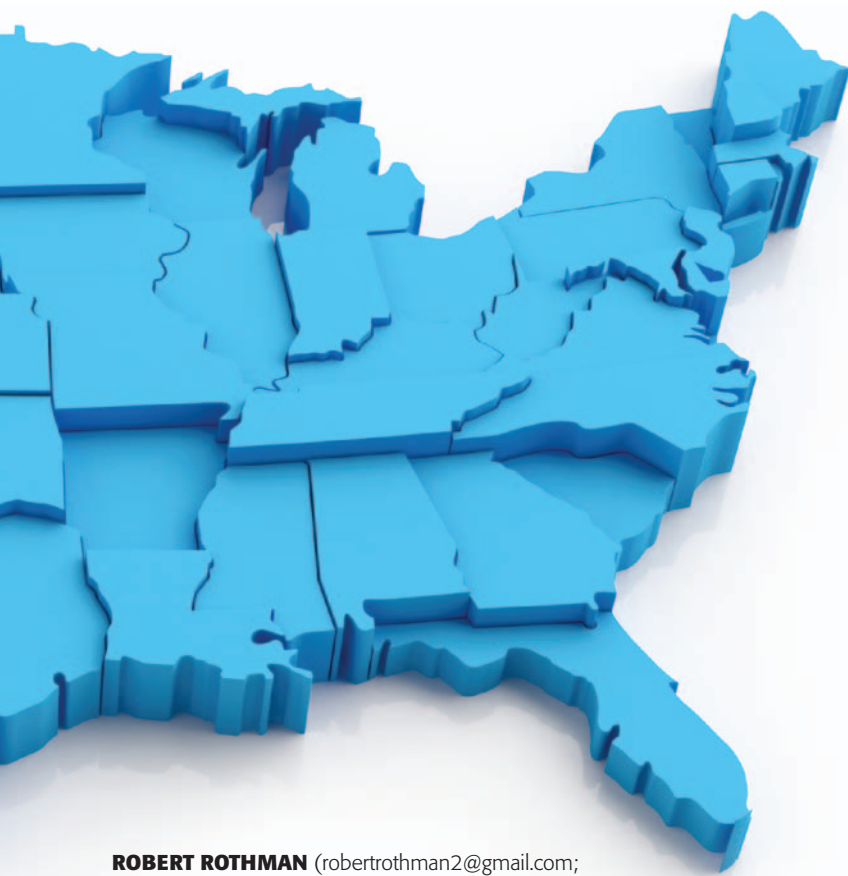


The next generation of state assessment and accountability

The Every Student Succeeds Act enables states and districts to seize the opportunity to develop assessments that go beyond paper-and-pencil bubble sheets. What's involved in stepping up to that challenge?

By Robert Rothman and Scott F. Marion



ROBERT ROTHMAN (robertrothman2@gmail.com; @RothmanR) is a Washington, D.C.-based education writer.
SCOTT F. MARION (smarion@ncea.org; @ScottFMarion) is executive director of the National Center for the Improvement of Educational Assessment, Dover, N.H.

Assessment has long had a prominent — and controversial — role in American education, and that's never been more true than today. But a new federal law gives states and districts an opportunity to move assessment forward.

Since the passage of the No Child Left Behind Act 15 years ago, state tests have played an outsized role in schools. The goal of the law was to create incentives for educators to focus on student achievement, but the law defined tests as the sole measure of outcomes. As a result, schools faced enormous pressures to raise test scores, which shaped classroom practice as teachers focused on the material in the tests and spent time on test preparation activities.

The heavy emphasis on end-of-year summative tests sparked a backlash. In recent years teachers have raised strong objections to using tests to evaluate teaching practice. Numerous parents in states like New York and Colorado “opted” their children out of taking certain standardized tests. Parents and educators alike decried “over-testing,” concerned that tests and test preparation were taking up too much classroom time. The chorus grew so loud that the U.S. Department of Education in 2015 proposed a cap — 2% of instructional time — on testing and encouraged states and districts to review their testing programs with an eye toward reducing test requirements.

