THE NEW TESTING LANDSCAPE

HOW STATE ASSESSMENTS ARE CHANGING UNDER THE FEDERAL EVERY STUDENT SUCCEEDS ACT

BY LYNN OLSON

SEPTEMBER 2019

FutureEd
GEORGETOWN UNIVERSITY

About the Author

Lynn Olson, a FutureEd senior fellow, is an award-winning writer and editor. She worked in the K-12 program at the Bill & Melinda Gates Foundation from 2008 to 2017, serving as a senior program officer, senior adviser to the director, and deputy director for knowledge management and strategic alliances.

About FutureEd

FutureEd is an independent, solution-oriented think tank at Georgetown University's McCourt School of Public Policy, committed to bringing fresh energy to the causes of excellence, equity, and efficiency in K-12 and higher education. Follow us on Twitter at @FutureEdGU

Usage

THE NEW TESTING LANDSCAPE

HOW STATE ASSESSMENTS ARE CHANGING UNDER THE FEDERAL EVERY STUDENT SUCCEEDS ACT


FutureEd
GEORGETOWN UNIVERSITY

FOREWORD

When Michigan launched the first statewide standardized testing program in 1969, it sought only to know if students were achieving "minimum competency" in core subjects. Like other states, Michigan at the time had no statewide academic standards, and there were no explicit consequences for educators if their students performed poorly.

But since then, standardized testing has played an increasingly central role in the nation's ambitious effort to educate a much wider range of students to high standards. The rise of the postindustrial economy—with its requirement for brains over brawn—and the nation's emerging commitment to racial equality required that the kind of rigorous academic curriculum traditionally reserved for the few henceforth be taught to the many.

By 2001, the federal No Child Left Behind Act demanded that states set standards, measure student performance against the standards, and hold schools accountable for the results. A decade later, the Common Core State Standards and the PARCC and Smarter Balanced state testing consortia sought to establish standards and assessments shared across states that were rigorous and relevant, that more accurately measured students' grasp of demanding knowledge and skills, and that resulted in fewer assessments of higher quality.

But today, an unlikely right-left alliance of anti-Washington, Tea Party conservatives and accountability-averse teacher unions has thrown much of the work on standardized testing into reverse, even as the Common Core standards have survived in some form in many states. The political opposition and the opt-out movement it spawned have turned testing into a third-rail for policymakers. And while annual testing survived under the federal Every Student Succeeds Act (ESSA), the national consensus on testing's importance in school reform has largely dissolved, and the testing consortia have waned.

Yet standardized testing's importance—as a backstop to ambitious standards, as a window into school performance, as a driver of improvement, and as a linchpin of educational equity—remains undiminished. The education sector abandons its commitment to effective assessment at its peril, even as it must address the flaws of standardized testing and test-based accountability.

This report, part of a wider body of forthcoming FutureEd work on the future of assessment, examines the current testing landscape in the states under ESSA—the demand side of the testing equation. We explore states' testing priorities, their responses to the political winds that have buffeted standardized testing, the strengths and weaknesses of today's state testing regimes, and the prospects for innovative new systems that support both school accountability and quality instruction.

FutureEd Senior Fellow Lynn Olson has done an impressive job of capturing the changing state testing landscape, drawing on dozens of interviews with state and industry leaders who generously gave of their time and expertise. Research Associate Rachel Grich led efforts to track down a wide range of information on state testing systems with support from research associates Margarita Arguello, Kendell Long, and Jacqueline Turcios. Editorial Director Phyllis Jordan, Policy Associate Brooke LePage and the rest of FutureEd's editorial team did a great job producing the report. We hope the report is a helpful starting point for a way forward on testing, one that helps the nation's students get the high-quality education they deserve, a path that Michigan started us down decades ago.

Thomas Toch

Director, FutureEd

State testing systems are in transition. **Buffeted by anti-testing sentiment on the left and right, budget battles, and renewed debates over the purposes of standardized testing and the role of technology in testing, the recent, unprecedented push for states to collaborate on high-quality, standards-aligned assessments has given way to an increasingly fragmented marketplace.**

An increasing number of states are designing new tests for grades 3 through 8 that reflect their individual state content standards and also meet accountability requirements under the federal Every Student Succeeds Act (ESSA). At the high school level, nearly half the states have adopted commercial college-admissions exams as their accountability tests under ESSA.

Although it's too soon to tell if these shifts will lead to diminished test quality, they have already resulted in less transparency and comparability across states. With many states' testing contracts set to expire in the next few years, the churn is likely to continue.

At the same time, states have shown growing interest in designing assessment systems that better reflect and support the daily work of students and teachers in classrooms. These include faster turnaround of test results, as well as greater use of end-of-unit tests, performance-based tasks that ask students to apply what they know and can do, and tests that are more closely linked to the curriculum. Such efforts could provide better ongoing information about student progress, while giving teachers more guidance on how to adjust instruction. But to date, except for in a handful of states, there has been more talk than action.

The large-scale shift to online testing (virtually all states now administer their annual tests online) offers opportunities for automated scoring, computer-adaptive testing, and technology-enhanced performance tasks that weren't possible at scale a decade ago. Yet with limited federal money for innovations in state testing, and limited philanthropic dollars compared to 10 years ago, states have few real incentives to push the envelope.

U.S. Secretary of Education Betsy DeVos recently challenged state chiefs to put forward more innovative testing proposals, saying, "Anything that is ultimately going to result in greater student achievement is going to be seen very favorably by the Department of Education."[1] Yet states are holding back until it becomes clear what changes the department will approve as it reviews state testing systems, including proposals under ESSA's innovative assessment pilot.

To get a handle on the state of state summative assessments under ESSA, FutureEd conducted a scan of state testing programs across the 50 states and Washington, D.C., based on publicly available data in the summer of 2019; looked more closely at half a dozen state systems; and conducted interviews with more than 20 experts in the field, including state assessment directors and the representatives of commercial test publishers. Except where otherwise noted, all of the quotes in this piece are from one-on-one interviews conducted between March 1 and April 30, 2019.

For now, a few trends are clear, although not all of them are necessarily tied to ESSA:

- The marketplace for summative assessments in grades 3-8 has fragmented. As of the 2019-20 school year, the largest vendors holding state testing contracts in these grades were American Institutes for Research (whose assessment division was recently purchased by Cambium Learning), Data Recognition Corporation, Pearson, and Measured Progress (now called Cognia, after it merged last year with AdvancedED, which was focused on education quality and accreditation).[2]
- More states are embracing the ACT and the SAT as their high school assessments despite concerns that the tests are not fully aligned with state standards. Language permitting this option under ESSA has accelerated the trend, enabling states to offer tests that parents and students actually care about.
- While states and the federal government still prioritize the ability to compare student test results within a state, comparing results across states has faded as a dominant goal.
- As a result of PARCC and Smarter Balanced testing consortia, there's evidence that both the quality and rigor of state tests have risen. To date, cut scores—or the benchmarks states set for student performance—appear to be holding the line in most places, but it's harder to get a handle on test quality.
- Churn in state testing systems is a big concern. While no one is tracking how many states have changed their tests or assessment vendors multiple times in the past five years, many of those interviewed mentioned it. Constant changes in state assessment systems make it harder to track performance over time, create problems for state and district accountability systems, and send mixed messages to educators, diminishing their morale and ability to focus instruction.
- Innovation is happening—particularly when it comes to technology-enabled testing. But despite growing interest among states in doing something different, especially in connecting state tests more closely to classroom work, there's still more talk than action. To date, four states—Georgia, Louisiana, New Hampshire, and North Carolina—have been approved for the federal innovative assessment pilot.

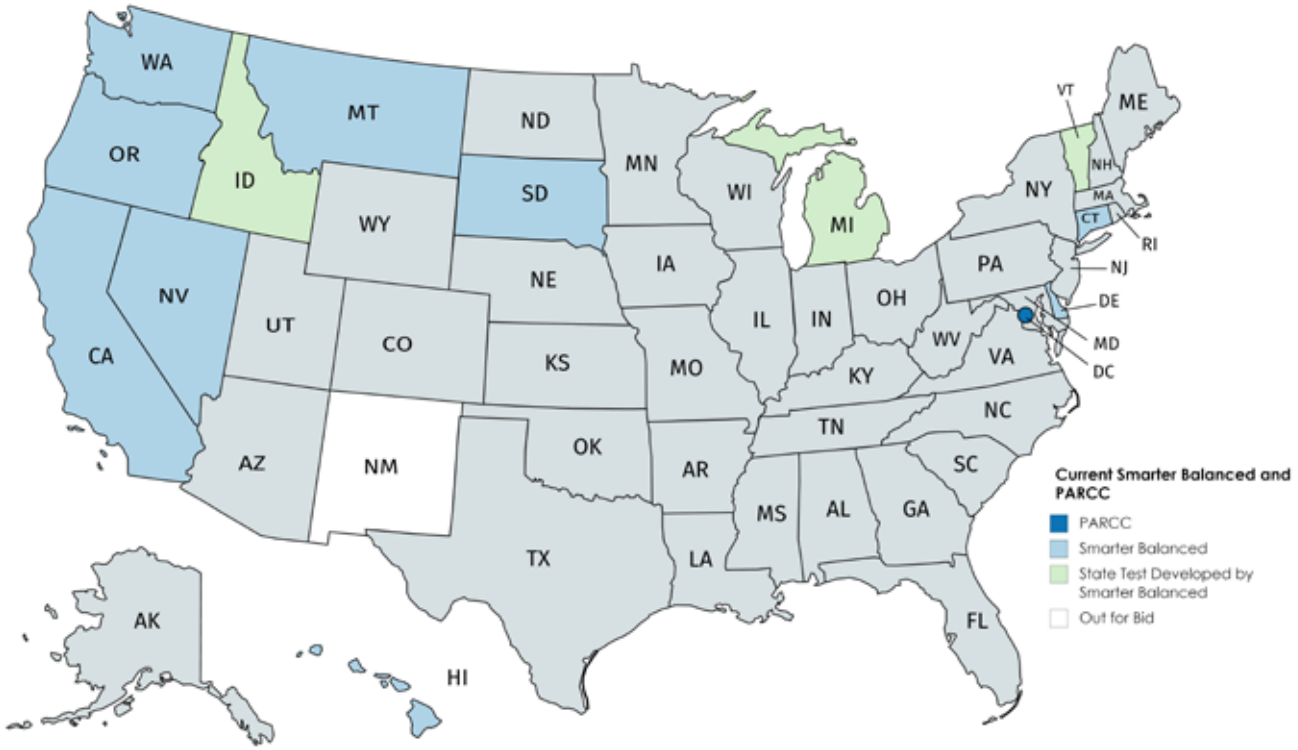
An Increasingly Fragmented Testing Market in Grades 3-8

The Common Core State Standards, and the creation of two state assessment consortia designed to measure student performance against those standards, promised to usher in a new era of quality, transparency, and comparability in state assessments. This was viewed as an important correction to states' use of multiple-choice and short-answer tests focused on basic skills under the No Child Left Behind Act, the prior reauthorization of ESSA. Yet since 2010, more and more states have decided to go it alone when it comes to state testing.

In 2010, most states (45) confirmed plans to use standards-aligned assessments developed by one of the two, federally funded consortia: PARCC or Smarter Balanced assessments in math and English/language arts. Today, a dozen states remain part of the Smarter Balanced consortium, while only D.C. is using PARCC, whose original model is all but defunct. Even so, several states plan to use PARCC items in their new tests in the 2019-20 school year.

There are many reasons for the rapid change in status. Policymakers trying to cope with pushback against the Common Core and teacher evaluations linked to

States that will Administar PARCC or Smarter Balanced in 2019-2020



test results frequently opted out of the consortia as a tradeoff to preserving state standards. The tests were long, contributing to a small but vocal movement by parents to opt out of state testing requirements. And PARCC's design, which required member states to use a single testing vendor, proved problematic; lobbying by competing, for-profit testing companies also contributed to PARCC's demise.

Smarter Balanced has retained a dozen consortia members, in part by maintaining control over test content and item design while enabling states to select their own vendors for test administration, scoring, and reporting. Smarter Balanced held firm on setting consistent cut scores across states. But early on, the consortium reduced its emphasis on publicly reporting cross-state comparisons of student results. Members were concerned that the percent of proficient students would differ between what the consortium and the states themselves reported based on variations in each state's accountability system, creating a communications nightmare. "We stepped away from that very quickly," says Tony Alpert, the executive director of Smarter Balanced. "We focused on within-state comparability and really helping the states have high-quality assessments, with an emphasis on fairness for all students, specifically around accessibility."

