Reconceptualizing Alignment for NGSS Assessments

Aneesha Badrinarayan, Achieve

Jill Wertheim, Stanford/SCALE

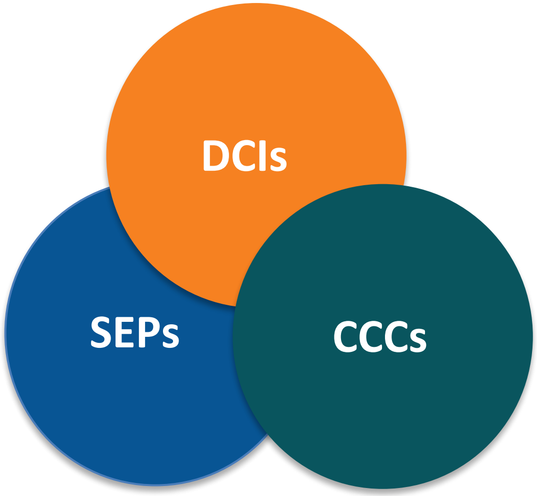
Bill Penuel, CU Boulder

TJ Smolek, Michigan DOE

Joseph Krajcik, MSU

NARST Annual Meeting
Baltimore, MD
March 31, 2019

What is an NGSS assessment?



The many meanings of "3D performance"

What are the key features of NGSS assessment?

Jill Wertheim.
Research Associate, Stanford Center for Assessment, Learning, and Equity (SCALE)
Stanford University.

TJ Smolek.
Science Education Research Consultant, Office of Educational Assessment and Accountability, Michigan Department of Education.

Bill Penuel.
Professor of Learning Sciences and Human Development, School of Education & Institute of Cognitive Science, CU- Boulder

Sara Cooper.
Science Education Specialist, Nebraska Department of Education.

Joe Krajcik.
Lappan-Phillips Professor of Science Education; Director, CREATE for STEM Institute, College of Natural Science, College of Education, MSU.

Aneesha Badrinarayan.
Director, Special Initiatives
Achieve

Goals of the project

Develop concrete examples, with explicit reasoning, of the essential qualities of 3D tasks

Provide a platform for experts **to be explicit about ideas about what it means to elicit a 3D performance.**

Begin **developing the "case law" for 3D assessments** and expected student performance.

Our process.

Assessment Task Evaluation

Foundation:

- A Framework for K-12 Science Education (NRC, 2012)
- Developing Assessments for the Next Generation Science Standards (NRC, 2014)
- Criteria for Procuring and Evaluating Science Assessments (Achieve, Inc., 2018)
- Criteria for High-quality Assessment (Darling-Hammond et al., 2013)
- Knowing what Students Know (NRC, 2001)

**and the expertise of the leadership team

Task Screener

A. Tasks are driven by high-quality scenarios that focus on phenomena or problems

B. Tasks require sense-making using the three dimensions

C. Tasks are fair and equitable

D. Tasks support their intended targets and purpose

Science Task Screener

Using the Task Screener. Use this tool to evaluate tasks designed for three-dimensional standards. For each criterion, record your evidence for the presence or absence of the associated indicators. After you have decided to what degree the indicators are present within the task, revisit the purpose of your task and decide whether the evidence supports using it.

Before you begin: Complete the task as a student would. Then, consider any support materials provided to teachers or students, such as contextual information about the task and answer keys/scoring guidance.

A. Tasks are driven by high-quality scenarios that are grounded in phenomena or problems.	B. Tasks require sense-making using the three dimensions.
i. Making sense of a phenomenon or addressing a problem is necessary to accomplish the task. ii. The task scenario—grounded in the phenomena and problems being addressed—is sufficient, engaging, relevant, and accessible to a wide range of students.	i. Completing the task requires students to use reasoning to sense-make about phenomena or problems. ii. The task requires students to demonstrate grade-appropriate: a. SEP element(s) b. CCC element(s) c. DCI element(s) iii. The task requires students to integrate multiple dimensions in service of sense-making and problem-solving. iv. The task requires students to make their thinking visible.

C. Tasks are fair and equitable.	D. Tasks support their intended targets and purpose.
i. The task provides ways for students to make connections of meaningful local, global, or universal relevance. ii. The task includes multiple modes for students to respond to the task. iii. The task is accessible, appropriate, and cognitively demanding for all learners, including students who are English learners or are working below or above grade level. iv. The task cultivates or explicitly builds upon students' interest in and confidence with science and engineering. v. The task focuses on performances for which students' learning experiences have prepared them (opportunity to learn considerations). vi. The task uses information that is scientifically accurate.	i. The task assesses what it is intended to assess, and supports the purpose for which it is intended. ii. The task elicits student artifacts that provide evidence of how well students can use the targeted dimensions together to make sense of phenomena and design solutions to problems. iii. Supporting materials include clear answer keys, rubrics, and/or scoring guidelines that are connected to the targeted three-dimensional standards and provide the necessary and sufficient guidance for interpreting student responses relative to all three dimensions and the target as a whole. iv. The task's prompts and directions provide sufficient guidance for the teacher to administer it effectively and for the students to complete it successfully while maintaining high levels of students' analytical thinking as appropriate.

Task Screener Indicators

ii. The task scenario is engaging, relevant, and accessible to a wide range of students*.