The Every Student Success Act (ESSA), the successor to NCLB enacted in December 2015, contains several provisions that could address concerns about testing. The law allows states to use measures in addition to tests in school accountability systems, which could reduce pressure on schools to raise test scores at the expense of other actions that could improve learning.

Perhaps more importantly, the law authorizes a pilot program that would allow up to seven states to develop innovative assessment and accountability systems that could incorporate new measures of student performance. These new measures could support student learning rather than detract from it, as critics charge current state tests do.

Innovations in assessment

“New” assessments are not really new. States and districts have experimented with alternatives to traditional standardized tests for decades, with mixed results.

Perhaps the most active period for assessment innovation was in the late 1980s and early 1990s, when a number of states put in place new performance-based assessments and other models. These assessments were aimed at tapping a broader range of student abilities than conventional tests and at creating incentives for teachers to focus on abilities like critical thinking and problem solving.

For example, Vermont and Kentucky introduced portfolios to assess students on the basis of classwork over time; Maryland, Wyoming, and Connecticut developed

tests that asked students to engage in complex projects; and a number of states added open-ended items and extended writing prompts to their assessments.

While these assessments produced some improvements in student learning, they also encountered significant challenges. Specifically, many assessment systems were challenged on technical quality issues; some were not feasible or affordable on a large scale, and some faced political opposition from critics who considered them too subjective or not rigorous enough. In the face of these challenges and NCLB requirements that mandated tests at every grade level — from grade 3 through 8 — states dropped these alternatives and reverted to more conventional tests.

In recent years, amid growing concern about the limitations of these tests, there has been a renewed interest in alternatives. The two state consortia that developed assessments designed to measure progress against the Common Core State Standards — the Partnership for Assessment of Readiness for College and Careers (PARCC) and the Smarter Balanced Assessment Consortium — have put some innovative features in assessments. For example, both are administered on computers, which enables them to use innovative items that tap higher order thinking skills rather than having students choose among pre-selected responses. In addition, the assessments include extended performance tasks that ask students to read several passages (in English language arts) and write extended essays that draw evidence from them.

At the same time, the Stanford Center on Assessment, Learning, and Equity (SCALE) and the Council of Chief State School Officers Innovative Lab Network (ILN) have developed a task bank that collects and curates performance tasks that eventually will be available nationwide. States will have to determine how to use the items, though, and many intend to make them available for classroom use rather than incorporate them into statewide assessments.

ESSA demonstration

ESSA potentially could expand the use of innovative assessments dramatically. Although the ESSA pilot program is limited to seven states, the program could expand to other states after three years if it is successful.

The statute describes the types of assessments that could be used in the ESSA pilot:

- Competency-based assessments, instructionally embedded assessments, interim assessments, cumulative year-end assessments, or performance-based assessments that combine into an annual summative determination for a student, which may be administered through computer adaptive assessments; and

- Assessments that validate when students are ready to demonstrate mastery or proficiency and allow for differentiated student support based on individual learning needs.

The assessments must be used for accountability and must be able to provide annual determinations of academic performance for every student. The assessments must generate results that are valid, reliable, and comparable for all students and all subgroups of students, and be developed in collaboration with a broad range of stakeholders, including parents, educators, and civil rights organizations.

States in the pilot must report annually how they will ensure that all students receive the instructional support to meet state standards, provide the necessary technological infrastructure for the assessment, and hold schools and districts accountable for results.

Initially, states can implement the pilot in a subset of districts demographically comparable to the state as a whole, but the state must have a plan for scaling up the pilot to the entire state.