"I think one of the lessons from Smarter Balanced is there are places where states need independence from one another, and Smarter Balanced gave them that in a way that PARCC never did," notes Joanne Weiss, an independent consultant who formerly served as the chief of staff for then-Secretary of Education Arne Duncan. "PARCC was tight, tight, tight on everything."

Yet even states that have withdrawn from the consortia are not proceeding entirely alone. American Institutes for Research (AIR) draws from a common bank of test items to customize assessments for states, including an increasing number of technology-enhanced performance

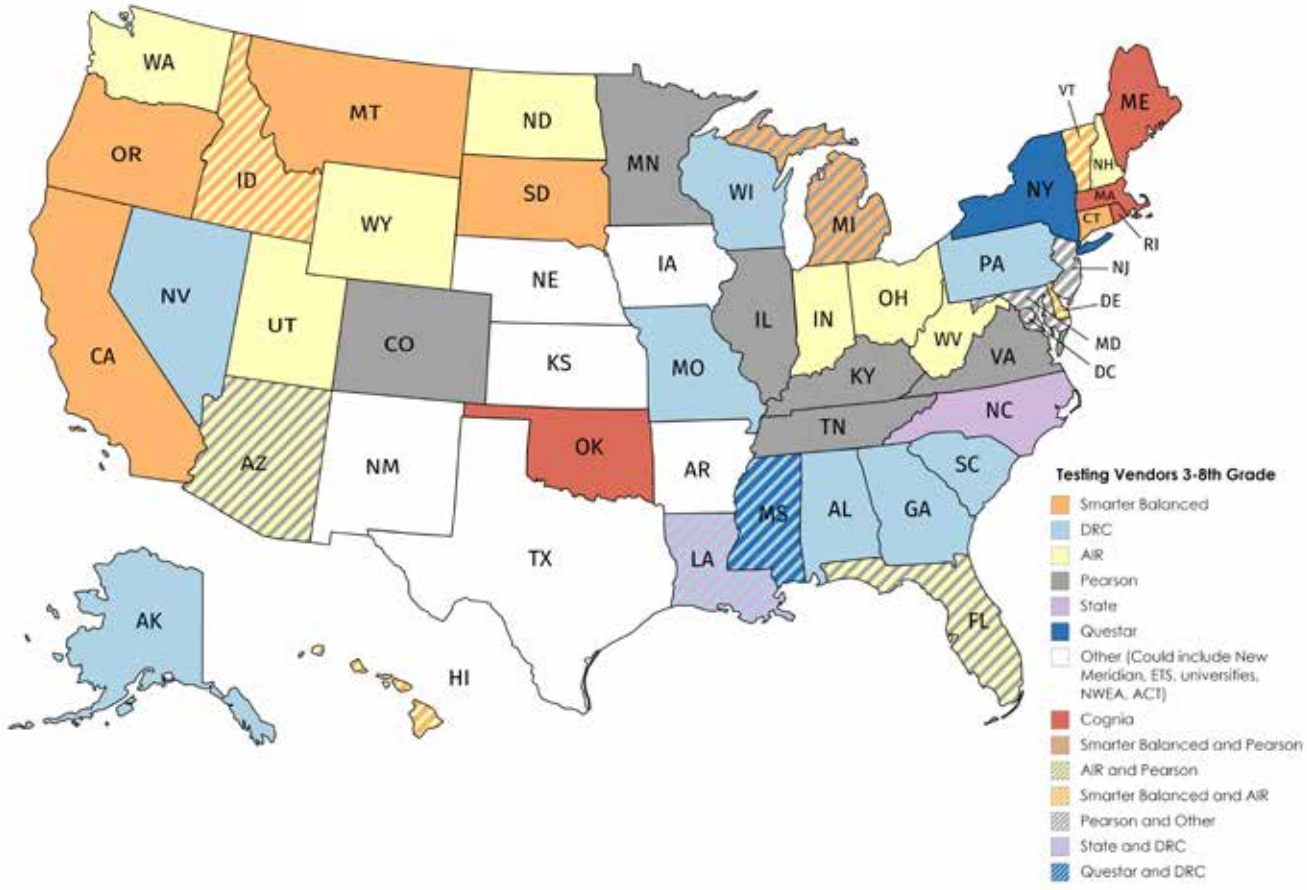
tasks. New Meridian, which in 2017 got the contract to sustain the PARCC item bank after the consortia stepped away from its all-or-nothing model, now has an even more extensive New Meridian item bank, which nine states currently license to help develop their state tests. "I think that things like New Meridian and AIR, to some extent, are potentially promising models," says Weiss, "especially if New Meridian is really going to have very high-quality items, stay sort of innovative and cutting edge, and be able to mix that with custom items that a state needs and wants."

"Any state that wants to license PARCC or Smarter Balanced items, I think that's going to create pressure to make sure the rest of your test measures up as well," says Michael Cohen, the president of Achieve, which formerly managed the PARCC consortium. "But otherwise there's not a ton of pressure on quality as far as I can see."

Illinois, for example, put out a request for proposals more than a year ago to develop a new computer-adaptive assessment. It will use New Meridian's content but work with a separate vendor on test administration, scoring, and reporting, similar to the Smarter Balanced approach.

Arthur VanderVeen, the president and CEO of New Meridian Corporation, notes that as states have moved away from the consortia, they've realized it is more expensive to develop tests at the same level of quality: "We've been seeing an increasing interest among states in licensing this high-quality content, but incorporating it into their own custom design … It's a strong trend and it makes a lot of sense. As a country, we spend ridiculous amounts of money every year developing new custom content for each individual state, basically covering the same learning standards."

States Testing Vendors Grades 3-8



Notes: Measured Progress merged with AdvancedEd to form Cognia; AIR's assessment division was recently purchased by Cambium Learning

Adopting a Hybrid Approach in Michigan

Michigan has developed a hybrid testing system that reflects cross-state work on standards and assessments, while designing something that is uniquely its own.

In June 2010, the Michigan State Board of Education adopted the Common Core State Standards as the state's K-12 content standards for math and English Language Arts. But subsequent pushback against standards made it politically difficult to use any test perceived as coming from out-of-state.

As a result, Michigan uses items from the Smarter Balanced testing pool to develop the customized Michigan-Student Test of Educational Progress, or M-STEP, in grades 3-7 in math and ELA. Data Recognition Corporation and Measurement Incorporated have the current contracts to score and administer the exams.

"From the beginning, we were never really allowed to use the word Smarter Balanced publicly or to say, 'Now it's time to take the Smarter Balanced test'," notes Andrew Middlestead, the director of the office of educational assessment and accountability in the Michigan Department of Education. "We needed a unique name to satisfy some of the concerns that were out there." The tests are given online, which Middlestead says students like better.

Since 2007, Michigan has used a college-entrance exam as the main part of its 11th grade test, the Michigan Merit Exam, in response to a legislative requirement. The state gives the PSAT in grade 8 in ELA and math, followed by the Michigan Merit Exam in high school, which includes the SAT, M-STEP in science and social studies in grade 11, and ACT WorkKeys, to measure career-readiness skills. The state switched from the ACT to the SAT when its existing contract was up, based on state procurement rules. "We didn't want to change because we knew it was going to be an absolute political nightmare, but the rules were the rules," says Middlestead. "This was a lot cheaper, and they had a better proposal. At the end of the day, it was bumpy for a few months, but it's actually worked out far better."

In the past, the state has offered districts a set of benchmark assessments in grades K-2 as optional tools for schools to use to supplement the end-of-year tests. But in 2019 the state legislature allocated money for districts to purchase their own benchmark assessments in grades K-2 and 3-8. These grants have led to intense lobbying for business by vendors.

Michigan applied for a waiver from the federal government to allow some districts to develop an alternative to M-STEP. Some Michigan districts are working on competency-based education and the state superintendent at the time was trying to respond to their needs. But that waiver, which Middlestead describes as very vague and open, was declined. The state is working with those districts to help build out some competency-based education models. "We're still waiting to see where we might go from here," says Middlestead, "because the Innovative Assessment Demonstration Authority was not what our chief had in mind. It's a much more rigorous process that takes years of effort."

"A state like Michigan is bigger than New Hampshire, with so many more kids and schools" he adds, referring to the Granite State's competency-based assessments pilot. "So it's not a one-to-one model."

VanderVeen predicts that "the decoupling of test design and administration is going to open up greater flexibility for states to continue to competitively bid the large bulk of their testing program—administration, scoring, and reporting—while maintaining continuity in their scale and test content." This could provide states with greater control over test content while potentially saving money. Yet the jury is still out on whether using common item banks as part of a broader state test will have big uptake.

Shifting to College-Admissions Tests at the High School Level

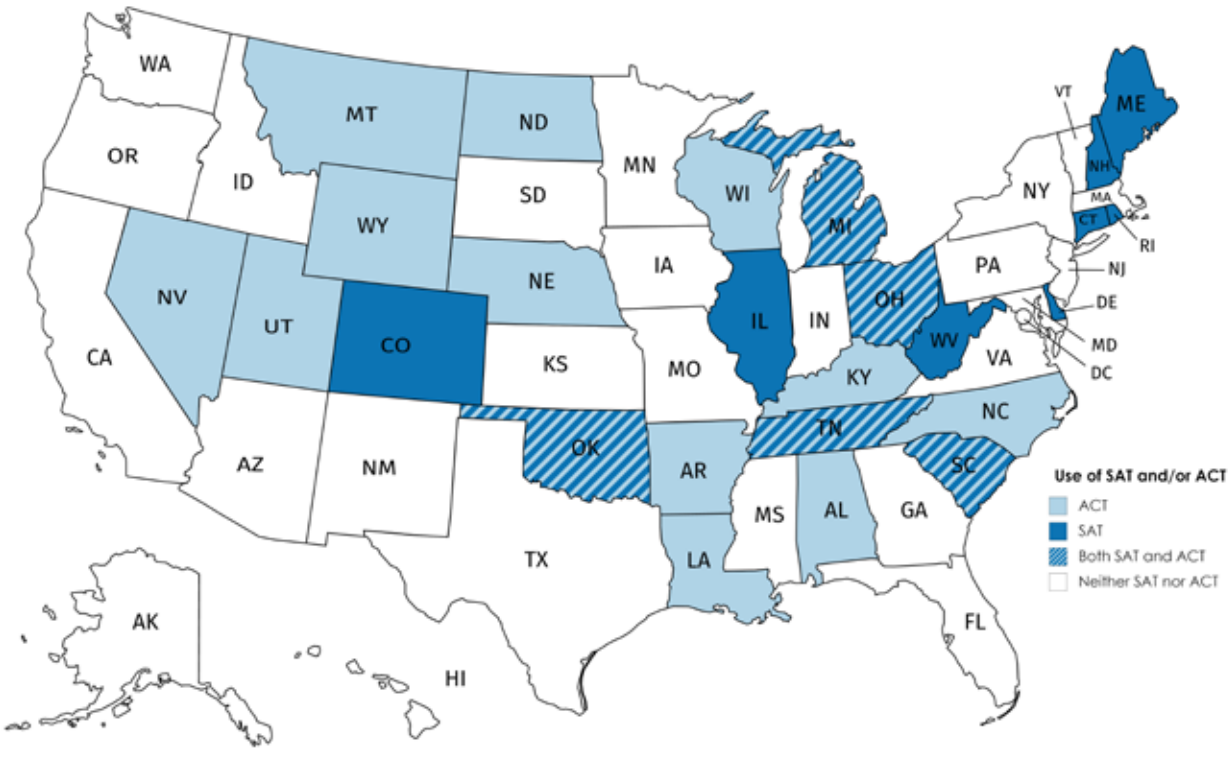
A new option under ESSA allows states to offer districts the option of using a nationally recognized college-entrance exam in place of their state high school test if the entrance exams meet certain technical requirements. This new option has accelerated a trend by states to replace their own exams with either the ACT or the SAT.