Features of engaging, relevant, and accessible tasks (Check the appropriate box, then describe rationale with evidence)

Features of scenarios	Yes	Somewhat	No	Rationale
Scenario presents real-world observations	☐	☐	☐	
Scenarios are based around at least one specific instance, not a topic, statement	☐	☐	☐	
Scenarios are presented as puzzling/intriguing	☐	☐	☐	
Scenarios create a "need to know"[1]	☐	☐	☐	
Scenarios are explainable using grade-appropriate SEPs, CCCs, DCIs	☐	☐	☐	
Scenarios effectively use at least 2 modalities	☐	☐	☐	

Analysis of existing NGSS assessments

Solicited a variety of tasks from a wide range of sources

Analysis of existing NGSS assessments

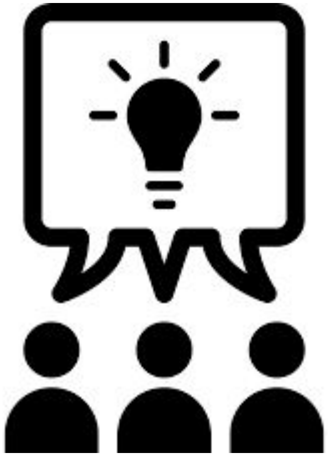
Solicited a variety of tasks from a wide range of sources

Prescreened 72 tasks for baseline features of 3D assessments.

Analysis of existing NGSS assessments

Solicited a variety of tasks from a wide range of sources

Prescreened 72 tasks for baseline features of 3D assessments.



Over 40 diverse experts engaged in a rigorous evaluation process to evaluate and annotate assessments.

Task Analysis

- 31 (43%) of classroom tasks met baseline requirements for analysis
- Each task analyzed by 3 reviewers
- Each reviewer analyzed 3-4 tasks
- Reviewers were asked to provide both a criterion-based evaluation of tasks as well as a separate quality judgement of strengths and weaknesses
- All annotations and consensus reports underwent an additional round of review by 3 expert reviewers.
- Reviews underwent secondary analysis for consistency by leadership team
- Reviewers completed a survey to explore common themes that emerged from reviews

Achieve

SCIENCE TASK ANNOTATION

ANNOTATION KEY

EQUITY	SCENARIOS	SEPs	DCIs	CCCs	SENSE-MAKING	ASSESSMENT PURPOSE
Supporting a wide range of diverse students.	Information provided to elicit performances.	Opportunities to demonstrate science and engineering practices.	Opportunities to demonstrate understanding of disciplinary core ideas.	Opportunities to demonstrate understanding of crosscutting concepts.	Opportunities for reasoning about phenomena and problems.	Highlights how the task features connect to intended assessment use.

Targeted Performance Expectation:

HS-LS-4-3: Apply concepts of statistics and probability to support explanations that organisms with an advantageous heritable trait tend to increase in proportion to organisms lacking this trait.

[Clarification Statement: Emphasis is on analyzing shifts in numerical distribution of traits and using these shifts as evidence to support explanations.]

[Assessment Boundary: Assessment is limited to basic statistical and graphical analysis. Assessment does not include allele frequency calculations.]

EVOLUTION OF SWALLOWS

In the 1970's along the I-80 highway in Keith County, Nebraska, drivers started noticing large numbers of dead swallows on the road. This led to a 45-year long study on swallow roadkill to figure out why this was happening.

Cliff Swallows traditionally built their nests on vertical cliff faces. However, with the expansion of roads, they have adopted many bridges, overpasses, and culverts as their colonial nesting sites. Their nests are grey or brown with openings at one end. Cliff Swallows zoom around in complicated aerial patterns to catch insects for food.

Image source: http://www.cell.com/cms/attachment/2021743115/2041577164/gr1_lrg.jpg

Source of data: Brown, C. R., & Brown, M. B. (2013). Where has all the road kill gone? *Current Biology*, 23(6), 233-234.

This is a specific, puzzling phenomenon that grounds student thinking. The phenomenon is set up to be comprehensible to a wide range of students, with limited words and sufficient detail to help students visualize what is going on. The images are helpful, and the scenario could be improved if it included videos of swallows' aerial behavior, so that students could more easily visualize this—especially as the aerial behaviors are most relevant to explaining the phenomenon as set up by the task.

The scenario text includes:
1) the specific observation that a large number of dead swallows were observed on the road,
2) some cuing that this is unusual and relevant to a community (language like "started noticing" and "45-year-long study to figure out why this is happening", and
3) identification of a "gap" or uncertainty [why this is happening] that students need to fill with their understanding of science SEPs, DCIs, and CCCs.