The New Hampshire model

Even before the pilot launches, New Hampshire provides a model of what this next phase of assessment might look like. In March 2015, the U.S. Department of Education gave New Hampshire permission to pilot an accountability system designed to support deeper learning for students and powerful organization change for schools and districts. The Performance Assessment of Competency Education (PACE) is a competency-based educational approach designed to ensure that students have meaningful opportunities to achieve critical knowledge and skills. Richard Elmore's concept of reciprocal accountability is at the core of New Hampshire's model:

For every increment of performance I demand from you, I have an equal responsibility to provide you with the capacity to meet that expectation. Likewise, for every investment you make in my skill and knowledge, I have a reciprocal responsibility to demonstrate some new increment in performance (Elmore, 2002, p.5).

For PACE, this means local educational leaders are involved in designing and implementing the assessment and accountability systems and receive intense technical, policy, and practical support and guidance from the state department of education.

PACE also is designed to capitalize on the latest advances in understanding how adults and children learn. The goal is to structure learning opportunities that let students grapple with meaningful knowledge and skills, represented as competency statements, at a depth of understanding to facilitate

The law authorizes a pilot program that would allow up to seven states to develop innovative assessment systems that could incorporate new measures of student performance.

transfer of learning to new real-world situations. PACE attempts to foster organizational learning and change by appealing to the intrinsic motivation of adults to improve their work rather than relying on top-down accountability and compliance strategies.

Assessments

The core of the PACE assessment system is locally developed, locally administered performance assessments tied to grade and course competencies determined by local school districts. Additionally, in each grade and subject without a state assessment (a total of 17 subject and grade combinations), there is one common, complex performance task administered by all participating districts. But this common assessment is not a state test. Rather, participating districts develop it collaboratively and use it to ensure that each teacher's evaluation of student performance is comparable to evaluations made by other teachers. Finally, Smarter Balanced is administered in 3rd grade (English language arts), 4th grade (math), and 8th grade (ELA and math); the SAT is administered to all 11th-grade students. In other words, "state" assessments are administered in only six grades/subjects and local assessments in 17.

New Hampshire and its technical advisers are aware of the well-known challenges with implementing performance assessments as part of accountability systems. As noted above, some of those challenges to technical quality have been due to a misunderstanding about the distinction between individual student and aggregate level (e.g., school) reliability, but New Hampshire is still concerned about ensuring the quality (both technical and usefulness) of the PACE assessments.

The state uses Universal Design for Learning (UDL) to minimize accessibility barriers and ensure that all students are able to show what they know. Instruction and assessments based on UDL address the needs of the broadest range of students while avoiding potentially distracting information such as unnecessary wordiness and/or visually cluttered page layouts. Even with UDL, some students will need accommodations and supports to be able to meaningfully participate in instruction and assessment. While this is a civil rights issue, PACE must ensure that participating districts use a common approach in order to safeguard comparability. Therefore, participating districts and the state education agency created guidelines that define accommodations for students with disabilities and English language learners. This guide closely follows accommodations guidelines used for the previous state assessments and is now used for the Smarter Balanced assessments.

Other possible models

While the New Hampshire PACE system is now under way in eight districts and is closely watched by educators nationwide, other approaches are possible under the ESSA pilot. At least three states — Colorado, Maryland, and New York — have indicated an interest in applying, and others are likely to follow.

One possible approach is the performance assessment system used for decades by a consortium of schools in New York state. Under that system, students demonstrate that they are ready to graduate by completing performance tasks, written reports, and oral presentations. In addition, performance assessments are embedded in the curriculum throughout the school year.

States might also consider a "through-course" assessment. Under that approach, states would administer assessments throughout the year and roll them up into a final score for students and schools. Such a system is distinct from the familiar end-of-course assessments and lets states test a broader range of knowledge and skills. In its original design, PARCC had proposed a form of the through-course system but dropped it amid concerns that it might restrict states' flexibility in curriculum.

While the ESSA pilot allows these and other models, the law itself relaxes many requirements for state tests used for accountability, which means states can potentially pursue alternatives even without the pilot program.

Issues to consider

In taking part in the pilot program, states need to ensure that the assessments they are developing are sound and avoid problems encountered by earlier attempts at alternative forms of assessment. New Hampshire's experience shows how states can address some of the most critical issues.

