

Methodology and Approach

The [Task Annotation Project in Science \(TAPS\)](#) asked expert reviewers to use an evidence-based process, detailed below, to surface important features of science tasks that are designed for the NGSS. This approach can be modified to be used in professional learning communities, to annotate tasks for state and district communication purposes, and to review science assessment tasks to determine how well they are designed for standards based on A Framework for K-12 Science Education.

Project Launch

In February 2018, Achieve convened two meetings with experts to discuss task-level alignment considerations for science assessment. A common, urgent need that emerged from those meetings was for concrete examples that showcased features of science tasks that are designed for three-dimensional assessments: both strengths and opportunities for improvement. As a result of those meetings, the Task Annotation Project in Science was launched, with the goal of bringing experts together to 1) collaboratively define what science tasks designed for the NGSS look like, 2) providing educators and developers with concrete examples of the kinds of performances students should be demonstrating during and by the end of instruction to show progress toward new science standards, and 3) helping educators, developers, and decision-makers develop a common and shared understanding of the key features of NGSS assessments that can be used to inform professional learning, instruction, and both classroom and external assessment efforts.

A Task Annotation Leadership Group was formed to guide the work, including prescreening tasks, developing an evaluation and annotation tool and protocol, developing and delivering training for all reviewers, and providing critical thought leadership throughout the review, annotation, and finalization processes.

Task Selection

Tasks were selected from a pool of tasks that included publicly available assessments (e.g., those that are part of units that appear on [nextgenscience.org](#)) as well as those submitted by developers, with the agreement that they may be annotated and publicly shared. All submitted and collected tasks first went through a prescreen process, during which at least two members of the task annotation leadership group prescreened tasks relative to a set of agreed-upon criteria. Those criteria were refined based on the process and shared as the [Science Task Prescreen](#).

Of 72 classroom tasks that were prescreened, 31 tasks that passed the prescreen were assigned to a reviewer group for evaluation and annotation. Released sample statewide summative assessment items from 8 states (including one common set across several states) were also reviewed and annotated; lessons learned from those assessments are included in the summary resources, but annotated tasks are not being released at this time.

Reviewers and Training

40 expert reviewers were recruited based on their prior experience with the Next Generation Science Standards (NGSS) or similar three-dimensional standards. Reviewers included classroom assessment developers, national, state, and local science education leaders; research partners; and classroom teachers and administrators. All reviewers first received virtual training on an evaluation and annotation protocol, including an opportunity to use the protocol on sample tasks and have mock consensus conversations.

Task Evaluation and Annotation

Following virtual training sessions, reviewers used a three-part process that that involved individual criterion- and evidence-based review using a pre-published version of the [Science Task Screener](#), group consensus conversations about the evaluation of the task, and group annotations based on the consensus evaluation. Every task was evaluated by 3 expert reviewers, and each group evaluated 2-5 tasks and generated a summary report for each. Following the initial review, all annotated tasks and summary reports were peer-reviewed by the reviewer group as well as additional external experts before being finalized. While the reviews and annotations from all tasks (classroom and statewide summative) were used to develop the summary resources described below, some tasks will not be publicly released per the developers' request. The Science Task Screener was modified based on feedback from users during the reviews prior to its release.

Development of Additional Resources

Several themes, trends, and lessons learned emerged from looking across annotated assessment tasks and through participating in the range of group consensus conversations. The Leadership Group documented these ideas through a series of short resources designed to be easily used by educators, developers, and leaders in support of professional learning, research, materials development efforts, and as part of efforts to support new science standards implementation.



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.



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 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.



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



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.