SCENARIOS EQUITY

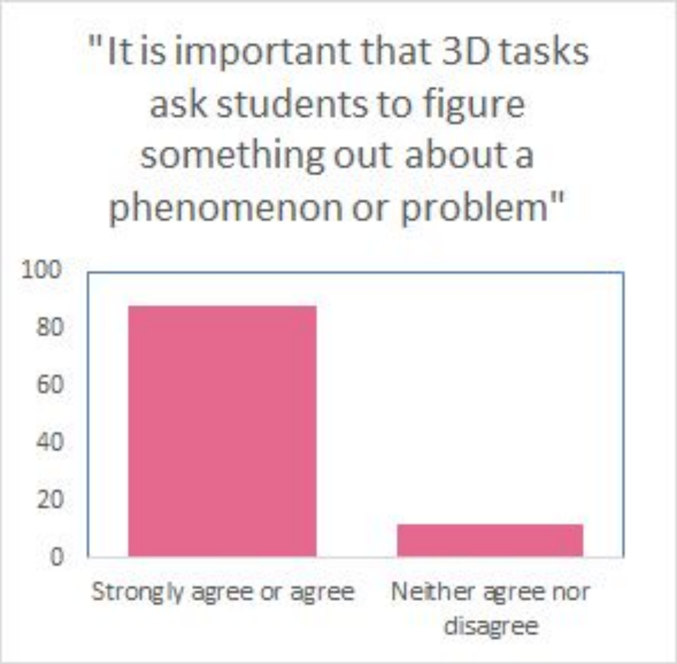
1

Identifying common themes and divergent viewpoints.

- Emergent themes from task analyses
- Surveys exploring emergent themes
- Direct conversations with researchers and educators

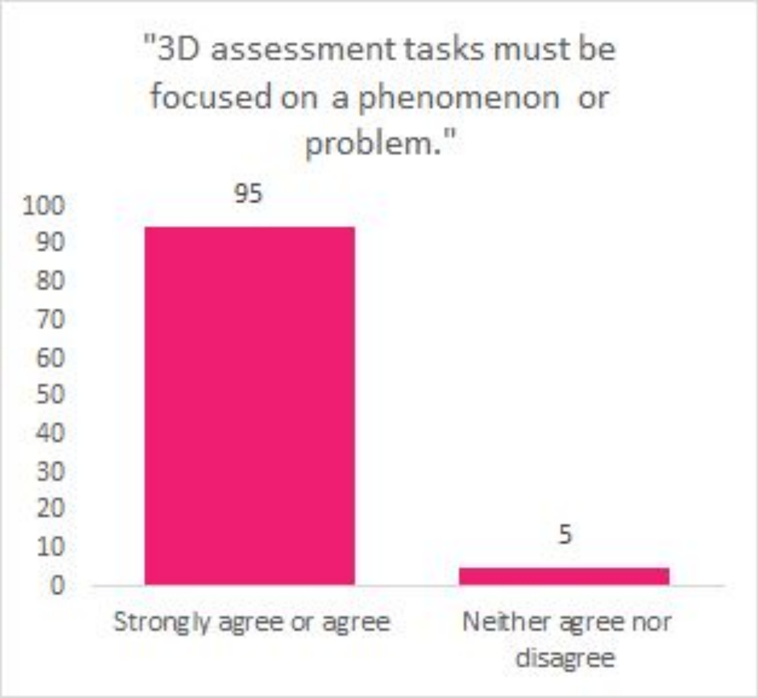
There are some important features of 3D tasks about which there is overwhelming agreement.

Sense-making or "figuring out" is the litmus test for NGSS assessments.



- Reviewers consistently identified that assessments claiming to be designed for 3D standards should focus on asking students to use their understanding of science ideas and practices to "figure something out" about a phenomenon or problem.
- An emergent practical definition of sensemaking is the construction of an understanding of a phenomenon or problem using the DCIs, SEPs, and/or CCCs being measured.
- Importantly, tasks were consistently critiqued if DCIs, SEPs, or CCCs were engaged ***without*** being tied to sense-making (i.e., these were identified as weaknesses).

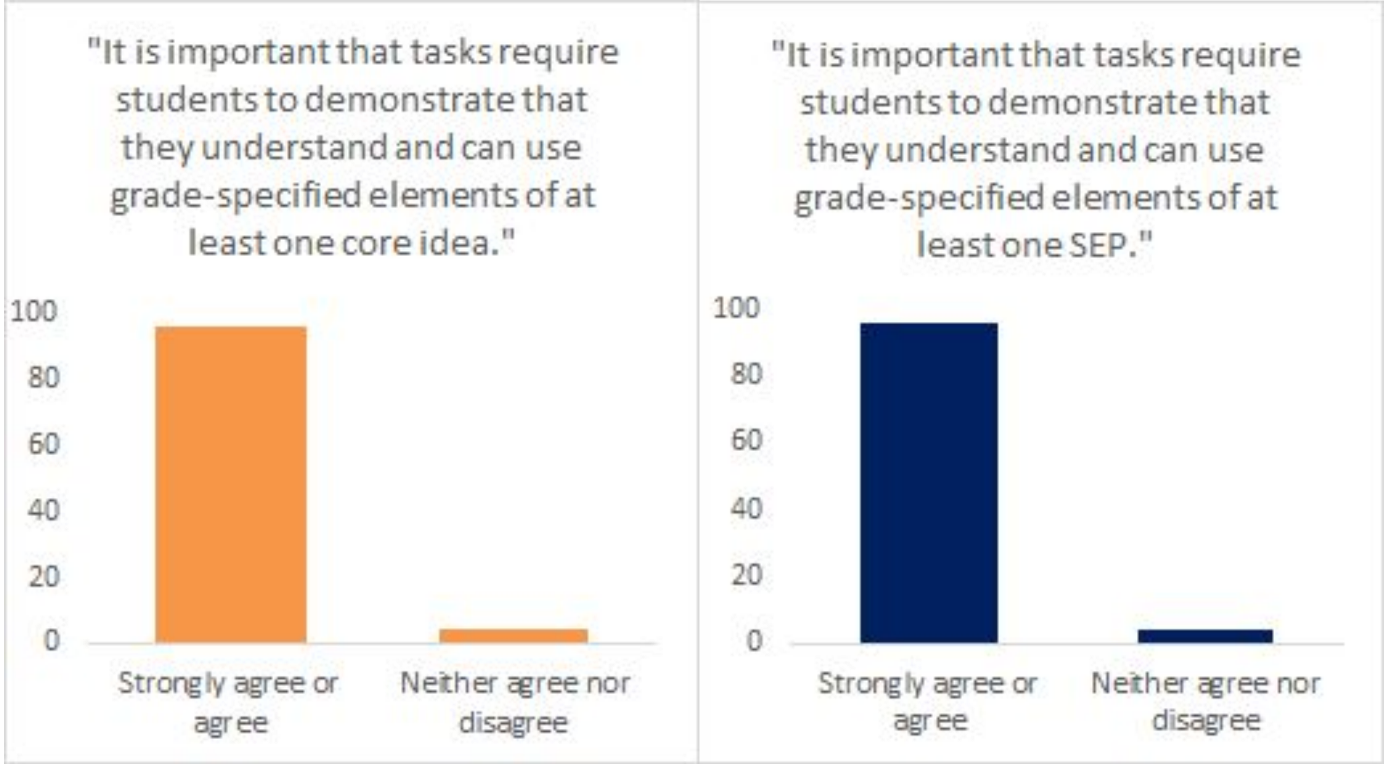
Accessible phenomena or problem-driven scenarios must motivate student responses.



