shape RFPs and negotiate with vendors.¹⁰ The National Network of State Teachers of the Year created a process for teachers to examine different tests, focusing on how well these different tests reflect high-quality instruction. Fordham and HumRRO collaborated on a state test review protocol that any state can now use.¹¹

2 Ask for evidence about test quality from independent reviewers you can trust.

Don't settle for promises and assertions from vendors and agency staff—ask for independent proof. Expect them to show you how they allowed others to “look beneath the hood” at test design and test questions—and what these reviewers found.

Or, if that review doesn't yet exist, commission one. Done by an acknowledged expert, such as the national nonprofit Achieve, a thorough and independent review is the only sure way to guarantee a state test is high quality and fully aligned to state standards for what students should know and be able to do.¹² A review becomes even more important after a state reviews and updates its K-12 standards.

3 Ask leading college and university faculty and admissions staff to review the assessments, the standards they measure and the quality of test questions.

They will be able to gauge if the assessment matches their requirements for college readiness and tell you if they think the new test has value.

4 Ensure teachers have an opportunity to review and inform state assessments.

Good teachers know the characteristics of a good assessment. Give excellent teachers a forum to review state tests and help strengthen them. Ensure that the assessment reflects great teaching. In addition to the National Network of State Teachers of the Year, TeachPlus, Student Achievement Partners and other teacher networks have led reviews in several states and can be a resource on how to design these efforts.

“As a teacher, I sympathize with parents who have grown frustrated with the number of tests their children face. But I can say with confidence these new assessments are the kind we should want our kids to take.”

Pam Reilly, Illinois State Teacher of the Year & Reviewer in NNSTOY Study of Old vs. New State Tests

10. See: <http://www.ccsso.org/resources/programs/assessments.html>.

11. The “Assessment Quality Evaluation Methodology” can be downloaded from the Center for Assessment (<http://www.nciea.org>).

12. Learn more about Achieve's state test review service here: <http://www.achieve.org/our-agenda/assessments/state-assessment-reviews>.



Education First collaborated on the research, design and writing of this policy brief. Education First is a strategy and policy organization that helps policymakers design and accelerate policies and plans that help all young people succeed in college, careers and life. Published February 2017.

CLICK HERE TO LEARN MORE > www.education-first.com/tests