"We saw states interested in using the SAT for accountability before ESSA passed," says Michelle McNeil, senior director for K-12 policy analysis at the College Board, SAT's parent company. "I always say ESSA kind of rolled out the red carpet to proceed."

The tests have numerous attractions for states, according to Achieve: nearly all higher education institutions use them for admissions decisions; they are shorter in length than most state-designed and consortia tests; they have brand-name recognition to parents and students; and they are known to predict first-year college performance.[3] In states that use them as their annual high school exam, students get to take them for free.

In 2019-20, eight states will use the SAT to meet ESSA's high school testing requirement in math and English language arts. Eleven states use the ACT as a federal accountability test at the high school level. Five states will allow districts to choose between the two exams. Michigan uses the SAT alongside ACT Work Keys, which

Use of SAT/ACT for ESSA Accountability



Notes: New Mexico's college entrance exam is out for bid; Arkansas uses ACT Aspire (9th/10th grade), but not the ACT (11th grade) for accountability; Ohio uses the SAT/ACT for some students for accountability

Delaware Creates Next Generation Science Assessments

Delaware is changing how students are assessed in science to emphasize applying scientific knowledge to real-world problems. Delaware adopted the Next Generation Science Standards in October 2013. The standards set expectations for what students should know and be able to do in physical science, life science, earth and space science, and engineering, technology, and science application. The science standards differ significantly from prior ones by integrating academic content, science and engineering practices, and cross-cutting concepts. To measure such skills, Delaware is developing a comprehensive system of assessments in grades 3-10 that includes:

- Teacher-developed quizzes to provide information about student learning in real time. The quizzes, developed by teacher leaders from across the state, will become part of an open access item bank that teachers can use at their discretion throughout the year.
- Voluntary paper-and-pencil tests at the end of units (approximately three per year) aligned to a specific content area. Teachers can choose to use the tests to inform instruction, as well as program evaluation at the classroom, school, and district levels. Teachers can use the end-of-unit tests as a course grade, in place of existing classroom unit tests.
- End-of-year, online assessments in grades 5, 8, and high school biology that go beyond multiple-choice and short-answer items to include technology-enabled performance tasks meant to capture the ways that students integrate, transfer, and apply science knowledge and skills learned throughout the year. The assessments will be used to meet the requirements under the federal Every Student Succeeds Act (ESSA). They were field tested last school year. The testing window this school year is between April 1 and May 30.

Delaware is working with WestEd, a San Francisco based technical-assistance provider, on content development of the unit and end-of-year tests, and with Pearson Assessment on administration. The Delaware Science Coalition—which includes educators, district and charter leaders, and representatives from the higher education and business communities—has been key to development of the state's plan.

Performance on the statewide, end-of-year science assessments in grades 5, 8, and high school biology are one of several indicators used to measure school quality under the state's accountability system, along with chronic absenteeism, college/career readiness, and on-track-to-graduation in 9th grade. Together, these indicators account for 20 percent of a school's overall rating for elementary and middle schools, and 40 percent for high schools.

"Delaware has actually been really involved in science education for a long time," observes April McCrae, an education associate for science assessment in the Delaware Department of Education. "We were already in that innovative, forward-thinking mindset from years ago."

In 1995, the state received a local systemic change initiative grant from the National Science Foundation to bring all of its stakeholders together around science teaching. This led to the creation of the Delaware Science Coalition and a long history of supporting teacher professional development through the use of hands-on curriculum materials and formative assessments based on diagnostic rubrics.

measures career readiness. New Mexico, which recently dropped the PARCC assessment, is in the bidding process to use either the ACT or SAT as its high school test. Other states are using the tests as part of a college-readiness indicator, often as one of multiple measures.

States using one of the college-entrance tests for their high school exam under ESSA must ensure it measures progress toward state standards at least as well as the original state test, subject to USDOE approval. There are indications the tests may not be fully aligned. In 2017, Achieve conducted an independent alignment study of the ACT with the Common Core State Standards. Overall, the review found significant weaknesses across both ELA (particularly in writing) and mathematics for both content and depth. In both ELA and math, fewer than half of items were judged to be aligned. The ACT's technical documentation indicates that 40 to 43 percent of math items are designed to measure pre-8th grade math content.[4]

In 2016, Delaware and Maine, which adopted the SAT as their statewide summative high school test, commissioned an alignment study from the Human Relations Research Organization, which found the SAT is "reasonably aligned to the high school reading and writing portions of the CCSS, but less so for the math portions."[5] The study found a number of below-high school-level math items and concluded the exam did not adequately assess geometry or statistics. Recently, the Florida Department of Education funded a set of studies to examine the alignment of both the ACT and the SAT, as well as the accommodations they provide, to determine if the tests are likely to meet federal peer review criteria. The study found that neither test was fully aligned to Florida Algebra 1 standards and would require at least some augmentation. For ELA, the study found the ACT would need major adjustments.[6]

The concern, according to a summary of the studies by Achieve, "is that many high school teachers will be driven to devote scarce course time to middle school topics, water down the high school content they are supposed to teach in mathematics, or too narrowly focus on a limited range of skills in ELA."[7]

"The feds are going to require a certain amount of alignment," notes Julie Woods, a policy analyst at the Education Commission of the States. "There haven't been a ton of studies that are publicly available about the alignment, but the ones we have seen haven't been overwhelmingly positive. So, I tell folks there's always a tradeoff."

College Board officials admit that the SAT is not totally aligned with state standards but measures the "core of the core." Stefanie Sanford, chief of global policy and external relations for the College Board, says the argument that the SAT is not designed to measure mastery of state standards, and that one test cannot serve multiple purposes, does not resonate with the public. "The general public assumes that a college-readiness test measures what matters in high school because the goal of high school is college readiness."

"No test can cover all the standards, especially a three-hour test," adds McNeil. "Interestingly, our states are OK with that. They know it's not perfect but they don't have an opt-out problem anymore."

Another concern is that state results on the two college-entrance tests may be skewed because more-affluent families pay for test prep on the college-admissions exams, something they didn't do on the previous state high school tests. The College Board has partnered with Khan Academy, an online, personalized-learning resource, to make free SAT prep available for all students. A 2017 study by Khan Academy and the College Board found a positive relationship between at least 20 hours of practice on Khan Academy and score improvements on the exam. Score gains were consistent across gender, family income, race, ethnicity, and parental education level.[8] McNeil says the two organizations are exploring how to improve the site further "to help clear a path for kids."

While some see the shift to college-admissions tests as a point of tension still troubling to states, others view it as a fait accompli. "I think we just need to embrace that trend and figure out a way to make it work," says Scott Norton, deputy executive director of programs at the Council of Chief State School Officers (CCSSO). States could pressure the ACT and the College Board to add a set of questions to their state administration of the exams to more fully measure state standards, but they don't, adds Weiss.

Michigan used the ACT as its state test for nearly a decade before switching to the SAT when the existing contract was up, because the College Board submitted a cheaper, better bid. Andrew J. Middlestead, the director of the office of educational assessment and accountability in the Michigan Department of Education, notes, "I think people are looking for a test that students actually value and that they think [is] important. All this other stuff doesn't mean anything to students. A Smarter Balanced test would be just as rigorous, but until higher education accepts it as a college-entrance exam, that's not going to go anywhere." While there had been hope that higher education systems in many states would accept scores on the consortia tests for college admissions, so far only a handful do so.

Using the tests for federal accountability purposes has pushed both the ACT and the SAT to offer more accommodations for students who need them to comply with federal law. In early 2017 the College Board announced that it would automatically approve accommodations that are part of a student's IEP or 504-plan rather than having the College Board separately consider each accommodation request.

States Focusing on Comparing Scores Within States

ESSA requires every student in the same grade to take the same test (excepting students entitled to take alternate assessments because of a disability or limited English proficiency). This key requirement promotes equity by ensuring that all students and schools are measured against the same expectations. So far, states and the USDOE remain focused on within-state comparability, but are less interested in comparability across states.

The USDOE recently notified Arizona that it could lose $340 million in federal funding because it hadn't followed the testing rules under ESSA. The state passed a law allowing its schools a choice of tests at both the elementary and high school levels, including the ACT, SAT or the state test, AZMerit, in high school.

In a letter to state officials, Frank Brogan, the assistant secretary for elementary and secondary education, cited the failure to offer students the same test statewide as the reason the state had been put on high-risk status; the state must select a single high school exam or risk losing federal Title I funding for disadvantaged students. For the 2019-20 school year, Arizona plans to require 9th graders to take AZMerit, rather than giving high schools a range of options. "In the future, it may be possible to have a menu that allows for the use of a different assessment instead of the statewide assessment," superintendent Kathy Hoffman wrote districts, "but it will take time to meet the federal requirements outlined in the letter" from USDOE.[9] Oklahoma also allows districts to choose between the ACT (with the writing section) and SAT in high school, but has met the law's technical requirements and received federal approval.

While states and the federal government continue to prioritize within-state comparability, the desire to compare results across states seems to have faded. "It definitely mattered a lot more to people who sat in positions like the one I was in than to district superintendents," says Chris Minnich, the former executive director of CCSSO and now chief executive officer of NWEA, a national assessment vendor. While NWEA's computer-adaptive tests can provide states and districts with national norms without a lot of effort, he notes, customers are more interested in knowing how their district compares to others in the state.

The state testing consortia promised three things, according to Scott Norton of CCSSO: alignment to the standards, better quality, and comparability across states. Norton continues, "Honestly, my sense is two of those three really happened. I think comparability faded a little bit. I think people would still like it if they could have it, but I don't hear a lot about that anymore."

"I think it was something that was lost in the fray," agrees VanderVeen of New Meridian. "Many education policy people I speak to lament that, but for most state political leaders and even educational leaders, it's not top of their priority list."

Yet while in-state comparisons are important, they arguably do nothing to ratchet up standards in low-performing states like Alabama and Mississippi.

As states move away from offering the full Smarter Balanced or PARCC testing blueprint, it also gets harder to know whether their state tests are comparable. Michigan, for example, is the only Smarter Balanced state that currently uses Smarter Balanced content in addition to its own material to develop a customized state test. It has stopped administering the full Smarter Balanced blueprint in ELA. Washington has shifted the Smarter Balanced 11th grade tests in ELA and mathematics to grade 10, which included removing some Algebra 2 items that students are unlikely to be exposed to by that grade. Smarter Balanced has contracted with the National Academy of Education to produce a set of papers to fully investigate such comparability issues and the implications for policy.

Quality and Rigor Have Improved, But Will These Improvements Last?

A number of studies have found the quality and rigor of state tests have improved since the advent of the two state testing consortia in 2010. Now, some worry whether that upward trend will continue.

One of the most important features of state tests today is their focus on college and career readiness, rather than lower-level knowledge and skills, notes Laura Slover, who helped launch PARCC and is now CEO of CenterPoint Education Solutions. PARCC and Smarter Balanced set such advances in motion, she argues, by establishing common performance levels across states within each consortium through a process that engaged both K-12 and higher education leaders.

A recent study from the National Center for Education Statistics found that cut scores for what states consider proficient have risen when compared to performance levels for the National Assessment of Educational Progress, a nationally representative test often known as the "nation's report card." The difference between the states with the lowest and highest performance standards narrowed between 2007 and 2017. Most state performance standards, however, still correspond more with the NAEP basic than proficient level in grades 4 and 8.[10] A more recent NCES study confirmed these trends through 2017. An analysis by *Education Next* also found a higher bar for student performance in most places, though the finding was not consistent from state to state.[11]

"States haven't gone back to the days of low-cut scores when almost all kids were proficient," says Scott Marion, the president of the nonprofit National Center for the Improvement of Educational Assessment (the Center for Assessment). "That's been, perhaps, a lasting positive effect that they've held the line pretty well."

But he and others worry the move away from the consortia will reduce test quality, given the pressure on states to reduce testing costs and time. In particular, it could reduce the use of open-response items that ask students to actually write, which take longer to complete and are more costly to score. "Many states have said goodbye to the writing components or are cutting down on the writing components," says Slover. Massachusetts, for example, reduced the number of longer writing passages when it moved away from PARCC. "If states are looking to shorten the test, which all of them are," she adds, "the fastest way to do that is to lose some portion of writing, which is a step backwards."