Across tasks analyses, features that were identified as critical for 3D tasks included:

- Tasks must be based on a specific instance (not topic or statement)
- Tasks must be grounded in real-world observations
- Phenomena used must be "problematized"
- Some kind of interpretative reasoning about the provided phenomenon must be required to respond to the question
- Aspects of the phenomenon or problem targeted by the task must be explainable using grade-appropriate SEPs, CCCs, and/or DCIs

It is critical that DCIs and SEPs are part of 3D tasks.



The most consistent and harshest critiques of tasks identified when tasks **were not requiring the demonstration of SEPs or DCIs**.

Scoring guidance must provide sufficient support for interpreting student responses.

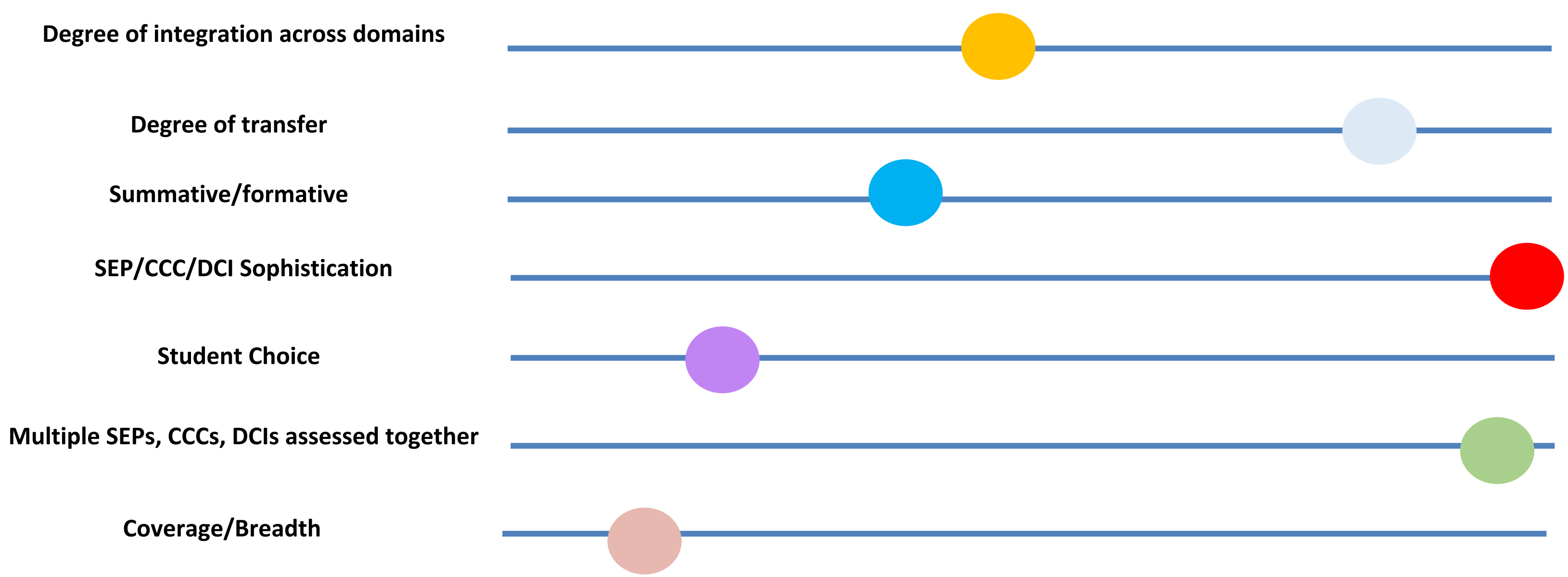
Common pitfalls identified included:

- Mismatch between claims and the scoring guide
- Vagueness about what is being evaluated
- Vagueness about how to interpret student data
- Inaccuracies about what is being evaluated/elicited

There were some features that elicited divergent quality judgements.

