

Progress Report on Testing Innovation

The Promise and Perils of Performance Assessments

by Lynn Olson

Performance assessments address the desire of students, parents, and educators to use assessments that are more engaging, more supportive of high-quality instruction, and designed to measure students' capacity for real-world problem solving and application of skills. But scaling such assessments—which ask students to demonstrate what they know and can do through projects, experiments, and other more direct measures—has proven difficult.

“A good performance assessment or a good portfolio assessment may be able to give information to a teacher while also yielding a score that goes into a summative assessment,” says Marianne Perie, senior research director of assessment and accountability at West Ed, a research, development, and consulting organization. “The problem is, those are so labor intensive, they tend to collapse before they get very far.”

In the 2018-19 school year, New Hampshire received approval to implement its Performance Assessment of Competency Education (PACE) under the Innovative Assessment Demonstration Authority (IADA), a federal assessment pilot. Teachers helped design and score common assessment tasks, which they could give when it best fit their local curriculum. To ensure reliability and comparability of scoring across schools and school districts, all

pilot districts held grade-level calibration sessions with teachers. The state also audited a sample of common tasks scored by different teachers to check for consistency. A formative evaluation of nine PACE districts, published by the Human Resources Research Organization in March 2017, found that the assessments had a “substantial positive impact on both teaching practice and student learning.”¹

But in March 2022, New Hampshire requested to withdraw from the federal pilot, citing implementation challenges. “Unfortunately,” wrote Commissioner of Education Frank Edelblut, “PACE assessment task development, implementation, and scoring placed significant administrative burden on educators and school leaders that detracted from instructional time and was costly to implement. PACE also fell short of expectations for comparability and assessment quality.”²

“Almost all states are moving away from what PACE was trying to do in New Hampshire,” says Aneesha Badrinarayan, director of state performance assessment initiatives for the Learning Policy Institute (LPI), an organization that promotes performance assessments. Instead, states are looking for ways to incorporate performance assessments as one component—but not the sole component—of their state testing systems.

As an example of such a system, Massachusetts is piloting machine-scored, technology-based performance tasks as part of its state science test. The computer simulations, developed with test publishers Pearson and Cognia, ask students to

conduct investigations and model scientific data to solve real-world challenges. For example, a 4th-grade task asks students to determine the type and arrangement of barriers that would best reduce the amount of sediment flowing into a local lake from a construction site. The state has piloted the tasks for three years, alongside professional development for teachers and sample demonstration tasks to help teachers deliver high-quality science instruction.

Students in the pilot complete a shortened version of the existing state science test alongside the

When it comes to comparing performance across students and schools, “It’s really hard to make the argument that a performance assessment is worth the burden of administering it.”

performance tasks. Only the former counts toward an official student score, which has enabled Massachusetts to compare scores between its old and new assessments while transitioning to the new test. The state plans to roll out the performance tasks statewide in spring 2026.

As part of a 2022 contract with the Educational Testing Service, the California Department of Education is exploring innovations to the California Assessment System in science that would include developing classroom-embedded performance tasks and further revising the test blueprint to shorten testing time for the end-of-year assessment.

The proposed system would include classroom-based performance tasks that would be available for teachers in all grade levels to use throughout the year to inform instruction, as well as common performance tasks given over the course of the year to supplement the end-of-year summative tests in grades 5, 8, and high school. Students

would also complete a mini-performance task on the end-of-year tests. In addition, the report from ETS recommends opportunity-to-learn indicators in science, such as access to lab equipment and qualified teachers, to help place test scores in context. The state has not determined the timeline for any possible future actions.

“There’s a lot of interest in performance assessments for all the reasons that people have been interested in them: their authenticity, their meaningfulness, their ability to test things like critical thinking and problem solving,” says Badrinarayan of LPI. That will continue to make them hugely attractive as a way to provide ongoing feedback about teaching and learning and to signal what high-quality instruction should look like. But when it comes to comparing performance across students and schools, Badrinarayan adds, “It’s really hard to make the argument that a performance assessment is worth the burden of administering it.”

Game-based assessments, which have been used to provide rapid feedback to students about their learning, pose similar problems if used for accountability purposes. Kristen DiCerbo, chief learning officer at Khan Academy, which offers personalized learning and AI-powered tutoring for students, says, “I did a lot of work on simulations and games. And where they fell down was that it takes a student a long time to engage in a simulation activity and it would produce only three or four data points that really told us what students knew. It extended the amount of time that it took for the assessment and who wants more assessment time? The benefit wasn’t as big as we would have liked to see in some of that work.”

Performance assessments are a potentially valuable component of comprehensive assessment systems and states will continue to incorporate them in state tests to measure knowledge and skills—such as scientific and mathematical practices—that are hard to capture in other ways. But current challenges make it unlikely they will become the primary component of statewide standardized tests.

ENDNOTES

- 1 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.
- 2 Frank Edelblut, New Hampshire Commissioner of Education, March 9, 2022, Letter to Dr. Donald Peasley, Office of Elementary and Secondary Education, U.S. Department of Education, <https://oese.ed.gov/files/2022/04/NHIADAWithdrawal3.9.2022.pdf>.