

Massachusetts Department of Elementary and Secondary Education





Introduction

The vast disruptions caused by the Covid-19 pandemic prompted states and school networks to pause standardized tests and ask hard questions about the role and influence of traditional, once-a-year exams. This amounted to an opportunity to reimagine and redefine K-12 assessment and accountability alongside the students, families, teachers, and school leaders most familiar with the possibilities and pitfalls of our current approach.

The Innovations in Assessment and New Measure Grant Program, founded in 2020 by the Bill and Melinda Gates Foundation, Walton Family Foundation, and Education First, was developed to support this work.

The grant program supports new assessments that measure learning across multiple domains and capture student readiness for college or career, while continuing to provide the detailed data that an equitable approach to accountability requires. The grant program also supports a new strategy to create assessments: one that uses human-centered design principles, to more broadly share planning and decision-making power and build on the expertise of the people most proximate to the problem.

Five grantees are developing new measures to assess student learning across multiple domains. In contrast to traditional, top-down program design, these grantees are engaged with this work at the grassroots level, with a particular focus on how institutional, systemic, or ideological biases influence the experiences of some individuals and groups. These grantees are using **human-centered equitable design**, which builds on the expertise and experiences of students, families, and educators (informed by [Christine Ortiz's methodology](#)¹).

The experiences of these grantees can inform assessments that more fully meet the needs of students, families, teachers, and school leaders. While no one test will perfectly assess all that we need to know about student learning, a richer menu of options and base of knowledge rooted in the direct experiences of the people most proximate to the problem can support the high-quality measures of learning that all students and educators deserve. Such measures can capture the breadth of student achievement, with explicit connections to the broad range of knowledge and skills that nurture thriving adults across diverse cultures and communities.

Equitable Design in Action



¹Ortiz Guzman, Christine M., *Just Design, The equityXdesign Book*.

Deeper Learning for All: A Cohesive Approach to Revise Assessment, Accountability, and Classroom Experiences

Fifth-grade scientists in Massachusetts are running a series of water-purifying experiments through a hands-on, computer-based simulation. First, they learn how to cut and bend a plastic soda bottle to form a dome and reservoir. Then, they place virtual glasses of salty water inside the dome under various conditions and note the results: a large glass under direct sunlight, a small glass under a cloudy sky, and variations in between. They watch a video that shares instructions and encouragement from a diverse group of peers. Finally, these 10- and 11-year-olds are asked to review weather conditions in various regions of the world. Where would similar moisture collectors have the biggest impact on communities that need more fresh water?

It's a new type of science task for students in Massachusetts, which historically has had one of the nation's highest-quality statewide exams. In years past, questions on the 5th-grade test have featured diagrams of scientific phenomena and asked students to identify or

explain the roles of component parts, often through multiple-choice questions. By contrast, this new computer-based simulation puts students in an active role as scientists and engineers, conducting experiments and investigating models in a relevant context that reflects real-world problems.

Massachusetts is moving toward "Deeper Learning for All," a statewide push to expand equitable access to in-depth, meaning-making learning activities where students apply skills and knowledge to demonstrate mastery. In 2018, after a quarter-century of education reform in the Commonwealth, state leaders reviewed achievement trends in depth and engaged in a listening tour at schools to assess progress and opportunities to improve. Despite overall strong academic achievement, BIPOC students, students from families experiencing poverty, students with disabilities, and English-language learners were far less likely to be on grade level, graduate high school, or persist through college. In addition, many teachers reported feeling

pressure to spend time preparing students for annual exams and move away from problem-solving activities.

Creating an innovative test to include performance-based tasks complements two related initiatives to promote deeper learning: expanding access to high-quality curriculum and supporting professional development for teachers. In this three-part vision of change, standardized tests can meet the state's needs for measurement and accountability, while also sparking and encouraging important updates in school and teacher practices.

In keeping with the state's focus on equity, the development and design of these academically rigorous and technically complex assessment items is rooted in the direct experiences of students and teachers, who are helping shape and test these new tasks. By engaging these co-designers early and along the way, Massachusetts is working to ensure that these end-users have a voice in shaping the design of the new test.



Key Facts

Who:

Massachusetts Department of Elementary and Secondary Education

What:

As part of ongoing work to more equitably incorporate deeper learning into science instruction and assessment, create innovative pilot tests in 5th- and 8th-grade science to measure student performance through in-depth, simulated science and engineering activities.