All assessment tasks involve some tradeoffs, based on purpose and goals for the assessment. However, we also found that some reviewers and developers made consistent trade-offs across assessment purposes, indicating some underlying differences in philosophies, perspectives, and values about science tasks.

Knobs to turn based on philosophy, priority, purpose and goals



Thought Experiments

Jack

Jill

Do you agree with Jack or Jill?

1. Think.
2. Choose.
3. Write.
4. Share.
5. Discuss.

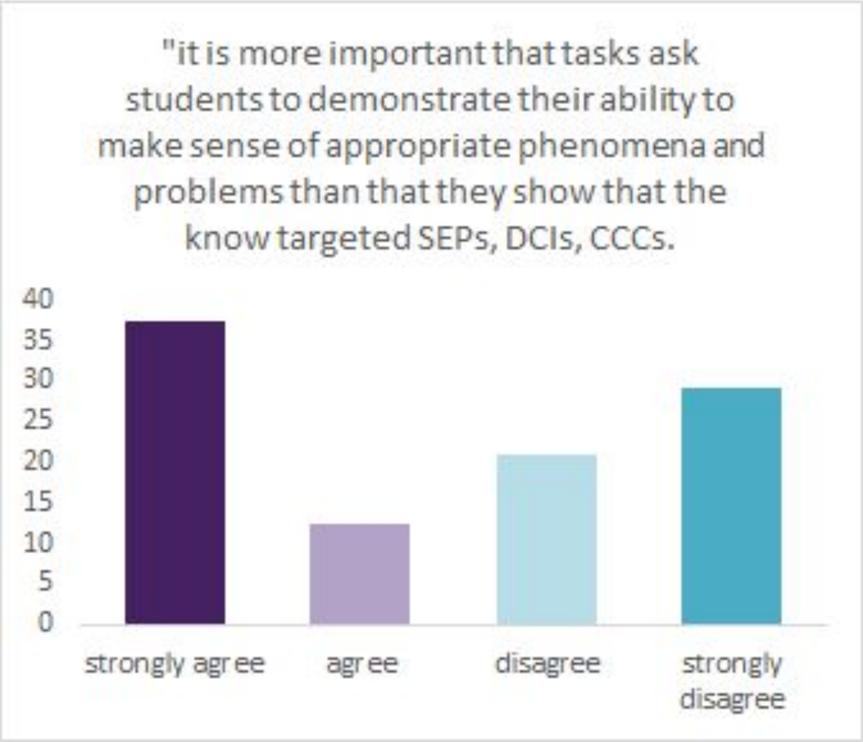
Focus of the assessment.

Points of Divergence: The focus of the assessment.

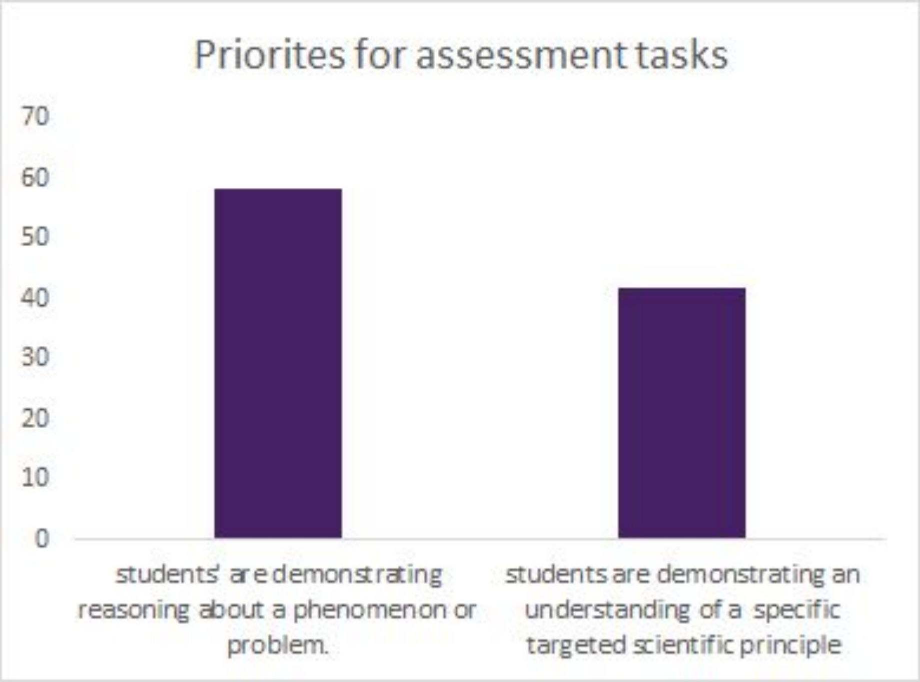
Jack	Jill
The focus of the assessment task should be on making sense of an authentic phenomenon or solving a problem, using parts of the DCIs, SEPs, and CCCs as needed.	The focus of the assessment task should be on determining whether students understand the targeted conceptual ideas and approaches--if a contrived scenario is the most direct tool to surface this, then that is acceptable.

Points of Divergence: The focus of the assessment.

Jack	Jill
Primary goal is assessing understanding of science principles.	Primary goal is assessing reasoning using the three dimensions.



