What does it look like to monitor student progress toward rigorous science standards?

Three-dimensional science standards based on *A Framework for K-12 Science Education* raise several questions about how to monitor student progress toward those learning goals. The Task Annotation Project in Science (TAPS) was launched to provide an answer to the questions "what does it look like to ask students to demonstrate progress toward three-dimensional standards?" and "what are the most important features of high-quality science tasks?" By asking a diverse set of experts to identify concrete features of three-dimensional assessment tasks across multiple domains of science across the K-12 spectrum, this project was designed to 1) collaboratively provide concrete, annotated examples of different kinds of three-dimensional assessments, highlighting features of high-quality science assessments and opportunities for improvement, and 2) surface lessons learned from looking across grades, domains, and task purposes. The criteria, examples, models, and other resources developed can support teachers, developers, and decision-makers put better assessments in front of students.

Overall Takeaways

✓ There are many different and appropriate ways to assess three-dimensional standards.

Supporting student progress toward three-dimensional standards requires a variety of assessments designed for different purposes. All of these can be appropriate, as long as they include the "must-have" features of 3D assessments , are designed to support students in meeting three-dimensional science goals, and we're clear about the intended uses and interpretation of student responses.

Bottom line: Surfacing students' understanding and ability to use the three dimensions to make sense of phenomena or address problems requires that tasks use relevant, engaging, and sufficiently rich phenomena- and problem-based scenarios; require students to make sense-making—using the appropriate core ideas, science and engineering practices, and crosscutting concepts—visible; intentionally support diverse learners; and embody both the major shifts of the NGSS as well as support the use they are intended to serve.

✓ Equitable and fair tasks are better three-dimensional assessments.

Three-dimensional tasks are increasingly real-world, specific, and complex. These features ensure that assessments surface what students know and can do. It is important to develop tasks that are relevant, engaging, accessible, and motivating to students; that are rich enough to motivate student thinking through the entirety of the task; that offer diverse students ways to access and find meaning in the task at hand; and that truly value student thinking. Attention to these features of equity helps make sure both that the tasks elicit student thinking from all students as well as making it more likely that students will have to sense-make using the three dimensions.

✓ We are still figuring out how to talk about student progress.

One consistent observation across TAPS was that while many of the tasks themselves emphasize multi-dimensional sense-making and problem solving, it is challenging to figure out the right ways to convey student progress meaningfully in terms of grades, points, and rubrics. This requires balancing several features, like monitoring progress along all three dimensions while emphasizing their use together, not separately; making sure rubrics connect to how the task is designed to be used while being clear about what conclusions can and can't be made; and navigating the desire to provide meaningful feedback with the time it takes to do so for large numbers of students

✓ Continuous improvement is key.

We have learned so much about three-dimensional assessments—what they look like, how they surface student thinking in different contexts, and the challenges we face developing and using them. Every task that was annotated as part of TAPS had both substantial strengths and some opportunities for improvement associated with it, including those that can be addressed more immediately (e.g., more engaging phenomena, more emphasis on making student thinking visible), while others may take longer to wrestle with (e.g., balancing open-ended tasks with time needed for grading). As further lessons learned and evidence from student work emerges, it is imperative that assessment systems be designed with continuous improvement in mind.

TASK ANNOTATION PROJECT in SCIENCE: TAKEAWAYS

Classroom teachers

✓ Make students' three-dimensional thinking visible. Emphasize assessment opportunities that ask students to show you what they know as they figure out a phenomenon or problem—shift the "right answer" away from facts and definitions and toward evidence-based reasoning.

✓ Embrace a range of artifacts as evidence of student learning. Remember that many different artifacts students produce—writing, discourse, models, presentations, etc.—can be important assessment opportunities, both formatively and during more summative assessment experiences.

✓ Progress should always be celebrated. Be discerning about the assessments you're using and how they are being used, but don't feel like you need to "throw out" all of your current assessments—instead, think about what tweaks you can make over time to better elicit three-dimensional performances from all of your students. The Science Prescreen and Science Task Screener can help identify next steps for modifying assessments.

Assessment Developers

✓ For classroom and external assessments alike, focusing on getting the phenomenon or problem right will lead to more equitable assessments that elicit grade-appropriate three-dimensional performances.

✓ Focus on student reasoning with evidence using the three-dimensions—not simply right or wrong answers facts or descriptions—as the targeted student responses and the basis for rubrics and scoring guidance. Be creative about ways students can make facets of their thinking visible.

✓ Be transparent about what assessments measure—and what they don't. With the range of expectations associated with new science assessments, it is especially important to be transparent about how each assessment balances those tradeoffs. Be honest and evidence-based about what tasks and tests do and do not tell you about student performance, and how that information should be used as part of a system of assessments to support student

Administrators

✓ New science standards ask teachers and students to reimagine how students show us what they know and can do, and this influences how students are assessed regularly in the classroom (e.g., a shift away from content-focused quizzes and toward incremental modeling as regular assessment opportunities). School-and district-level policies about grades and assessments should account for these shifts, providing teachers with the flexibility to use the assessments and feedback loops that are right for their students and parents.

✓ Help teachers do what they do best. Supporting students means ensuring that teachers have the time, space, and community to develop, modify, and use high-quality assessments, as well as discuss student work, lessons learned, and new ideas with their peers.

✓ Coherent systems matter. The assessments students see in the classroom should complement and be coherent with school-, district-, and state-wide assessments. The major lessons learned here—including the non-negotiable features of NGSS assessments, the criteria reflected in the task screener, and possible tradeoffs to consider—can support school- and district-wide assessment efforts. See some models here.

Policy-makers

✓ Doing better at the classroom requires signals and incentives from leaders. Rigorous and high-quality tasks are imperative for student progress. To be used regularly, teachers, parents, students, and administrators need external signals that validate, incentivize, and provide feedback on these efforts. This means advocating for better aligned state- and district-wide assessments.

✓ How students interact with tasks matters, especially for those students who often feel left out of science. Supporting student success and achievement for all students requires creative approaches to systems and structures that value and incentivize evidence of student performance that is grounded in the classroom.

✓ Science can be an innovation zone for other content areas. While TAPS focused on science assessments, the ideas surfaced here—about the kinds of scenarios we ask students to address, what counts as meaningful and rigorous performances, how to balance tradeoffs and implications for innovative systems of assessment—are relevant and can inform similar innovative approaches in other content areas.