

## ESSA's Flexibility on Assessment Elicits Qualms From Testing Experts

By [Catherine Gewertz](#)

The Every Student Succeeds Act invites states and districts to use interim assessments in a new way: by combining their results into one summative score for federal accountability.

But testing experts say it can be difficult to produce valid scores that way, and warn that the approach can limit teachers' curricular choices.

The No Child Left Behind Act required states to give "high-quality, annual tests" that measure students' year-to-year progress in math and English/language arts. ESSA, the latest revision of the Elementary and Secondary Education Act, signed into law Dec. 10, says that "at the state's discretion," student achievement and growth can be measured "through a single summative assessment" or "through multiple, statewide interim assessments during the course of the academic year that result in a single summative score."

Summative tests are typically measurements of learning taken when instruction is complete, while interim tests more often measure students' progress toward learning goals, after specific sections of instruction.

No one knows yet whether states will use the new option in the law, but it has the potential to make a big market bigger. [Industry analysts estimate that total U.S. spending on "classroom assessment"](#)—which includes formative, interim, and benchmark tests—totals \$1.3 billion per year, outdoing the \$1.1-billion-per-year statewide summative assessment market.

States that explore the new option might consider interim tests from the commercial market, or they could design their own tests by buying—or writing—questions or tasks and administering them statewide at specified points in the year.

Which interim tests the U.S. Department of Education will consider acceptable for summative results is an open question, since regulations and guidance on the new law haven't been written yet. And states will have to prove to the Education Department that their tests are valid for their intended purpose.

But states that choose to go the interim-testing route will have to grapple with key issues affecting their validity, and their power to shape curriculum, assessment experts say.

### **Bid for Better Assessment**

In adding the new language, lawmakers wanted to recognize educators' desire to use "more authentic" kinds of assessment, according to a former Senate staffer who worked on early versions of the bill.

"It came from the idea that [assessment] is not just a one-time, one-day, multiple-choice exam, that doing these authentic performance-based tasks over the course of the school year should be recognized as equivalent to a one-time, one-day evaluation," the staffer said.

Designing such a system for valid statewide results, however, is rife with challenges.

Psychometricians cringe when policymakers and educators use a test to measure things it wasn't designed to measure because it compromises the validity of the results. Combining interim test results into a summative score runs that risk, assessment experts said, but it depends on what tests states use and how.

"It seems like this could be the full employment act for psychometricians arguing back and forth about how to combine interims properly into a single summative," said Laress L. Wise, immediate past president of the

National Council on Measurement in Education, the association that sets standards for best practice in assessment. “There are ways to do it, but there aren’t good examples of it yet.”

States considering exploring the law’s new option must be particularly careful if they’re thinking of using off-the-shelf interim tests to produce a single summative result, said Derek C. Briggs, the chairman of the research and evaluation and methodology department at the University of Colorado-Boulder.

“It sounds appealing to use the interims you already have for that, but I don’t think people appreciate that the current system of interim tests isn’t designed to be a replacement for summative tests,” Briggs said.

The Northwest Evaluation Association, maker of one of the most widely used interim-testing systems, the Measures of Academic Progress, or MAP, agreed.

“Our interims as they’re designed today are not appropriate for use as a summative,” said Donna McCahon, the company’s assessment director. NWEA is currently refining a computer-adaptive testing system that blends interim and summative elements, and would be more appropriate for that use, she said.

### **Need to Reflect Curriculum**

To be combined for a valid summative score, interim tests “need to be written with a particular curriculum in mind,” Briggs said. “If they’re misconnected to the curriculum, there are all kinds of problems.”

For instance, one school might focus deeply on a topic for the first few months of the school year, and its students will do well on that interim test, Briggs said, but a school that chose to spread the same subject matter out across the year would risk having its students do poorly on that first interim.

To avoid those kinds of inequities, and to create results that are valid and comparable statewide, schools would have to teach shared curriculum topics in the same order, Briggs and other experts said.

“When people realize this, they’ll be concerned. It could introduce a kind of conformity most people would balk at,” said Gregory J. Cizek, a professor of educational measurement and evaluation at the University of North Carolina-Chapel Hill.

To avoid pushback against “lock-step” instruction, a state would have to engage teachers, parents, policymakers, and others in a dialog—and build a strong consensus—about what should be taught and when, experts said.

Fear of that kind of backlash, amplified by anger about federal intrusion on instruction sparked by the common core, was one reason that the Partnership for Assessment of Readiness for College and Careers, or PARCC, one of the two federally funded testing consortia, **abandoned its initial “through-course” design**, which would have spread out summative testing across the year.

Some districts and states are already working on blending different types of assessment into a year-end result.

**New Hampshire has drawn attention for a pilot program** that uses a mixture of tests from the Smarter Balanced Assessment Consortium and locally developed performance tasks. About 10 other states are investigating variations on testing that include features such as blending year-end summative tests with competency-based tests given throughout the year, said Jennifer Davis Poon, the program director of the Innovation Lab Network at the Council of Chief State School Officers, which is working with those states.

But states that venture into such projects should recognize that there are “a lot of technical hurdles to overcome,” Poon said. A particular challenge in New Hampshire is figuring out how to get comparable results across locally developed tasks that vary from one district to another, she said.