"it is more important that tasks ask students to demonstrate their ability to make sense of appropriate phenomena and problems than that they show that the know targeted SEPs, DCIs, CCCs.
40
35
30
25
20
15
10
5
0
strongly agree
agree
disagree
strongly disagree



Priorites for assessment tasks
70
60
50
40
30
20
10
0
students' are demonstrating reasoning about a phenomenon or problem.
students are demonstrating an understanding of a specific targeted scientific principle

"I think students should figure out phenomena with SEPs and CCCs because that is what they will do outside of the classroom for the rest of their lives. DCIs, while important to an extent, should not be the primary focus of 3D assessments."

"I believe the deep learning comes through an emphasis on reasoning (via the practices and the CCCs). If students are successful at these they can reach the deep learning and figuring out of the phenomenon."

"While SEPs and phenomenon or problem-based contexts may help students better understand the science ideas, the purpose of science instruction and science assessments is to ensure students understand science ideas."

"Teaching and learning in science needs to prepare students for STEM careers long-term--this means that students need to understand the ideas in fields they might want to pursue."

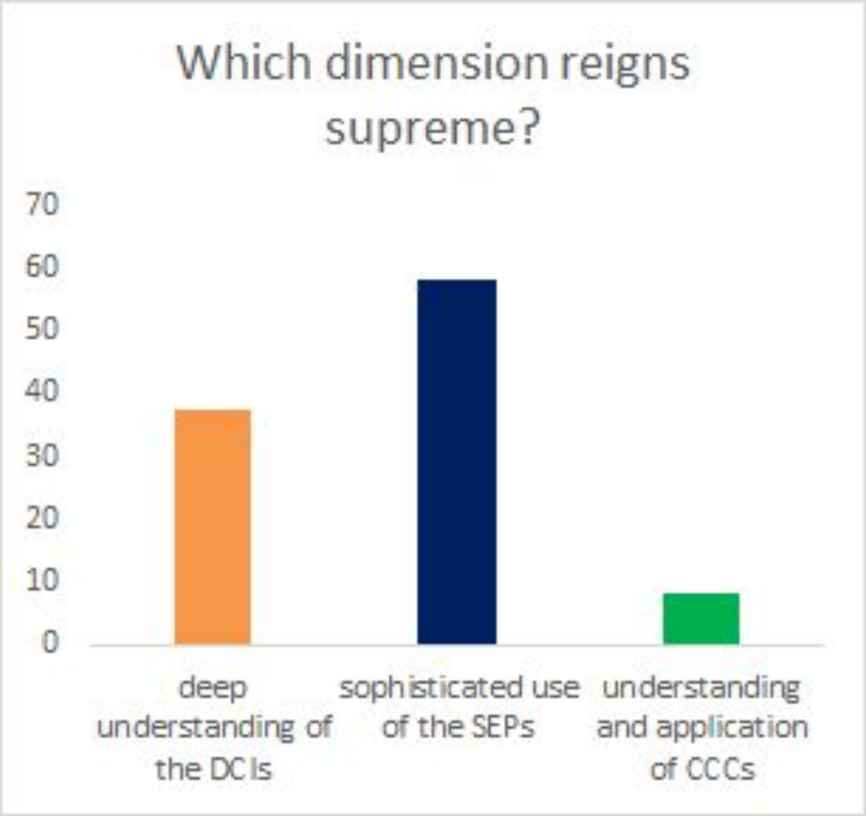
Points of Divergence: Relative priority of SEPs vs. DCIs

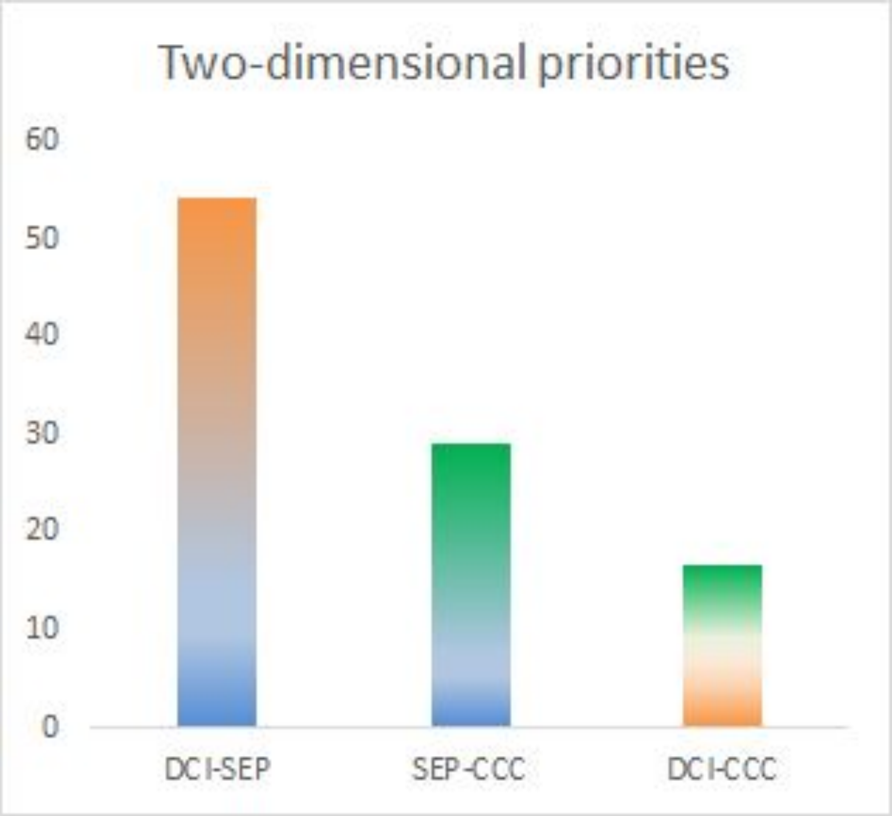
Jack

Assessments should include all three dimensions, but the most important aspect of student learning is their understanding of the DCIs because they are ideas needed to explain the world around them.