There's also less transparency about what's actually happening with state tests. PARCC and Smarter Balanced released a large number of test items and publicly released information about test quality and alignment to standards. Smarter Balanced even makes its item specifications available online. But there's far less pressure on states to provide a similar level of transparency. "I think we literally have no idea about the quality of some of these tests," says Weiss.

Such transparency is foundational to building family and educator trust in state tests, argues VanderVeen. But he adds: "It's expensive to release test items. You've got $1,000 to $10,000 an item walking out the door every time you release an item, and that's expensive to sustain." In

Georgia's Innovative Approach to Assessment for Learning

In 2018, the Georgia legislature passed a bill (SB 362) to create an Innovative Assessment Pilot that gives local districts the opportunity to develop alternatives to the state's annual tests. It also directed the state education department to pursue maximum federal flexibility to implement the pilot. The state received federal approval in July 2019 as part of the Innovative Assessment Demonstration Authority.

Georgia's approach stands out because it will be piloting not one but two alternatives to the existing state tests. While the two approaches differ, both replace end-of-year state tests with more periodic assessments that will "roll up" into a final score.

- The Georgia MAP Assessment Partnership, which includes more than a dozen districts, is working with NWEA to use the nonprofit's computer-adaptive assessments to track progress over the course of a year, giving tests in the fall, winter, and spring that would produce a summative end-of-year score. Similar to the work underway in Nebraska, the through-year assessment would leverage interim, adaptive tests to provide timely insights into students' command of grade-level standards, measure academic growth, provide norm-referenced test results, and produce summative proficiency results.
- The Putnam County Consortium, of about 12 districts, is using a through-year, web-based assessment system developed by Laine P. Bradshaw, a professor at the University of Georgia. Bradshaw, the founder of Navvy Education LLC, developed the classroom-based, diagnostic assessment system to provide real-time feedback for teachers about students' competencies in grades 3-8 and high school math and ELA. The system uses novel psychometric methods to provide ongoing, actionable information about what students understand and where they need additional help. Teachers can give students mini-assessments on one or more standards whenever they are ready, as long as students cover all the grade-level standards by the end of the year.

Lawmakers appropriated $175,000 for the state Department of Education to provide technical assistance to the districts and establish a technical advisory committee to work with them. "Districts have definitely been interested because, like most places, they've expressed frustration with the traditional assessment model, where it just happens once at the end of the year," says Allison Timberlake, the state's deputy superintendent for assessment and accountability. "They wanted something throughout the year where they could get more timely feedback to adjust instruction for individual students."

addition, he notes, "smaller states working with vendors often can't afford to really engage large numbers of teachers in the content-review process."

Alpert, the executive director of the Smarter Balanced Assessment Consortium, particularly worries about holding states accountable for making their tests accessible to students with disabilities or limited English. "I think transparency in that regard is incredibly important," he says, "and I'm not sure that all assessments are being held to the same standard that Smarter Balanced adheres to."

Meanwhile, as philanthropies have moved away from funding state-level work related to standards, assessments, and accountability, there's less money for advocacy groups that once kept a strong focus on the quality and rigor of state exams.

But Minnich, whose nonprofit NWEA has become successful by offering districts computer-adaptive assessments that measure academic growth and reveal student learning, says the two consortia focused too much on providing a Cadillac model that optimized on delivering the best test possible—regardless of testing time—rather than on ease of use and quick feedback for teachers in classrooms. "I think we optimized on the wrong thing," he says. "I still think we need a lot of that content, but I don't think it's sustainable, in most states, to have a six-hour testing experience for a kid. It's not going to work."

In 2013, CCSSO and the Council of Great City Schools jointly released a set of principles for high-quality assessments.[12] Recently, CCSSO has been working with states on how to embed those principles into their testing RFPs as a means to ensure quality.[13] In spring 2019 it held a workshop for more than 20 interested states and hopes to provide states with individual technical assistance and follow up.

"Doing assessments in a high-quality way is hard and expensive," says Weiss, "and doing it with other states, both in terms of sharing the costs and getting more money and more expertise, really helps. Going it alone is antithetical to both of those things. I'm very worried it's going to push quality down, not up, if we're not careful." Particularly if states band together, she argues, they have more power with vendors than they think. "But they have to stay firm and wield it together" to move the market.

Here Today, Gone Tomorrow: The Churn in State Testing Systems

Unfortunately, turnover, not stability, is the mark of most state testing programs, as states shift from one vendor to another. "I think the biggest challenge that I've heard … is a continual change in some states in the testing program," says Norton of CCSSO. "It seems like many, even most states, are changing tests way more frequently than used to happen."

There are many reasons for such turnover, including modifications in state standards; problems with state test administration, scoring, or reporting; the desire to lower costs; and politics. But such turnover comes with a price. The disruption makes it hard to track trends in student learning, maintain stability in accountability systems, and build parent and educator trust and support as the benchmark keeps moving.

Often, it's the stuff of headlines. Tennessee withdrew from PARCC in 2014. What followed was three straight years of testing problems—under two different vendors. The snafus in test administration led the state to suspend elements of its accountability system and diminished support among educators. "Unfortunately, there is now, and there's going to be for some time, a mistrust in our state testing process at all levels," says Jerry Boyd, superintendent of the Putnam County School District. "We're going to have to overcome that. But it will take some time to rebuild that trust." In Spring 2018, Utah signed a $44-million contract with Questar Assessment Inc. In June 2019, state officials abruptly canceled the contract after a string of technological glitches.

Marion argues that the "Massachusetts Miracle," the state's reputation for more than a decade of educational

Louisiana Changes State English Language Arts Tests to Use Curriculum Texts

It's widely known that students with more background knowledge read at more advanced levels. Yet states have built reading and writing tests that are largely content agnostic and do not typically measure students' deep understanding of the books and texts they have studied previously. Now, Louisiana is building an innovative testing format that will cover both English Language Arts (ELA) and social studies to encourage standards-based instruction and content-rich curriculum in all Louisiana classrooms. The goal is to have teachers focus on helping students make meaning out of the texts they actually use in class.

In the past decade, Louisiana has adopted higher standards and developed an optional text- and content-rich ELA curriculum (ELA Guidebooks 2.0) with units built around general themes, knowledge domains, and "anchor texts." As students increase their background knowledge by reading multiple texts on the same topic within the same unit, they're better prepared to access more complex texts and subjects over time.

The state's current Louisiana Assessment of Education Progress (LEAP) measures the state's ELA standards, including such skills as asking students to summarize passages and locate main ideas. But it does not go above that to measure whether students have developed a base of knowledge to support reading comprehension. That has encouraged schools to focus on discrete reading skills. The new assessments, being developed with NWEA, will bring ELA and social studies standards, curriculum, and assessments into full alignment. "So, the effect of the assessment is that teachers will have an incentive now to focus on the meaning of texts and on students making meaning of texts," says John White, state superintendent of education.

The pilot, which has been approved under the federal Innovation Assessment Demonstration Authority, is starting in grade 7 and will build out to the rest of middle school next. The online assessments will be given three times a year at the conclusion of curriculum units. School systems will have a choice of units, rather than requiring a single curriculum. (About 85 percent of Louisiana districts use ELA Guidebooks 2.0.) Students will complete a series of writing tasks—some based on texts they've already studied in class and others that are new to them, but closely aligned with the content they've been studying. The final assessment, in particular, will allow students to bring in evidence from texts they've read throughout the year, not just respond to a single body of knowledge and texts. The mini-assessments will be scored by teachers and rolled up into a final, summative score.

DRC is the vendor for Louisiana's statewide test, LEAP, which includes PARCC items that enable Louisiana to compare student performance with that in other states, while maintaining greater state control over the assessment. At the high school level, Louisiana gives both LEAP end-of-course tests and the ACT. The state is working with the Center for Assessment, Johns Hopkins University, NWEA, and Odell Education to evaluate the pilot assessments.

progress, stems less from the quality of its standards and tests than from its ability to keep the same system for 18 years, so that educators could focus on aligning teaching and learning, professional development, and curricula over time.

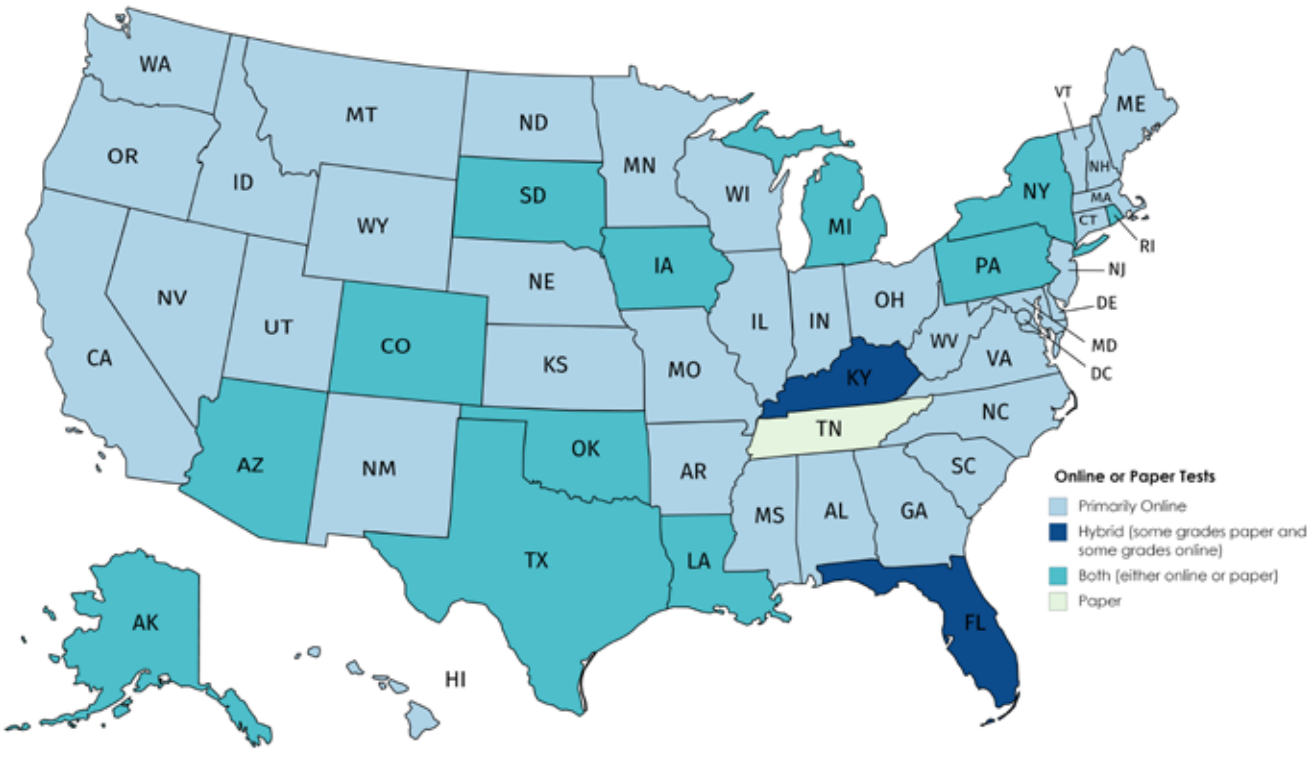
Yet more change is on the horizon. The 2018-19 school year was the last year for administering PARCC. Twenty states had contracts with testing vendors expiring in 2018, with some states renewing such contracts annually.[14] "As a lot of the contracts that were put out to do the first round of Smarter Balanced and PARCC testing run out," says Middlestead of Michigan, "I'll be interested to see if there's a lot of vendor change around those states." The next few years will tell a lot about where state tests are heading. While the picture for state assessment budgets varies state by state, officials in most states said those budgets have held steady. Barry Topol, Managing Partner at Assessment Solutions Group—a for-profit firm that provides consulting on assessment costs, management, and accountability systems—estimates that in the 2018-19 school year, states spent an average of $25 per student for their math, reading, and writing tests, and about $46 per student if other subjects (such as science) and alternate high school tests (such as the ACT and SAT) are included.

The bigger problem may be the decreased size and capacity and increased workloads for state education departments since the agencies are responsible for making so many assessment decisions.