Why:

Massachusetts has a long history of high student achievement, and its statewide MCAS tests are considered among the nation’s most rigorous. Since 1998, the exams, which include both multiple-choice and open-response questions, have helped identify schools and districts in need of support and intervention. But many classroom teachers report investing a lot of instructional time on memorization that they believe will prepare students for the MCAS, especially at schools where average scores are relatively low. Meanwhile, Massachusetts has adopted new science standards that emphasize deeper learning through problem solving, mastery, and real-world applications of relevant content knowledge. To expand access to deeper learning—particularly among students who are BIPOC, from families experiencing poverty, are English-language learners, or have disabilities—state leaders are piloting science tests in grades 5 and 8 to include the types of tasks that support meaningful learning experiences. This work accompanies related efforts to support schools to choose new, high-quality curriculum and provide related professional development to teachers to enhance classroom practice.

How:

For the past several years, state education leaders have been developing approaches to expand access to deeper learning in science classrooms across Massachusetts. Building on this work, the innovative science assessments project team consulted with committee groups to capture the needs and perspectives of students with disabilities, English language learners, and civil-right leaders, and incorporated those perspectives in its work with its assessment vendor. The team also conducted empathy interviews with the people most proximate to the problem: students, parents, and teachers at schools with traditionally low performance on the MCAS. Then, they convened teacher committees and focus groups to review and revise multiple versions of task simulations, to assess their alignment to grade-level knowledge and skills and ensure tasks were inclusive and relevant to diverse student groups. Students tested six rough-draft prototypes in cognitive labs that captured their feedback and experiences in real time, and the state then conducted two rounds of piloting in 2021 and 2022. As the work continues, Massachusetts will continue to pilot these refined test items and engage teacher committees focused on content and potential bias.

No. 1 for Some

In 2018, Massachusetts marked a quarter-century of accountability and standards-based reform. The state passed its landmark Massachusetts Education Reform Act of 1993, which set high and uniform academic standards, mandated statewide testing, and held schools accountable to meet goals. The law also established Chapter 70, a new school finance tool to help equalize spending across the state’s local and regional school districts, including by directing more state resources to lower-income communities. Chapter 70 required districts to meet “foundation” spending levels, which lawmakers deemed necessary for provide an adequate education that met the learning goals set forth by the broader law.

By many measures, the results were strongly positive. In 2017, Massachusetts ranked No. 1 in the United States on the National Assessment of Educational Progress (NAEP), though its rate of growth was slowing. If it were a country, Massachusetts would have the highest average scores in the world on the reading portion of the 2015 Programme for International Student Assessment (PISA) exam. And in 2021, nearly one in three high-school



graduates in the Commonwealth passed an Advanced Placement test—more than in any other state.

But a detailed look at the data showed that achievement gaps persisted, despite early progress in equalizing outcomes for students who are BIPOC, from families experiencing poverty, English-language learners, or have disabilities. These students earned higher scores than their counterparts in other states. Looking across Massachusetts, however, revealed stark differences when disaggregating data by demographic on the 2017 MCAS, for example. In 3rd-grade reading, 54 percent of white students met or exceeded expectations compared to 29 percent of Black and Latinx students. In 8th-grade math,

26 percent of students experiencing poverty met or exceeded expectations compared to 58 percent of students not experiencing poverty.

Meanwhile, the state was moving toward a more forward-looking set of goals and measures for student learning. In 2016, it adopted new science standards that prioritize scientific inquiry and hands-on practice across topics. In 2017, the MCAS was revised to emphasize critical thinking and applying knowledge, though it remained a primarily multiple-choice exam with a few types of open-response questions. And in 2018, the new Commissioner of Education, Jeffrey C. Riley, embarked on a listening tour and visited more than 100 schools across the state.

26 percent of students experiencing poverty met or exceeded expectations compared to 58 percent of students not experiencing poverty.

A New Focus on Deeper Learning

The common feedback was clear: teachers at schools that typically earned lower scores on the MCAS tests reported spending a lot of time preparing students for the exams. Helping students to memorize the broad range of information likely to appear on the test often came at the expense of more exploratory, meaningful learning experiences. This hyper-focus on educational outcomes on the MCAS tests contributed to inequitable opportunities for instruction.

The way forward was “Deeper Learning for All.” In Riley’s 2019 [report](#) to the state Board of Elementary and Secondary Education, he wrote:

“There is growing awareness not just in our schools—but also in the research community—that we must more closely match students’ daily experience in school with the expectations they will encounter in college, in their careers, and as citizens navigating a complex world. This means asking students to work in ambiguous contexts, on meaningful projects with larger purpose, and both independently and in teams—all while connecting these activities to our state standards.”