Quality control. Professional standards for educational tests require evidence of validity, reliability, and fairness, and these apply to alternative forms of assessment as well as to conventional standardized tests.

New Hampshire's PACE system employs five layers of quality control relative to technical quality:

Professional learning and collaboration. Teams of teachers from all PACE districts collaborate and challenge one another in creating the assessment's common performance tasks. Their task development includes having students try the tasks before they're piloted and having teachers score the results to ensure consistency among raters.

Content-area leads. The state education department provides content experts in English lan-

guage arts, math, and science to help ensure that all tasks meet key design principles.

Technical review. The National Center for the Improvement of Educational Assessment (Center for Assessment), a nationally recognized nonprofit assessment firm, conducts technical reviews of each PACE common task. The content-area leads handle minor revisions, but more significant revisions must go back to the teacher teams and be resubmitted to the center for another review.

State review. The state education department does the final review and signoff on all PACE common assessment tasks.

Data review. After students have performed the tasks, the Center for Assessment leads teams from the assessment development group to review data from the assessments and examines student work to ensure that the assessments are eliciting the evidence of student learning that is intended.

Comparability and fairness. Assessments also must be administered and scored consistently and fairly so all students are held to the same standards and no group of students is at an unfair disadvantage.

In New Hampshire, scoring the common assessment tasks follows a rigorous protocol developed to ensure a high degree of scoring consistency within each district and as a first step for ensuring cross-district comparability. This scoring framework identifies “anchor papers” (prototypical student work at each competency level) and monitors consistency among teacher scorers.

But psychometric comparability (i.e., “interchangeability” of student scores) across districts administering different systems of assessment is not the goal. Rather, the state wants to ensure that students in PACE districts are held to comparable expectations. For example, students who are judged to be at a certain level in one district should receive similar judgments of quality in every other PACE district. Further, the state education department requires that students in PACE districts are held to the same or higher expectations as students from non-PACE districts. Students deemed proficient in a particular grade or content area should reasonably be expected to perform at the proficient level across different types of assessments.

Scaling up . . . carefully. The New Hampshire waiver and the ESSA pilot both require that states plan to scale up the pilot system statewide within a defined time frame. In that way, the new system can benefit all students and all districts. But accomplishing this can be challenging, since locally-determined performance-based assessments require much greater involvement of teachers.

New Hampshire has pursued a careful strategy for scale-up. The PACE system relies on volunteer high-capacity schools, but improving chronically low-performing schools will be a priority over time. The state education department is committed to supporting the development of local leadership and capacity to help low-performing schools implement the PACE system with fidelity.

On the other hand, while bringing PACE or any other innovative approach to scale is the long-term goal, doing so must be done deliberately to avoid setting schools up to fail. To that end, New Hampshire instituted several expectations to determine when districts are ready to participate in the pilot. Districts must:

- Have adopted the state model graduation and grade-span competencies and developed a coherent set of K-12 course and grade competencies.
- Have demonstrated the leadership and educator capacity to participate effectively in the pilot.
- Have developed or be close to completing the development of a comprehensive assessment system designed to appropriately measure student learning of required competencies.
- Be willing to participate in a peer and expert review process where they submit their system of performance-based assessments for evaluation based on clear and rigorous criteria.
- Administer Smarter Balanced or SAT in high school at least once at each grade span (e.g., 4, 8, and 11) to serve as both an internal and external audit regarding school and district performance.

Conclusion

For almost 15 years, states have used a single — and narrow — form of assessment that has failed to capture all the competencies students need to develop and in some cases has constrained instruction and created a backlash against all forms of testing. The recently passed ESSA now provides states with an opportunity to try new approaches. As New Hampshire’s experience shows, this new opportunity can open possibilities for improving instruction and learning — as well as new challenges. The next moves are up to the states. **K**

Reference

Elmore, R. (2002). *Bridging the gap between standards and achievement: The imperative for professional development in education*. Washington, DC: Albert Shanker Institute.

ESSA potentially could expand the use of innovative assessments dramatically.