Technology-Enhanced Testing: No Going Back

Despite periodic snafus with online testing in some states, there's clearly no going back to paper and pencil. The change cannot be attributed to ESSA; a far greater impetus was likely the additional dollars available to states under the American Assistance and Recovery Act during the last recession, including $650 million for educational technology. But reduced testing time and cost, quicker results, greater access for English language

Paper vs. Online Test Administration



New Hampshire's Performance Assessment of Competency Education

New Hampshire is piloting classroom-embedded performance tasks in English language arts, math, and science to replace annual state tests with the goal of providing students a richer, more individualized learning experience.

Schools and districts participating in the Granite State's innovative assessment pilot—Performance Assessment of Competency Education, or PACE—supplant much of the traditional end-of-year state testing with teacher-developed performance tasks. These include one common task in each grade and subject combination without a state test, which participating districts agree to collaboratively develop and administer. Districts also commit to using other, locally developed performance tasks that can vary across jurisdictions.

Teachers are at the center of developing and scoring PACE tasks. After receiving training from experts at the Center for Assessment, the state's lead technical and policy partner, a core group of 60 "content lead" teachers across the three content areas and grade spans helps lead the task-development work. These content leads, who also are trained to facilitate adult learning, then work with almost 400 teachers from across PACE districts to develop the actual tasks. Teachers are developing performance tasks for a variety of high school courses, as well, but these tasks are not part of the state accountability system.

Teachers design the PACE Common Tasks using a template (based on Evidence-Centered Design principles) that asks them to first specify what students should know and be able to do based on the state's academic standards. Teachers then determine what kinds of evidence would indicate that students had mastered the intended learning targets. Finally, they design the task itself to elicit that evidence. Students with disabilities and English language learners can use the same accommodations as those on the regular, statewide tests. The tasks are not given in a specific testing window; teachers can give them when it best fits the curriculum. PACE tasks are complex, asking students to write and revise, apply math to real-world problems, or conduct science experiments.

For example, an open-ended math question might prompt students to design a budget for a school dance by using pricing models and accounting for varying levels of attendance. A final project for a geometry unit asks students to prepare a proposal for building a water tower capable of holding 45,000 + 2,000 cubic feet of water, which can be submitted to the town planning committee, complete with a cover page, models or scale drawings, calculations and mathematical strategy, and communication of the student's analysis and recommendation.

The New Hampshire Department of Education and the Center for Assessment check whether participating districts' performance tasks are aligned with state standards by collecting and reviewing local summative assessment maps that specify which standards and competencies are being measured by each task. In addition, the state collects a sample of local assessments for peer review across all participating districts to check for alignment and quality. Districts receive feedback to help them improve the tasks. The PACE Common Tasks go through a rigorous technical review by the Center for Assessment prior to use.

To further ensure reliability and comparability in scoring, all PACE districts hold grade-level

(continued)

learners and students with disabilities, individualized questions, automated scoring, and technology-enhanced performance tasks that can assess more complex skills mean the shift is here to stay. Students also like it better, notes Middlestead.

"Once you're online, there's a ton of stuff you can do with a test," says Jon Cohen, the president of AIR Assessment. "Everyone now is looking at going online."

In 2011, the State Education Technology Directors Association reported that 33 states offered some type of online testing; only five of these states required that students take the end-of-year assessment online.[15] Today, most states give their annual tests online as the default option, with others offering a choice of computer or paper-and-pencil administration. While not responsible for this shift, Smarter Balanced and PARCC helped accelerate it by offering their own tests online.

New Hampshire's Performance Assessment of Competency Education *(continued)*

calibration sessions. During these sessions, teachers bring samples of student work on the PACE Common Tasks representing the range of achievement in their classrooms. They then work together to come to a common understanding about how to score papers and identify prototypical examples of student work at different levels for each dimension of the scoring rubric. The educators annotate each of the anchor papers to document the groups' rationale. The state also audits a sample of PACE Common Tasks scored by different teachers to check for consistency in scoring.

For accountability purposes, the state employs a calibration approach using samples of student work from the Common Tasks. Calibration involves having peers evaluate student work samples from other districts to evaluate the leniency or rigor with which each district tends to rate student work. These calibration results are then used by the state to make statistical adjustments to district competency scores, if necessary, to ensure they are comparable. It then uses students' end-of-year competency scores and teacher judgments about which achievement level best describes each of their students, based on PACE achievement level descriptors that mirror those for the statewide assessment, to define performance levels across districts.

PACE began with four districts in the 2015-16 school year under a waiver from the No Child Left Behind Act's testing requirements. There are now 14 districts in full implementation (all three subjects in all relevant grades), with 16 additional districts phasing in implementation by starting with as little as one content area at one grade span. Given the current rate of growth, the state is confident that it will be able to have all New Hampshire school districts participating in PACE by the end of the demonstration period.

A formative evaluation of nine PACE districts engaged in full implementation, published by Human Resources Research Organization (HumRRO) in March 2017, found considerable evidence that PACE has had a "substantial positive impact on both teaching practice and student learning."

The evaluation included visits to schools to conduct interviews or focus groups with administrators, teachers, parents, and students; classroom observations; observations of cross-district meetings, including task-development, scoring, and calibration sessions; a review and analysis of scoring and calibration data; and a survey of district teachers.

The study found strong buy-in for PACE and a high degree of collaboration across districts. Teachers

One clear advantage is the ability to offer more accommodations that make the tests accessible to students with disabilities and to English learners, often sparing these students the stigma of "visible" accommodations. Smarter Balanced, for instance, now offers its tests in 10 languages and Unified English Braille and is working on illustration glossaries to help even more students access the tests using pictures as a thesaurus. "What we saw in PARCC, with students being able to take accommodations without leaving the classroom, was such a shift," says Slover. "Kids appreciated it; teachers liked it. Just the sense of being able to clandestinely turn on and off whatever you needed, but not have that be a public conversation."

Some states, such as Alabama, are using computer-adaptive tests, where questions get harder or easier based on a student's initial responses. The goal is

New Hampshire's Performance Assessment of Competency Education *(continued)*

feel PACE has increased their subject knowledge and given them real-time feedback they can use to make 'on-the-spot' adjustments to their instruction to better meet student needs. They also report higher student engagement and deeper learning of the content. The majority of students report that they would rather take a PACE assessment than an end-of-year comprehensive test, noting how closely the tasks were linked to their curriculum and how strong a measure of their abilities the tasks were. The study found considerable evidence that students' scores and annual determinations are accurate and reliable.

While the study found that PACE was largely working as intended, it noted that contextual factors—such as district capacity, size, and prior experience with competency-based education—can influence implementation. Small districts may struggle to have teachers develop and give performance tasks in multiple subjects, while larger districts may struggle to score all the tasks. "PACE requires a tremendous amount of work on the part of teachers," the researchers note. "While most teachers were very supportive of PACE, it was not uncommon for them to comment on the time and effort required to implement the program."

Survey results indicate about one-fourth of respondents do not think the time and effort required are worth the benefits. Figuring out how to assess district readiness and provide the right level of support and capacity building for districts and teachers as the program expands will be critical to sustainability, the report concluded.

The testing system is currently supported by a combination of state and external resources, which will continue to be required, at least in the near-term. In the 2018-19 school year, the total budget for PACE was $991,700, including $360,000 provided by the New Hampshire Learning Institute, a training and funding partner, and $364,000 that the department is raising from foundations, primarily to support the local share of project costs.

---

[1] Julia Freeland Fisher, "From Policy to Practice: How Competency-Based Education is Evolving in New Hampshire," May 2014, Clayton Christensen Institute for Disruptive Innovation: Cambridge, MA, https://www.christenseninstitute.org/publications/from-policy-to-practice/

[2] Arthur Thacker and D.E. Becker, *Formative Evaluation of New Hampshire's Performance Assessment of Competency Education (PACE): Summary Report,* March 10, 2017, HumRRO: Alexandria, VA, http://reachinghighernh.org/wp-content/uploads/2017/04/VI-A-1-Summary-PACE-Formative-Eval-Report_10-March-2017_v2.pdf

to provide a more accurate picture of what students know and can do, as well as instant scoring for use in instruction. For purposes of ESSA, the tests still have to show whether a student is on grade level for their current grade, even if some questions are below or above grade level. Companies such as AIR and NWEA have figured out how to deliver computer-adaptive testing reliably.

Online testing is also spurring the creation of a new generation of technology-enabled performance tasks that ask students to draw, write, conduct lab experiments, and solve multi-step math problems, sometimes with simulations that would not otherwise be possible. Working with about 10 states, AIR has been taking classroom activities and turning them into machine-scorable performance tasks that can be given as part of a test, says Jon Cohen. The work began in science and has expanded to mathematics tasks that are being field tested this spring. "Hopefully, we'll start using it operationally in at least a couple of states next spring on a very small scale, for just a few items in a few tests," he adds. "Being able to administer and reliably score from authentic tasks that look like instruction has been the holy grail of the testing industry and I think we're getting closer and closer to that."

These efforts are gaining the most traction in science owing to the Next Generation Science Standards. Science lends itself to such assessments, and the National Academy of Sciences' Board of Testing and Assessment explicitly recommended that states develop a comprehensive assessment system that goes beyond annual tests to adequately capture the science standards.[16] Significantly, because accountability pressure is less intense in science than in math and language arts, there is more room to innovate.

"There is much more room to maneuver there because you only have to test once per grade span [under federal law], rather than every grade level, and the stakes are a lot lower," says Michael Cohen of Achieve, "so you've got more flexibility in what you do." *(See Delaware profile.)* Achieve recently published a set of annotated performance tasks in science on its website.

Despite the interest in technology-enabled performance tasks, Marion of the Center for Assessment warns they may still constrain item design in a way that actual, teacher-scored performance tasks do not.

"One of the things that happened over the last 5 or 6 years is the field pushed testing companies to become more innovative," says Weiss. "I really worry we will take the foot off the pedal on that, and we still have a long way to go. Tests are much better than they were, but technology is moving so fast—there's so much stuff one could do to have higher-quality assessments in cheaper ways … but that takes a sector that is powered up around this and doing all the design work."

The Next Frontier: Assessments Closer to Classroom Teaching and Learning

Teachers have long complained about annual tests that reduce time for instruction, don't deliver results until the following year, and do not adequately reflect or support classroom instruction. Today, there's growing interest in creating state testing systems that better support the work of students and teachers in classrooms. So far, there's been more talk than action on this score.

ESSA's Innovative Assessment Demonstration Authority was designed to encourage states to explore new alternatives to existing end-of-year assessments. But there's no money attached. And some of the requirements have discouraged takers, including a provision demanding that states be able to scale the innovation statewide within five years and that the new measures be as rigorous, comparable, valid, reliable, and aligned to state standards as the existing state test.

So far, four pilots have been approved. New Hampshire is using performance tasks given periodically throughout the year and embedded in classroom instruction in English language arts, math, and science as a way to encourage schools to provide richer, more individualized learning for students. *(See New Hampshire profile.)* Schools and districts participating in the Granite State's

innovative assessment pilot—Performance Assessment of Competency Education, or PACE—supplant much of the traditional end-of-year summative testing with these teacher-developed assignments, including performance tasks that are shared across the districts as well as ones that each site commits to develop and use locally. The pilot has generated a lot of interest, despite concerns that the time and money required for teacher professional development, task design, and scoring may not be replicable in larger states.

Louisiana, in collaboration with NWEA, is developing ELA/social studies tests to be given three times a year at the end of units that can be rolled up into an end-of-year summative score. The online assessments will enable districts to choose from a set of texts organized around key topics. Students will respond to writing tasks that require them to make meaning of texts they've already studied or of texts that are new to them, but closely related to the curriculum. "We're trying to incent a deeper focus on making meaning as the central activity in reading comprehension," says John White, the superintendent of public instruction. *(See Louisiana profile.)* The state is building the pilot tests in ways that don't necessitate using Louisiana's curricula so that other states might follow.

Nebraska is working with NWEA to offer three, computer-adaptive tests throughout the year, as well as performance tasks that teachers would use solely for instructional purposes and not as part of the state measure of student performance. "Essentially, it's one long adaptive test across the year," says Minnich of NWEA, "where by the end of the year, we cover all the standards and we give both summative scores at year's end and a growth metric from the beginning of the year to the end of the year." The system, which is not being designed under the federal innovation pilot at this point, will begin piloting in 2020.