In this vision of deeper learning, students gather, interpret, and apply knowledge, work collaboratively, and master content through active engagement with real-world

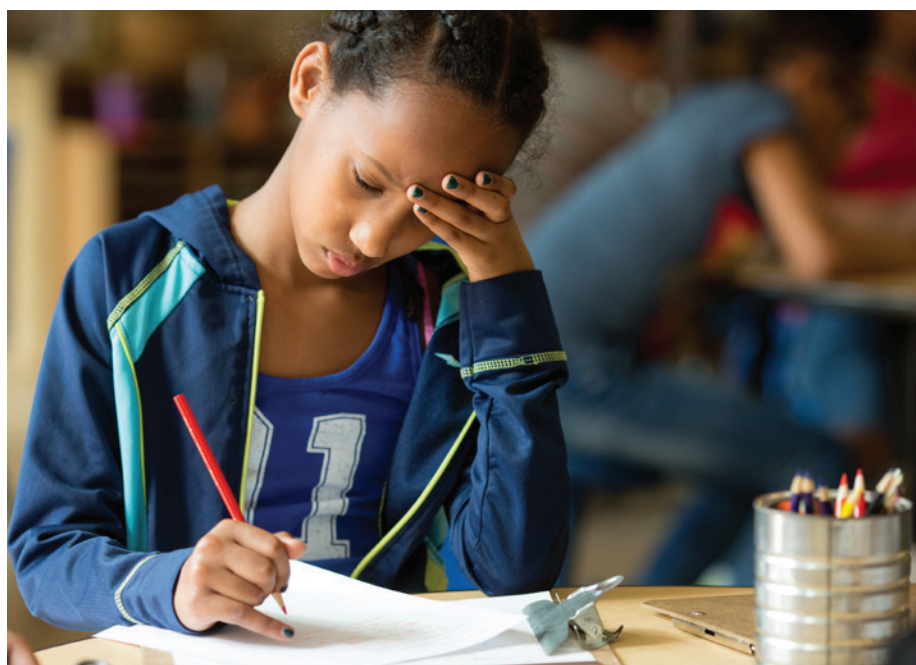
.....

“The students identify themselves as the ones doing the sense-making and engaging real-world phenomena.”

.....

information. In science classes, learning facts about science is the foundation for students to act as scientists in open-ended investigations that call on academic content, critical thinking, and communication with peers, all under a teacher’s guidance. This type of learning is sophisticated, highly engaging, and designed to prepare all young people with the knowledge and skills to flexibly navigate college and career in the 21st century—including students who have not been well-served by current modes of instruction.

“The students identify themselves as the ones doing the sense-making and engaging real-world phenomena,” said Corrine Steever, Innovative Assessment Project Manager on the Massachusetts team. “Through deeper learning, they can truly understand and do





Innovative Assessment and Demonstration Authority (IADA)

program and the Innovations in Assessment and New Measures led by Education First.

Taken together, the work is cohesive and focused on equity and inclusion. New assessments are intended to “encourage equitable access to deep learning experiences that will prepare students for success in and beyond school,” according to the project team.

“Our goals for deeper learning fit well with all of these pieces,” said Steever. “How are we identifying high-quality materials? What does this learning mean for individuals coming from different experiences and backgrounds? How can we build tools so students are driving learning and becoming college and career ready? How are we creating an assessment that really aligns with those instructional practices?”

science through multiple modes. For example, instead of defining a ‘habitat,’ students show true understanding by creating a model habitat and helping meet the needs of plants and animals in it.”

It’s a dramatic shift, one that rests on three complementary conditions: school districts selecting high-quality curriculum, teachers receiving professional development to lead these new lessons, and the state revising standardized assessments to assess deeper learning within existing accountability measures. Massachusetts is working to advance all three.

Its initiative **CURATE**, or CUrriculum RAtings by TEachers, convenes panels of Massachusetts teachers to review and rate curricula against state standards for English, math,

and science, to inform curriculum choice at the district level. The **Kaleidoscope Collective for Learning**, a wide-ranging network of schools and educators, provides resources and professional-development opportunities for Massachusetts teachers looking to incorporate deeper learning in their classroom practice. And the state joined two programs to develop new performance-based tasks that better reflect deeper learning: the federal

.....

The shift to deeper learning rests on three complementary conditions: school districts selecting high-quality curriculum, teachers receiving professional development to lead these new lessons, and the state revising standardized assessments to assess deeper learning within existing accountability measures.

.....

Engaging in Equitable Design

Innovating Alongside Teachers and Students

The work of the project team was already well underway when Massachusetts entered the Innovations in Assessment and New Measures grant program. Not only had state education leaders investigated and framed the problem a few years earlier, but because Massachusetts was already part of the federal IADA program, “we had already done a lot of strategic planning and discussing,” said Sam Ribnick, Special Advisor for Innovative Assessments at the state education department. The team had already planned to focus on science in grades 5 and 8 and test items with teachers and students as part of its design process.

