FutureEd
Independent Analysis, Innovative Ideas

Progress Report on Testing Innovation

The Limitations of Through-Year Assessments

by Lynn Olson

Through-year assessments address the desire of parents and educators for more timely, actionable information by testing students multiple times throughout the school year. Ideally, the results can inform instruction and also yield a final, summative score.

Not all through-year assessments are the same, however. There are curriculum-agnostic tests, which measure all grade-level standards at each test administration (such as pilots in Florida, Texas, and Virginia); curriculum-specific tests (such as the Louisiana Innovative Assessment Program for ELA, which enables districts to choose between tests aligned with one of two popular curricula: the state's ELA Guidebooks 2.0 or Wit and Wisdom); and curriculum-relevant tests, which districts can align to their own scope and sequence by measuring only a subset of state standards at any one time (such as pilots in Montana and Indiana).[1]

Each variety of through-year assessment poses challenges.[2] Assessing the full set of grade-level standards with each test means that students will be tested on content they have not yet been taught (particularly early in the year). "How does that even make sense?" asks Scott Marion, the executive director of the Center for Assessment. While such tests may be useful for measuring growth over time from an initial baseline, they can be demoralizing for students and distort the pacing and content of teachers' instruction.

Louisiana's Innovative Assessment Program for ELA is tied most directly to the instructional materials students used in class. Students take three assessments throughout the year (fall, winter, and spring) at the conclusion of curriculum units. Some of the writing tasks are based on texts they've studied, while others require them to apply that knowledge to unfamiliar texts on the same topic.

But Louisiana's curriculum-specific tests have proven difficult to scale. "While the vast majority of classrooms in the state use Guidebooks or Wit and Wisdom, developing assessments that work for every curriculum is challenging," the Louisiana Department of Education wrote in a report to the U.S. Department of Education.[3] Even among districts using the two curricula, implementation varies in pacing and unit selection. As a result, the state is trying to develop a test that could span multiple curricula.

Tying Testing to Teaching

No other state has taken Louisiana's curriculum-specific approach, but several are piloting curriculum-connected assessments. Montana's Instructionally Aligned Assessment System for math and ELA relies on a modular approach to accommodate the variety of curricula used across districts. Each "testlet" (eight in ELA and 12 in math) measures a subset of state standards. Districts and teachers can choose to

give the tests one by one or in batches throughout the year based on the scope and sequence of their instruction.

Devising an assessment that could yield more granular classroom feedback across different curricula "was a market requirement," says Arthur VanderVeen, the president of New Meridian, Montana's test contractor. Montana's more than 400 districts each select their own curricula, with more than 17 different math curricula used among the districts participating in the pilot. "A one-size-fits-all does not fit Montana," says Elsie Arntzen, state superintendent of public instruction. "It was important for us to enable a tool that would be aligned with local scope and sequence, because in Montana we don't determine the scope and sequence or the curriculum."

Close to one-quarter of Montana's districts are participating in the pilot, representing about 30,000 students. Still, determining when and how to administer the tests has entailed lots of logistical complexity. Feedback sessions with teachers and school leaders have revealed a "range of readiness to effectively sequence their assessment administration with their local curriculum, particularly for mathematics," says Julie Murgel, the chief operating officer for the Montana Department of Public Instruction. "I do think it's going to require much more extensive professional learning opportunities and guidance than is normally the case."

Indiana has tried to create through-year assessments that reflect multiple local curricula by surveying its 415 school corporations about their instructional scope and sequence to identify commonalities. The state previously had approved vendors to provide interim assessments to districts if they met certain criteria, including being at least 85 percent aligned with state standards and providing a predictive measure of state summative scores.

But the state found that even approved measures, such as NWEA's MAP tests and Curriculum Associates' i-Ready, were not fully aligned with state standards. As a result, students could show growth and still not perform well on the end-of-year test, says Lynn Schemel, chief academic officer for the Indiana Department of Education.

A bill enacted during the 2022 legislative session required Indiana to dramatically reduce the number of state academic standards by 25 percent and identify the most "essential" standards. Based on the survey and feedback from local test coordinators and curriculum developers, the department of education worked with its test contractors—Smarter Balanced and Cambium—to determine which essential standards to test at three ILEARN checkpoints during the school year. Each computer-adaptive test can be flexibly administered during a nine-week window and measures four to seven standards to help assess learning closer to when those standards are taught. Only a shortened, summative test given toward the end of the year will count for accountability.

The department is working with Cambium to tie checkpoint results directly to the instructional resources teachers use in classrooms, including those from NWEA and Curriculum Associates. "Our checkpoint data will flow directly into those vendor systems, so schools can still use those products," says Schemel. "Teachers can still use those same vendor materials, but the vendor diagnostic is going to be replaced with our checkpoints." Even so, some local curricula may be better aligned with the checkpoints than others, and the tests could reduce some teacher discretion and agency, since the three checkpoints must follow a prescribed sequence.