In some ways, it is similar to the original PARCC model which also called for a "through-grade" or "through-course" assessment system that enabled states to roll up performance across the year. But PARCC proposed including performance tasks given mid-year as part of that summative measure, while Nebraska is prioritizing getting results to teachers quickly with the computer-adaptive tests and using the performance tasks for informational purposes only. PARCC ran into challenges because the through-grade assessments would have tied up computer labs and libraries needed for other purposes, and because of the public perception that the tests would so dictate the timing for teaching knowledge and skills throughout the year that it amounted to a mandated curriculum. "Both of those problems were solvable," says Slover. "Five to 10 years later, you see all these curricula that have been created with different pathways but they cover the same standards … but then, it was too early."

Georgia lawmakers passed legislation in spring 2018 to create a state innovative assessment pilot that would give districts or consortia of districts the opportunity to develop and try innovative assessments in place of the state tests. The state board of education held two rounds of competitions and advanced three proposals, two of which were approved by USDOE. *(See Georgia profile.)* "The common denominator that we saw was that all three are through-year or interim-assessment models," says Allison Timberlake, the deputy superintendent for assessment and accountability at the Georgia Department of Education. Because the state law conflicts with federal law requiring states implement a single assessment system, Georgia applied for the innovation pilot as a work-around and received approval in 2019.

North Carolina also has received federal approval for a model that consists of interim assessments, NC Check-Ins, administered three times throughout the year and based on the same item bank as the state's end-of-grade tests. In the 2018-19 school year, the NC Check-Ins were voluntarily administered to over 50 percent of North Carolina students in grades 3-8 for mathematics and 4-8 in English Language Arts for informational purposes, in addition to the state exam. The state wants to expand upon the NC Check-Ins to develop a comprehensive assessment system that would support the use of a through-grade model as the summative assessment.

Virginia's Long History of Going Its Own Way

Virginia adopted the Standards of Learning (SOLs) in the early 1980s and developed a statewide assessment aligned to the standards soon after they were revised in 1995. The tests were first given in 1998 and were originally given in grades 3, 5, 8, and at the end of key high school courses. The state expanded the reading and math tests to grades 3-8 to meet testing requirements under the federal No Child Left Behind Act. The state also has a writing test in grade 8 and high school that is not part of the federal accountability system.

When the Common Core State Standards were being developed, Virginia was in the process of revising its math and ELA standards, which it does every seven years. To ensure students would be college and career ready Virginia sought advice from Achieve and the College Board as part of the revision process. As a result, and because districts supported continuing with the SOLs, the Virginia Board of Education made a deliberate decision not to transition to the Common Core—a move that was criticized nationally at the time. The state did not join either state testing consortia.

"The Standards of Learning are the foundation of everything that is done in our public schools, and to pull that out, the state board of education felt, would be very disruptive," said Charles Pyle, the public information officer for the state department of education. The state did review its standards in both math and English against the Common Core, he said, and made some adjustments to fill gaps.

That longevity is also reflected in its choice of test contractors. Its most recent contract with Pearson, with ETS as a subcontractor for some test development, was signed in 2005. Teachers are extensively involved in reviewing test items; by the time one appears on a test it has been reviewed by Virginia teachers at least three times.

"We have been progressive in making changes as needed, but we've been very thoughtful about making those changes," said Shelley Loving-Ryder, assistant superintendent for student assessment and ESEA Programs in the Virginia Department of Education, "and I think that has been helpful to school districts because typically, unless we are required to by legislation, we don't make rapid changes."

The state was one of the first to go to online testing (more than 99% of students now take the tests online). That enabled the state to move to computer-adaptive testing starting about six years ago, which shortened state tests by about 30 percent and optimized the tests for each student's achievement level.

Three years ago, a new online test delivery system was implemented so that students can take the tests on iPads and Chrome Books in addition to desktop and laptop computers. Virginia also was one of the first states to use technology-enhanced items and now includes them in all tested subjects, including history. Its online writing assessments are scored by one human scorer and by artificial intelligence.

While they're not part of the state assessment system, performance assessments are being developed by districts for instructional purposes. The state department is working with districts to develop common rubrics that can be used across the state and a tool to help systems evaluate the quality of performance tasks.

"We are seeing states take this notion of instructional utility more seriously and recognize that they can't do it with an end-of-year test," says Marion of the Center for Assessment. "So, a lot of states are now trying to support local districts and others in helping build better local assessments, more supportive performance assessments, whether as part of the Innovative Assessment Demonstration Authority or just part of a local initiative. We're seeing much more interest in richer forms of assessment closer to the classroom."

Massachusetts, for example, recently adopted new social studies and civics standards and is exploring what more innovative assessments might look like in those areas. Colorado is considering using more performance-based assessments as part of the state's graduation requirements, but not as the federal high school accountability tests. States such as Arkansas and Alabama are focused on increasing educators' assessment literacy. Under Alabama's $26-million Math and Science Teachers' Initiative, for example, coaches in 11 regions are working to improve teachers' classroom assessment skills.

Smarter Balanced also is exploring how it could support competency-based instruction through a system of standardized assessments that teachers could administer to certify units of knowledge or content as students are ready, rather than through a single, summative test. This is important because the absence of meaningful standards and good measures of whether students are meeting them is currently a challenge for the competency-based-learning movement. "We're working with a couple of service providers to lay out some foundational documents about equity, accessibility, accountability, and legality as part of that decision-making process," says Alpert. Smarter Balanced is also building a bank of more focused interim assessments, called Interim Assessment Blocks, that will measure more specific bundles of content than the summative tests, and that teachers can use throughout the year to inform instruction. The first set of these assessments should be available to schools in the 2019-20 school year.

According to Marion of the Center for Assessment, the biggest challenge to test quality is not annual state tests but the interim assessments that many schools and districts use: "The commercialization of interim assessments has been a big barrier because it's de-skilling the teachers with useless information." A study by the Council of Great City Schools found that in the 2014-15 school year, 401 unique tests were administered across subjects in the 66 Great City School systems, with the biggest burden at the high school level.[17]

Despite the handful of proposals approved by USDOE under the innovation pilot, most expect the risk-adverse nature of state testing directors will remain a barrier to innovation. "There needs to be R&D funding to work with states to design new innovative assessment models that can achieve these goals of bringing testing back more closely to teaching and learning while maintaining rigor and quality and consistency," says VanderVeen. During their recent legislative day on Capitol Hill, the CCSSO was recommending a competitive grants program modeled on the federal Investing in Innovation, or I3, grants.

The Future of Annual State Tests

While state testing is in flux—in part because of ESSA—no one is predicting its imminent demise. A more immediate goal might be to reduce the footprint of annual state tests to make room for more innovation at the district and classroom level.

"I think the big question will be is there such a thing as end-of-year summative assessments in 10 years?" asks Slover. "Will we need that?" She argues that if a through-course or modular assessment model can be developed with a lighter footprint, using technology to make it more feasible, "I think we will see the waning of end-of-year testing. Maybe it will never go away, but it will be much smaller and less accountability-focused and more focused on tracking progress."

VanderVeen argues that for accountability purposes, states are still going to need annual state tests. "Can you shorten it and get more information from through-course stuff that feeds information back into the classroom?" he asks. "I think you can do more of that, but I don't think you're going to get rid of the end-of-year summative."

"There's a lot more we could do with summative assessment," agrees Marion. "But first we have to get off this insanity of testing every kid, every item, every standard, every year … If you could reduce the footprint of the state summative assessment and temper some of the accountability demands, you could make space for more innovative stuff at the district level."

One challenge, according to Lillian Pace, vice president of policy and advocacy for Knowledge Works, is to get the civil rights community on board. And that, she says, will require a shift away from the mindset "that the only way to drive equity in this country is to have an annual statewide summative assessment that looks the same for every student." Such tests, she argues, have not served historically underserved students well.

But Michael Cohen, who experienced the fight over the Common Core standards and aligned assessments firsthand, feels differently, "I don't think the time is right, in general, for a major effort to create better tests because nobody wants to talk about tests. People are tired of standards, tests, and accountability. They just don't want to deal with it anymore."

ENDNOTES

1 "DeVos Pushes State Chiefs to Consider ESSA Waivers, Particularly in Testing, Saying Anything that Helps Student Achievement Is Going to Be Seen Very Favorably," April 8, 2019, https://www.the74million.org/article/devos-pushes-state-chiefs-to-consider-essa-waivers-particularly-in-testing-saying-anything-that-helps-student-achievement-is-going-to-be-seen-very-favorably/

2 Jill Mullan and Julie Rowan Woods, "50-State Comparison: State Summative Assessments," April 11, 2018, Education Commission of the States: Denver, https://www.ecs.org/50-state-comparison-state-summative-assessments/

3 "What Gets Tested Gets Taught: Cautions for Using College Admissions Tests in State Accountability Systems," March 2018, Achieve: Washington DC, https://www.achieve.org/college-admissions-tests-accountability

4 *Ibid*.

5 *Ibid*.

6 *Ibid*.

7 *Ibid*.

8 Khan Academy, "Studying for the SAT for 20 Hours on Khan Academy Associated with 115-Point Average Score Increase," May 8, 2017, https://www.khanacademy.org/about/blog/post/160451329150/studying-for-the-sat-for-20-hours-on-khan-academy

9 Alyson Klein, "DeVos' Team: Arizona Could Lose $340 Million for Skirting ESSA's Testing Requirements," *Education Week*, April 9, 2019, https://blogs.edweek.org/edweek/campaign-k-12/2019/04/betsy-devos-arizona-essa-testing-act-sat-high-school.html?cmp=eml-enl-et-news1&M=58805476&U=2661408&UUID=145de2ceb135c741621f8b21d6e4869d

10 V. Bandeira de Mello, Taslima Rahman, and B.J. Park, *Mapping State Proficiency Standards onto NAEP Scales: Results From the 2015 NAEP Reading and Mathematics Assessments*, May 31, 2018, National Center for Education Statistics: Washington DC, https://nces.ed.gov/nationsreportcard/pubs/studies/2018159.aspx

11 Daniel Hamlin and Paul E. Peterson, "Have States Maintained High Expectations for Student Performance? An Analysis of 2017 State Proficiency Standards," *EducationNext*, Fall 2018, vol. 18, no. 4, https://www.educationnext.org/have-states-maintained-high-expectations-student-performance-analysis-2017-proficiency-standards/

12 *States' Commitment to High-Quality Assessments Aligned to College- and Career-Readiness*, October 14, 2014, CCSSO: Washington, DC, https://ccsso.org/topics/high-quality-assessment-systems

13 *Criteria for Procuring and Evaluating High-Quality Assessments*, March 2014, CCSSO: Washington, DC, https://ccsso.org/resource-library/criteria-procuring-and-evaluating-high-quality-assessments

14 Education First, *Whatever Happened to All those New and Better State Tests? The State of State Assessments*, February 2018, Education First: Seattle, WA, https://education-first.com/wp-content/uploads/2018/02/Education-First-What-Happened-To-State-Tests-Feb-2018-6.pdf

15 Douglas Levin, Geoffrey Fletcher, and Yen Chau, "Technology Requirements for Large Scale Computer-Based and Online Assessment: Current Status and Issues," June 22, 2011, State Education Technology Directors Association: Glen Burnie, MD, https://www.setda.org/wp-content/uploads/2014/03/large-scale_techrequirements_june22_combined.pdf

16 National Academy of Sciences' Board on Testing and Assessment, *Developing Assessments for the Next Generation Science Standards*, 2014, National Academy of Sciences: Washington, DC, http://sites.nationalacademies.org/DBASSE/BOTA/Developing_Assessments_for_NGSS/index.htm

17 Ray Hart, Michael Casserly, Renata Uzell, Moses Palacios, Amanda Corcoran, and Liz Spurgeon, *Student Testing in America's Great City Schools: An Inventory and Preliminary Analysis*, October 2015, Council of Great City Schools: Washington, DC, https://www.cgcs.org/cms/lib/DC00001581/Centricity/Domain/87/Testing%20Report.pdf