“The good news was that our thinking was really well-aligned with what Education First was encouraging as well,” he said. “If we are trying to do something innovative, it’s critical to get it in front of students and teachers as early in the design as possible, so we’re not just guessing but acting on real input from a representative group of students and getting real feedback from teachers.”

This stood somewhat in contrast to the “little bets” approach, which is somewhat more open-ended and driven by the perspectives of the people most proximate to the problem. “By the time we had written our application and been accepted, we had a certain amount of timeline pressure and follow-through on our plan,” said Ribnick.

The development process followed several core principles of equitable design. Members conducted empathy interviews with about two-dozen families and students at the outset of the grant work to explore its theory of action. They also met with advisory groups representing students with

disabilities, English language learners, and racial justice in education. Then came task design. The team started with nine potential ideas in each grade and enlisted teachers and students to help bring those ideas to fruition.

Diagnosing Challenges, Investigating Bias

Two teacher committees, each with 10 members paid for their time, contributed to a robust series of rapid-cycle prototyping for either 5th or 8th grade tasks. That process followed six rounds of development in two-week cycles: first, the team and committees reviewed and weighed in on initial concepts, storyboards,



multiple versions of rough-draft items, and finally, reviewed three working prototypes.



In addition, a Bias and Sensitivity Committee of 10 members reviewed all sample tasks in the early and near-final stages to ensure they were relevant across all student groups, identify potential bias due to different cultural experiences outside of school, and ensure that tasks did not reinforce negative power structures.

Each task included a storyline, a simulated experiment in which students could change variables in a simulated apparatus and run the experiment multiple times to record results and identify an optimal design, and questions that prompted them to draw conclusions and demonstrate their knowledge. The goals for each task included featuring diverse representations of students, teachers, and scientists; depicting scientific phenomena in the real world; and situating tasks within narratives that were relevant, meaningful, and purposeful to students.

It was a “very intense period,” said Ribnick. Teachers spent 2–4 hours each week reviewing drafts, while education department and vendor staff worked quickly to incorporate teachers’ feedback. In addition, these processes were unfamiliar, because sharing rough drafts outside of the department was a departure from typical practice.

“Our norm, and one of the ways that we like to sustain trust in our testing program, is that we make sure that what we put out there is as high quality as we can possibly make it,” Ribnick said. “It’s really important to define upfront the level of finished quality that you’re going to be comfortable with for prototypes and have really clear communication about that.”

The team also tested these sample tasks with students in “cognitive labs,” when 37 students in grades 5 and 8 reviewed tasks in depth and observers noted their reactions and comments. They also invited a few classrooms of



Teachers’ feedback helped developers make important changes to the tasks along the way.



students to try out more fully developed tasks and collected survey responses afterwards. While 91 percent of students said they could “definitely” or “somewhat” show their science knowledge on the test, only 45 percent said they had done real experiments in class that were similar to the sample tasks.

Teachers’ feedback helped developers make important changes to the tasks along the way. Many felt that early versions required students to do too much reading, or that the text was too complex in introducing the task. The storyline and activities of each task were streamlined so students could complete them more quickly. And teachers helped the team expand its understanding of inclusion and diversity in the characters to reflect age, body type, and non-binary gender.

Teachers shared positive reactions to the assessment approach. As one teacher said, “it is more about the skill of doing science. Look for patterns and connections. More authentic.”

Another teacher commented, “I think that the new... style in 5 and 8 grade moved in the right direction in being more equitable. I would love it to go even more in the direction of more open-ended and practice-based questions.”

Looking Ahead

Enlisting teacher committees to work alongside designers and the project team has been a critical component to this effort, Ribnick said. However, for committees to be most efficient and effective, teachers need some baseline knowledge about standardized assessment practices under federal law. The Massachusetts team is now convening committees for two-year periods so participants can build, and then build on, that foundational understanding.

“We have people truly co-designing, and they are with us for a long enough stretch of time that it’s worthwhile to do some upfront training,” said Ribnick. “As they spend more time with us, they really develop more expertise and understanding of assessment design.”



Working directly with teachers during the design process has helped reinforce the project team’s theory of action and set the stage for the next phase of work. Teachers have stressed that students are unlikely to be prepared to complete simulation tasks as part of a standardized test and noted that practicing scientific

investigations, using technology in the classroom, and providing professional development for teachers will be critical next steps—setting the stage for greater student access to storyline-based high-quality science curriculum, aligned professional development for teachers, and deeper learning opportunities for more students across the state.



.....

Enlisting teacher committees to work alongside designers and the project team has been a critical component to this effort.

.....