Jill

Assessments should include all three dimensions, but the most important aspect of student learning is their understanding and ability to use the SEPs because they are the ways students can explain the world around them.


Which dimension reigns supreme?
70
60
50
40
30
20
10
0
deep understanding of the DCIs
sophisticated use of the SEPs
understanding and application of CCCs

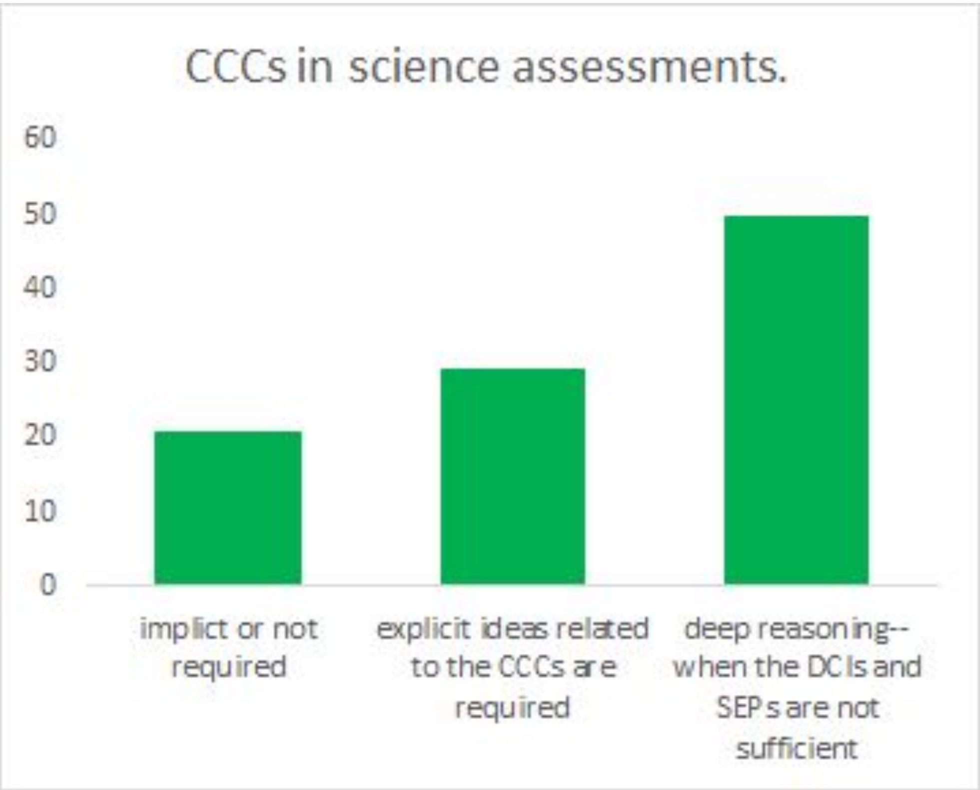

Two-dimensional priorities
60
50
40
30
20
10
0
DCI-SEP
SEP-CCC
DCI-CCC

CCCs in Assessment.

Points of Divergence: The Role of the CCCs.

Jack
It is imperative that assessments claiming to assess the NGSS or similar standards require students to demonstrate their understanding of one or more CCCs--CCC understanding must be necessary to respond to the task.

Jill
CCCs are always present when students are applying DCIs and SEPs--they do not need to be explicitly targeted in assessment.



CCCs were identified in tasks in 3 primary ways:

- CCCs were implicitly part of assessments, but understanding the concepts or crosscutting nature was not required to respond to the task.
- Explicit ideas related to the CCCs (e.g., what is a pattern?) are required to respond to the task, distinct from SEPs and DCIs
- DCIs and SEPs were necessary but insufficient to respond to the task successfully--application of CCCs was necessary.

Achieve

“The CCCs’ power is in how educators use them to connect to prior knowledge gained by the student from other classes or subjects. Them being assessed explicitly is not necessary.”

“One or more CCCs "fall out" of SEP use in the context of a DCI--it is not necessary for students to demonstrate a separate grasp of the CCCs. Such contextualized knowledge use has the potential to provide strong evidence that students have robust and flexible command of a discipline, but is not valuable in a vacuum.”

“CCCs are the single most important innovation of the NGSS--they are the connection to higher order thinking for ALL students, and not assessing them prevents us from signaling and supporting all students in developing the thinking skills they are capable of.”

“CCCs are likely the most transferrable ideas in the NGSS--they can help students approach situations and problems outside of science too. So we have to make sure they are developing them!”