APPENDIX

State K-12 Testing Systems

	Name of Test, 3-8	Name of Test, High School	Testing Contractor, 3-8	Testing Contractor, High School	Previous PARCC or Smarter Balanced
ALABAMA	Scantron Testing (2019) ACAP Summative (2020)	ACT with Writing (11th grade), ACT WorkKeys (12th grade), PreACT (10th grade)	Data Recognition Company (DRC) ACAP Summative	ACT	PARCC (pulled out in 2013)
ALASKA	Performance Evaluation for Alaska's Schools (PEAKS)	PEAKS (only grade 9-10)- (only science)	DRC	DRC	Neither
ARIZONA	AzM2, AIMS Science	AzM2, AIMS Science	Pearson for Science	Pearson for Science	PARCC (dropped out in 2014)
ARKANSAS	ACT Aspire	ACT Aspire (grades 9 & 10) ACT (grade 11) Not required for students to take but required for schools to offer	ACT	ACT	PARCC (2014-2015)
CALIFORNIA	Smarter Balanced	Smarter Balanced (grade 11 only)	Smarter Balanced, ETS	Smarter Balanced, ETS	Smarter Balanced
COLORADO	CMAS ELA and Math: Grades 3-8. Science: Grades 5, 8 & 11 Social studies: About 1/3 of schools will test grades 4, 7 & 11	PSAT, SAT	Pearson	Pearson for Science CMAS test College Board for PSAT and SAT	PARCC (largely dropped in 2017)
CONNECTICUT	Smarter Balanced	NGSS (grade 11-science only) SAT (grade 11-reading/writing, math)	Smarter Balanced	College Board	Smarter Balanced
DELAWARE	Smarter Balanced (also called DeSSA)	PSAT10 is required by the state, but not for accountability. SAT is given once in high school, typically grade 11	Smarter Balanced AIR (testing vendor)	College Board	PARCC (dropped out in 2014)
DC	PARCC	PARCC SAT (11th grade) DC covers the cost but this isn't a requirement for ESSA	New Meridian for content, Pearson for administration	New Meridian for content, Pearson for administration	PARCC

State K-12 Testing Systems | 1

	When Test Is Given	When Results Are Returned	How Results Are Shared With Parents, Students	Paper or Online Administration
AL	ACAP Summative: April ACT: March	ACAP Summative: Students will receive a printed report (timeframe tbd) ACT: Students will receive printed scores in the mail 3-8 weeks after exam; scores online about 1 week after printed scores	ACAP Summative: Students receive a printed report (timeframe tbd) ACT: Student receive printed scores in the mail 3-8 weeks after exam; scores will available online about 1 week after printed scores are received	ACAP Summative: online ACT: paper or online
AK	March-April	August-September of following school year	Depends on the school	Both
AZ	April	Electronic results to schools by end of May. Paper reports to districts by the end of June	Determined by the district/charter & the school	Both
AR	April-May	Results are returned to schools on July 1	District receive individual student reports that are sent home to families	Online
CA	Final third of the year LEAs may select their own testing window	Districts and schools monitor communication from the California Technical Assistance Center and/or the California Department of Education for information	Student scores to family within 20 days of receipt by the district, if the district receives reports after the end of the school year, families receive within 20 working days of the next school year	Online (computer adaptive)
CO	April	Results to school districts in June	Released publicly by school districts in August. Districts send out score reports to parents	Both
CT	March-June	Preliminary: June Final: August (for Smarter Balanced), depends on the school	Two paper copies of individual student reports are shipped to the districts (one for the parents one for the district)	Online (computer adaptive)
DE	Smarter Balanced: March-May SAT: March-April with a make up option	Smarter Balanced: Results to educators within 10 days of testing. SAT: Mid May-Educators receive them via the College Portal a week or two after students and parents	Smarter Balanced: Parents receive results when embargo lifts in August SAT: Online via College Board portal 2 weeks after the assessment is administered. A paper report is also mailed to parents in August	Smarter Balanced: Online (computer adaptive) SAT: Paper (some districts are piloting PSAT 8/9 online)
DC	PARCC: April-May (grades 3-8); Students must take one ELA and math assessment in high school, aligned with grade 10 testing	August	Schools receive individual student reports and then distribute them to families, results can also be viewed online	PARCC: online, paper for accommodations

State K-12 Testing Systems (continued)

	Name of Test, 3-8	Name of Test, High School	Testing Contractor, 3-8	Testing Contractor, High School	Previous PARCC or Smarter Balanced
FLORIDA	Florida Standardized Assessment (FSA) (grades 3-8), Florida Statewide Science Assessment	FSA and FSA end of course tests (Algebra, Geometry) (FSA ELA grades 9-10)	AIR for grades 3-8 ELA and Mathematics; Pearson for grades 5 & 8 Science	AIR for grades 9-10 ELA, Algebra, and Geometry; Pearson for Biology	PARCC (dropped in 2013)
GEORGIA	Georgia Milestones	Georgia Milestones	DRC	DRC	PARCC (Dropped in 2013)
HAWAII	Smarter Balanced	Smarter Balanced	Smarter Balanced AIR (testing vendor)	Smarter Balanced (consortia) AIR (testing vendor)	Smarter Balanced
IDAHO	The Idaho Standard Achievement Test (ISAT) developed by Smarter Balanced	Grade 10: ISAT Grade 11: State pays for the SAT for all students (a graduation requirement but not used for federal accountability)	AIR for test administration scoring and reporting, Smarter Balanced for test design/ development	AIR for test administration scoring and reporting, Smarter Balanced for test design/ development	Smarter Balanced
ILLINOIS	Illinois Assessment of Readiness exam (IAR)	SAT under ESSA, contemplating an amendment to add PSAT 8/9 and PSAT 10	Pearson	College Board	PARCC (dropped in 2019)
INDIANA	Indiana Learning Evaluation Readiness Network (ILEARN)	ISTEP+ Math & ELA; ILEARN Biology End of Course Assessment (Science)	AIR	AIR	PARCC (dropped in 2014)
IOWA	Iowa Statewide Assessment of Student Progress (ISASP)	ISASP	Iowa Testing Program within the University of Iowa	Iowa Testing Program within the University of Iowa	Smarter Balanced (adopted but never implemented)

State K-12 Testing Systems | 2

	When Test Is Given	When Results Are Returned	How Results Are Shared With Parents, Students	Paper or Online Administration
FL	FSA: April-May End of Course Assessments: Sept, Dec, May, July	Scores to schools in June, schools release to parents a few weeks later	Hard-copy individual score reports are provided by the state to school districts, districts distribute reports to parents. Electronic copies can be uploaded to parent portals	Paper for ELA reading and math in grade 3, ELA writing and reading and math in grades 4-6, and students with accommodations, all other assessments are computer-based.
GA	End of course work: November-January April-June June-July	July	Scores are released by state to school then by school to parents	Online administration as primary, paper-pencil as back-up for students with accommodations
HI	Smarter Balanced: February-June ACT: February	Smarter Balanced: Results to educators/administrators within 10 days of testing. SAT: Mid May-Educators and administrators receive them via the College Portal about a week or two after students and parents	Smarter Balanced: The state sends a printed "Family Report" to schools who distribute to parents ACT: Students receive printed scores 3-8 weeks after exam	Smarter Balanced: Online (computer adaptive) ACT: Online or paper
ID	ISAT: March-May SAT: April of 11th grade	ISAT: Student results to schools within 10 days of testing, summary data on a rolling basis Final results released in August-September SAT: Mid May for April test	ISAT: See above to ISAT: August-September SAT: Mid May if test is taken in April; tests returned online to students in College Board Portal	ISAT: Online (computer adaptive) or paper for accomodations only SAT: Paper
IL	IAR: March-April SAT/PSAT: April	Dates not finalized	IAR: A paper score for each student is sent to the school, which distributes to parents PSAT/SAT: Released to students and parents via mail and electronically	IAR: Online (state has granted a 1 year exception for paper if needed) SAT: Online or paper
IN	ILEARN Grades 3-8: April-May ILEARN Biology 1st Semester: December ILEARN Biology Trimester: February ILEARN Biology 2nd Semester: April-May 22 ISTEP+ Part 1: February-March Part 2: April–May	ILEARN 3-8 & Biology: To schools 12 days after students complete the online test. ISTEP+ High School: To schools by end of June	Schools are required to communicate individual student test results to parents and to manage the subsequent rescore request process locally	Online (paper for students with accommodations)
IA	March-May	Districts receive data as soon as possible after test completion.		Online and paper

State K-12 Testing Systems (continued)

	Name of Test, 3-8	Name of Test, High School	Testing Contractor, 3-8	Testing Contractor, High School	Previous PARCC or Smarter Balanced
KANSAS	Kansas Assessment Program (KAP)	KAP	Achievement and Assessment Institute at University of Kansas	Achievement and Assessment Institute at University of Kansas	Smarter Balanced (dropped before first testing)
KENTUCKY	Kentucky Performance Rating for Educational Progress (K-PREP)	K-PREP, ACT	Pearson	Pearson, ACT	PARCC (dropped in 2014)
LOUISIANA	Louisiana Educational Assessment Program (LEAP 2025)	LEAP End-of-course tests, ACT	State, DRC for ELA & Math, WestEd for science and social studies content	DRC for ELA & Math, WestEd for science and social studies content, ACT	PARCC
MAINE	Maine Educational Assessment (MEA) Grades 3-8 MEA math and ELA/literacy (eMPowerME) Grades 5 & 8 MEA science	High School math and ELA/literacy (SAT) – used for accountability	Cognia for eMPowerME, Science, MSAA DRC	Cognia for eMPowerME, science, MSAA DRC	Smarter Balanced (dropped in 2015)
MARYLAND	Maryland Comprehensive Assessment Program (MCAP) covers ELA, math, science & social studies	MCAP	ETS for content, Pearson for administration	ETS for content, Pearson for administration	PARCC (dropped after 2018-19)
MASSACHUSETTS	Massachusetts Comprehensive Assessment System (MCAS)	MCAS	Cognia	Cognia	Some PARCC items used in 2017 and 2018 for grades 3-8 ELA and math only
MICHIGAN	Smarter Balanced and Michigan Student Test of Educational Progress (M-STEP)	PSAT, SAT, ACT WorkKeys	Design: Mix of state and Smarter Balanced content, DRC for support Administration and scoring: DRC for online, Measurement Inc. for paper	Design: Mix of state and Smarter Balanced content, DRC for support Administration and scoring: DRC for online, Measurement Inc. for paper College Board, ACT	Smarter Balanced

State K-12 Testing Systems | 3

	When Test Is Given	When Results Are Returned	How Results Are Shared With Parents, Students	Paper or Online Administration
KS	March-April	About 2 weeks after the testing window closes	Individual student reports are made available electronically to the districts/schools who then disseminate to parents	Online
KY	Within the last 14 instructional days of a district's calendar. Online field tests occur January-March	ACT: July, State results: September, individual scores: by October	ACT results are sent directly to the students, state results through a School Report Card, individual results are printed and sent to parents	2019-2020 SY: All high school assessements are online, half of the 3-8 exams are online half are paper and pencil. By SY 2020-2021, all exams will be online and paper & pencil for those with accommodations
LA	March-May	By the end of June under legislative mandate	The state requests that a Parent Guide to the LEAP 2025 Student Reports be sent home with the results	Online or paper (two testing window time options)
ME	WIDA ACCESS: January-February SAT: April Science High School: March-April Science 5 & 8: April-May	Results to districts/schools to verify September	Test results to public and for parent dissemination, October. Results distribution process is district/school/local decision	MEA: Online (paper for kindergarten only) SAT: Online or paper
MD	High School: December-January	November for new test results in 2020, early in the school year thereafter	Fall	Online (paper for students with accomodations only)
MA	March-June	Preliminary results: within a few weeks of test completion	Final results including to parents: end of September	Online tests with paper for students with accommodations
MI	April-May	Before Labor Day each year. Data is available to schools earlier for review before public release of data	The state prepares an individual student report, which is sent to schools to be disseminated to parents	Paper or online

State K-12 Testing Systems (continued)