One goal among states piloting through-year assessments is reducing overall testing time. In addition to using a computer-adaptive format, Indiana shortened its final summative test in several ways: It replaced a performance task in math with open-ended items and eliminated one of two performance tasks in ELA. And while every student is tested in all essential standards, only a sample of students answer questions on other state standards on the end-of-year exam.

The Struggle to Reduce Testing Time

Despite efforts such as Indiana's, it's not clear that through-year assessments will lead to less testing time overall. The desire to produce more fine-grained information multiple times throughout the year could require students to spend more time taking state tests, not less. Moreover, unless states are willing to constrain district testing, districts may continue to give locally administered interim and benchmark assessments on top of through-year assessments. A study by the consulting company Bellwether of through-year assessment pilots in four states—Delaware, Florida, Nebraska, and Texas—found that districts were continuing to administer interim tests. "This is partly due to familiarity with the assessments and the data," the study found, as well as to concerns that districts would lose their ability to track student data over time if state tests changed.[4]

States also face the challenge of how to arrive at a summative score on through-year assessments. If a student doesn't master a standard initially but does so later in the year, or on a re-test, which score counts? And if a state rolls up all test results into one summative score, is it just creating multiple, shorter, high-stakes tests? Some states (like Florida) only count the final test administration to compute a summative score. Some (like Montana) are still figuring out how to arrive at a summative score across multiple tests.

Learning also progresses differently across different disciplines, with implications for through-year test design. Math involves more discrete sets of skills that build upon each other sequentially, so it makes sense to measure their acquisition over the course of a school year. Reading comprehension, in contrast, requires a more integrated set of skills that become more sophisticated over time as students apply their learning to increasingly complex texts. As a result, New Meridian plans to use different summative scoring models for Montana's math and ELA tests.

Through-year assessments also could blur the lines between low-stakes assessments used for instructional purposes and tests used for accountability in unintended ways. In February 2023, the Georgia Department of Education asked to withdraw from the Innovative Assessment Demonstration Authority (IADA), a federal program designed to encourage states to develop novel approaches to state testing, in part because of such tensions. Georgia received flexibility under the IADA in the 2019-20 school year to allow two consortia of local districts—the Georgia MAP Partnership and the Putnam Consortium—to pilot systems of assessment in place of the existing Georgia Milestones tests. The former used NWEA's computer-adaptive MAP assessments, while the latter used the Navvy ELA and math classroom assessments, offered by Pearson, which measure whether students have mastered individual state standards.

Allison Timberlake, deputy superintendent for assessment and accountability in the Georgia Department of Education, said participating districts loved the more instructionally focused tests but found it challenging to make them work for accountability. "Even if they liked the test design better and got data throughout the year," she says, "they started to realize they were really just expanding a Georgia Milestones-like assessment year-round."

In a status report to the U.S. Department of Education, Georgia noted that the consortia assessments would have to provide evidence in "six different categories with a total of over 30 separate criteria" to ensure they could support the same high-stakes decisions as the Georgia Milestone tests.[5] Georgia uses state test scores in grade retention and promotion decisions, as part of course grades in high school, in teacher and leader evaluations, and as a key component of school accountability metrics. Timberlake says that both consortia struggled to develop final, end-of-year scores for the assessments and to meet the high bar required for comparability

to existing state tests. The NAVVY assessments also proved hard to scale because they require districts to embrace a competency-based approach to instruction based on students' mastery of individual standards.

Despite the number of challenges, through-year assessments could offer an improvement over existing state standardized tests in timeliness, efficiency, and instructional utility. "The through-year assessments could be a really powerful opportunity from a parent perspective, because they are so connected to what the kid is learning and, done well, could inform how a teacher is differentiating instruction for their child," says Bibb Hubbard, president of Learning Heroes, a nonprofit that supports family engagement in education.

If such tests eventually replace the large number of district-administered tests of variable quality, they also could decrease the overall testing burden and increase coherence across the system, especially for many poorly resourced and smaller districts that cannot afford to invest in high-quality assessments.

The challenges facing through-year assessments would be fewer if the measures given throughout the school year were solely for instructional purposes and not tied to a final, end-of-year score that must be used for accountability decisions. That's one reason some states are opting to use only the final test administration to derive a summative score.

ENDNOTES

1 Education First, "What Are Through-Year Assessments? Exploring Multiple Approaches to Through-Year Design," 2023, Washington D.C.: Education First.

2 See, for example: Nathan Dadey, Carla M. Evans, and Will Lorie, "Through-Year Assessment: Ten Key Considerations," March 2023, Center for Assessment: Dover, N.H.

3 Louisiana Department of Education, "2023 IADA Annual Performance Report," 2023, https://oese.ed.gov/files/2023/10/LAIADA202223APR.pdf.

4 Michelle Croft, Titilayo Tinubu Ali, and Bonnie O'Keefe, "Testing the Waters: Insights into Parent Perspectives on Through-Year Assessment Implementation," November 2023, Washington D.C.: Bellwether.

5 Georgia Department of Education, "2021-22 IADA Annual Performance Report," September 29, 2022, https://oese.ed.gov/files/2022/09/GADOE-IADA-APR-21_22.pdf.