	Name of Test, 3-8	Name of Test, High School	Testing Contractor, 3-8	Testing Contractor, High School	Previous PARCC or Smarter Balanced
MINNESOTA	Minnesota Comprehensive Assessments (MCA)	MCA	Pearson	Pearson	Neither
MISSISSIPPI	Mississippi Assessment Program (MAAP) ELA and Math (Grades 3-8), Science (Grades 5 and 8)	MAAP	Questar (ELA and math), DRC for science	Questar (English II and algebra I), DRC (biology and U.S. history)	PARCC (2014-2015)
MISSOURI	Missouri Assessment Program (MAP) Grade-Level Assessments	MAP End-of-course tests, ACT	DRC	Questar	Smarter Balanced (2014-15)
MONTANA	Smarter Balanced	ACT (grade 11)	Smarter Balanced	ACT	Smarter Balanced
NEBRASKA	Nebraska Student-Centered Assessment System (NSCAS)	ACT	NWEA for NSCAS-General	ACT	Neither
NEVADA	Smarter Balanced, end-of-course tests	End-of-course tests, ACT	DRC	DRC	Smarter Balanced
NEW HAMPSHIRE	New Hampshire Statewide Assessment System (NHSAS)	SAT (grade 11), NHSAS for science	AIR	SAT, AIR	Smarter Balanced (dropped in 2019)
NEW JERSEY	New Jersey Student Learning Assessments (NJSLA)	NJSLA	New Meridian for content development, Pearson for administration	New Meridian for content development, Pearson for administration	PARCC (dropped in 2019)
NEW MEXICO	ELA & Math test out for bid, New Mexico Assessment of Science Readiness (NMASR) for grades 5 & 8	College entrance exam out for bid, NMASR for grade 11	Out for bid, Cognia for Science	Out for bid, Cognia for Science	PARCC (dropped in 2019)
NEW YORK	New York State Assessments	Regents Exams	Questar	End of course "in house development"	PARCC (pulled out before implementing test)
NORTH CAROLINA	End-of-grade tests	End-of-course tests, ACT within ESSA but not as an academic indicator	North Carolina State University	NC State, Questar for	Smarter Balanced (pulled out)

State K-12 Testing Systems | 4

	When Test Is Given	When Results Are Returned	How Results Are Shared With Parents, Students	Paper or Online Administration
MN	Early May	August	Schools issue Parent/Guardian reports each fall, the state provides a webpage for accessing scores	Online with paper for students with accommodations
MS	MAAP: Two times per year: November-December (English II and Algebra I, Biology and U.S. History) and April-May (all exams)	End of July	Districts provide one copy of student report to the parent/legal guardian	Online except for students with accommodations)
MO	MAP: April-May	Summer, within 15 days of districts receiving them	District's discretion	MAP: Online except for students with accommodations
	NHSAS: Available online to school district staff shortly after test SAT: May	State recommends that schools send results to parents by the end of the school year	Smarter Balanced: Online (computer adaptive) ACT: Paper or online	Smarter Balanced: Online (computer adaptive) ACT: Paper or online
NE	NSCAS: March- ACT: March-April	NSCAS: Within 48 hours of test completion. Individual student reports: Sent home in a "timely manner", ACT: District and parents 6-8 weeks from test completion	Paper results are sent to the students	NHSAS: Online ACT: Paper or online
NV	Smarter Balanced: February-May ACT: March-April	Late Spring		Smarter Balanced: Online (computer adaptive) ACT: Paper or online
NH	NHSAS: March-June SAT: March-April	NHSAS: Available online to school district staff shortly after test SAT: May	State recommends that schools send results to parents by the end of the school year	NHSAS: Online SAT: Paper or online
NJ	High schools: November-January and April-June, General administration: April-May	Summative results to districts by mid August, individual student results by early September		Mainly computer-based exams. There are only a few instances where a paper exam is used
NM	April-May	Grade 3-8 scores: July, Science scores: August, College Entrance: TBD	Parents receive individual reports on onine and in the mail typically at the beginning of the new school year	Online except for students with accommodations
NY	NYSA: March-May Regents: 3 times per year (Jan, June and August), but not all subjects are offered on each date	NYSA: Late summer; score reports must be sent to parents by school district. Regents Exams: Immediately. Results must be reported on students' permanent records	Discretion of schools, some mail test reports while others make them available online (e.g. parent portal)	Online (NYSA only) or paper (NYSA and Regents)
NC	High School: December and May-June Grades 3-8: May (final 10 days of the school year)	Results are available to schools the day after testing	Individual student reports are sent to families within 30 days of testing	Online except in cases of technology hardship or students with accommodations

State K-12 Testing Systems (continued)

	Name of Test, 3-8	Name of Test, High School	Testing Contractor, 3-8	Testing Contractor, High School	Previous PARCC or Smarter Balanced
NORTH DAKOTA	North Dakota State Assessment (NDSA)	NDSA (grade 10) or ACT (grade 11)	AIR	AIR (grade 10) or ACT (grade 11)	PARCC (dropped in 2013)
OHIO	Ohio State Tests (OST), given by grade level and subject	Ohio State Tests End-of-course tests, SAT/ ACT	AIR	AIR	PARCC (2015 only)
OKLAHOMA	Oklahoma School Testing Program (OSTP)	ACT, SAT	Cognia	SAT or ACT	PARCC (pulled out in 2013)
OREGON	Smarter Balanced	Smarter Balanced	Smarter Balanced	Smarter Balanced	Smarter Balanced
PENNSYLVANIA	Pennsylvania System of School Assessment (PSSA)	Keystone Exams	DRC	DRC	PARCC (pulled out in 2013)
RHODE ISLAND	Rhode Island Comprehensive Assessment System (RICAS) for ELA and math	PSAT in grade 10, SAT in grade 11	Rhode Island contracts with Massachusetts to use its MCAS assessment	SAT	PARCC (dropped in 2017)
SOUTH CAROLINA	SC Ready and SCPASS	End-of-course tests, ACT/SAT	DRC	DRC for end of course tests, ACT & College Board	Plans to use Smarter Balanced prohibited by state legislature
SOUTH DAKOTA	Smarter Balanced	Smarter Balanced (grade 11)	Smarter Balanced	Smarter Balanced	Smarter Balanced
TENNESSEE	TCAP Achievement	TCAP, end of course, ACT/SAT	Pearson	Pearson	PARCC (Dropped in 2014)

State K-12 Testing Systems | 5

	When Test Is Given	When Results Are Returned	How Results Are Shared With Parents, Students	Paper or Online Administration
ND	March-May	May	Mail or online	Online only (no paper accomodations except Braille and print-on-demand)
OH	OST: 2 times per year (October-November, March-May) End of course tests: End of each semester ACT: Paper - February-April, Online: February-March SAT: April	Required to return results within 45 days of the testing window, printed family reports are sent 30-45 days after that	Printed family reports	Online (paper testing only an option for 3rd grade or for districts that meet an exception)
OK	OSTP: April-May ACT: April SAT: April	Preliminary results in May	Parents receive preliminary results in May, written results in the fall	Paper for 3rd grade, online for other grades, by 2021, all will be online, except students with accommodations
OR	Schools pick testing window between January-June	Preliminary results are available to districts within 10 days after submission, final results are published in mid-September	Parents receive individual results within a year after test administration	Online (computer adaptive)
PA	PSSA:April-May Keystone: Offered 4 times per year (2 winter waves in December and January, 1 spring wave in May and 1 summer wave in July/August)	May-July	Individual student reports are mailed to schools in September. Schools distribute to parents	Online or paper
RI	April-May	Fall of the next school year	Families will receive paper copies of individual student score report that contain information about their student's results	Online or paper
SC	SC Ready: Last 20 days of district calendar. ACT: February SAT: March-April End of Course Assessments: December-January, May-June, June-July	Districts results SCPASS: July SC READY: July EOCEP: June	Districts disseminate individual student reports as soon as they are received. Summary district data is embargoed until early September	SC READY and SCPASS: Online except for students with accommodations and districts with waivers ACT: Paper or online SAT: Paper only End of course exams: Online with the exceptions listed above
SD	March-May	Smarter Balanced: Shortly after close of the testing window		Online (computer adaptive) or paper
TN	Achievement: April-May End of course: November-December, April-May	Raw scores to districts in May; printed individual student reports and class/school/district reporting in mid-summer		Paper

State K-12 Testing Systems (continued)

	Name of Test, 3-8	Name of Test, High School	Testing Contractor, 3-8	Testing Contractor, High School	Previous PARCC or Smarter Balanced
TEXAS	State of Texas Assessments of Academic Readiness (STAAR)	STAAR	ETS	ETS	Neither
UTAH	Readiness Improvement Success Empowerment (RISE)	Utah Aspire Plus (grades 9-10), ACT (grade 11)	AIR	Pearson for grades 9-10, ACT for grade 11	Smarter Balanced (Pulled out in 2012 before testing began)
VERMONT	Smarter Balanced, Vermont Science Assessment (VTSA)	Smarter Balanced (grade 9), VTSA (grade 11)	Smarter Balanced for item development, AIR for test administration and reporting	AIR	Smarter Balanced
VIRGINIA	Virginia Standards of Learning (SOL) Tests	End-of-course SOL Tests	Pearson	Pearson	Neither
WASHINGTON	Smarter Balanced for ELA and Math, Washington Comprehensive Assessment for science (WCAS)	Smarter Balanced for ELA and math; WCAS for science	AIR for the general assessment	AIR	Smarter Balanced
WEST VIRGINIA	West Virginia General Summative Assessment	SAT	AIR	College Board	Smarter Balanced (2015-2017)
WISCONSIN	Wisconsin Forward	ACT Aspire (grades 9-10), ACT with writing (grade 11)	DRC	ACT, DRC for grade 10 Social Studies	Smarter Balanced (2014-15)
WYOMING	Wyoming Test of Proficiency and Progress (WY-TOPP)	WY-TOPP (grades 9-10), ACT (grade 11 for readiness only)	AIR	AIR and ACT	Smarter Balanced (withdrew from the consortium in 2016)

	When Test Is Given	When Results Are Returned	How Results Are Shared With Parents, Students	Paper or Online Administration
TX	High School Course Assessments: December and May Grade 3-8: April-May		Student portal	Online and paper
UT	RISE: For courses that end before March, November-March. For full-year courses, March-June Aspire: March-May ACT: March-April	RISE: 24 hours after students complete a test Utah Aspire plus: August, ACT: August	Districts share results with parents within 3 weeks of availability either digitally or in person	RISE: Online (computer adaptive) Utah Aspire Plus: Online (fixed form) ACT: Online or paper
VT	Smarter Balanced: March-June VTSA: April-June	As soon as quality control is completed, generally before the start of the next school year	Schools are required to forward individual student results to parents, but they can determine the timing and format	Online (computer adaptive)
VA	Grades 3-8: April-June end of course/ High school: October-December, Spring: April-June, Summer: July-September	Test results to districts on a rolling basis throughout the test administration. Districts notify parents and guardians in a timely manner	The state provides printed copies to all districts, who decide whether to share digitally or on paper	Online for grades 3-8 and grades 9-12 (computer adaptive for grades 3-8 reading and math only); Paper for students with accommodations
WA	Interim: Offered the entire year at the school's discretion Summative: Last 12 weeks of district calendar	Preliminary individual results to schools within 3 weeks of the student taking the test, aggregate results are reported by September 10 each year	Each school/district determines how families will receive the individual student report (e.g., mail, parent/teacher conference, or electronically)	General Assessment: Online; paper testing to support large print, braille, and standard print forms for students with accommodations
WV	WVGSA: April -May SAT: 1 date and 1 make up date from College Board's SAT test dates	WVGSA Grades 3-8 results on a rolling basis about 10 days after students complete testing, SAT: June on line, printed report in August, individual student reports by the end of September	Printed student reports are provided to districts and schools in August for distribution at the opening of school. Individual student reports distributed by the end of September	WVGSA: Online (paper for accommodations) SAT: Paper
WI	Wisconsin Forward: March-May ACT: March-April ACT Aspire: April-May	Results are returned 3-8 weeks after testing	Online copies of individual student reports must be available to parents/guardians or paper copies can be mailed	Wisconsin Forward and ACT Aspire: online ACT: paper/pencil
WY	3 types of WY-TOPP tests: Modular (Optional): available all school year Interim (Optional): September, January-February, April-May , ACT: March-May	WY-TOPP: normally by August 1, ACT: normally by August 1	Teachers receive results in the online reporting system and schools receive individual student reports to be mailed home to parents	WY-TOPP: Online (computer adaptive), paper test for students with accommodations ACT: Online or paper

FutureEd
GEORGETOWN UNIVERSITY

THE NEW TESTING LANDSCAPE

HOW STATE ASSESSMENTS ARE CHANGING UNDER THE FEDERAL EVERY STUDENT SUCCEEDS ACT