Student Engagement

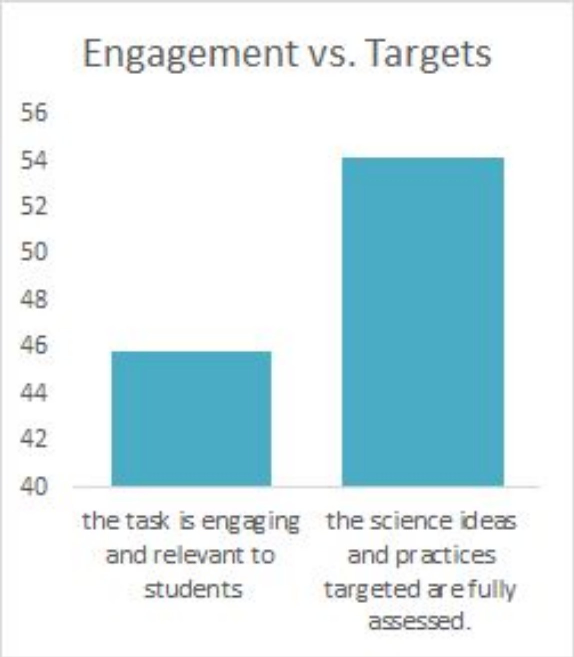
Points of Divergence: Student engagement.

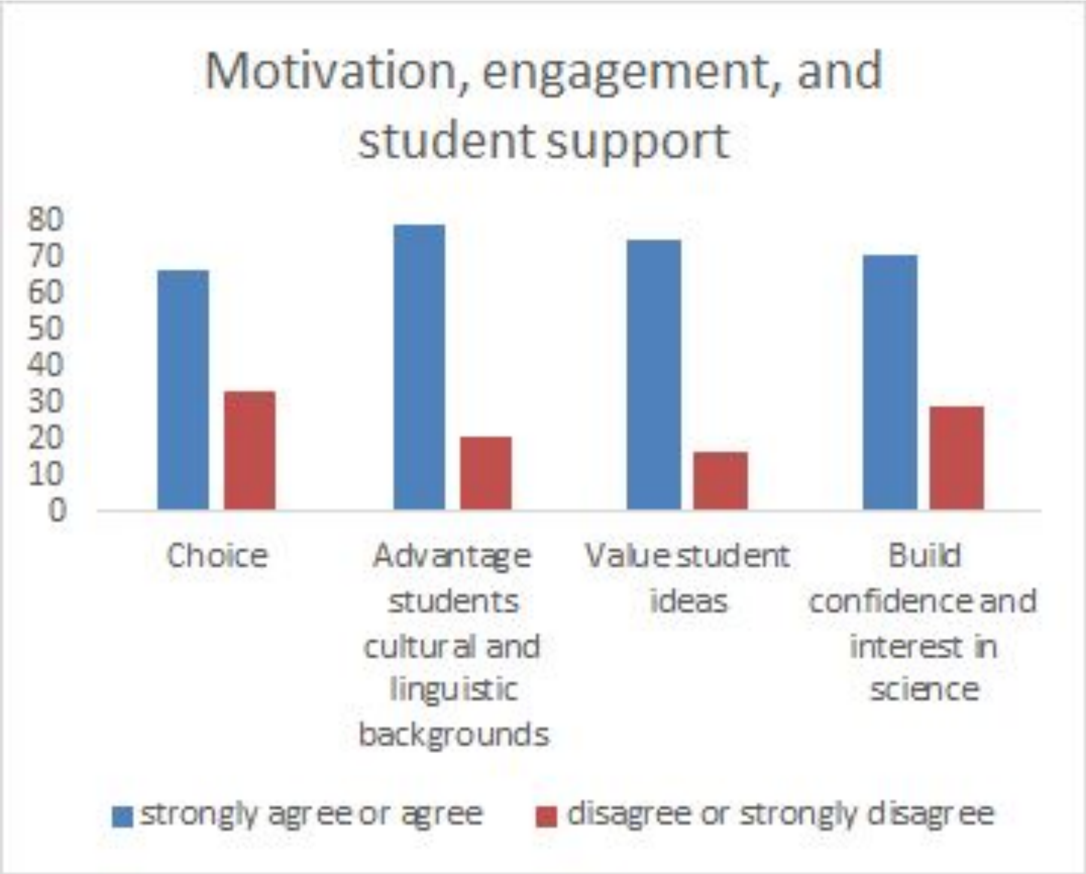
Jack
Scenarios should present a real, specific instance of a scientific phenomenon whose relevance can be made clear to students--even if students are only addressing a piece of the phenomenon/problem.

Jill
Scenarios should be designed to elicit the targeted performance, but they do not need to be relevant to students. Relevance should not required of a scenario to motivate student responses and is not a critical feature of NGSS assessments.

Points of Divergence: Student engagement.

Jack	Jill
To appropriately motivate and surface student thinking, it is important that tasks are relevant and engaging (e.g., including choice, valuing student ideas, respect and advantage students' cultural and linguistic backgrounds).	While features related to student engagement are important in instruction, assessments do not need to attend to them because there are other ways to motivate students inherent to assessment (e.g., they are required; grades).


Engagement vs. Targets
56
54
52
50
48
46
44
42
40
the task is engaging and relevant to students
the science ideas and practices targeted are fully assessed.


Motivation, engagement, and student support
80
70
60
50
40
30
20
10
0
Choice
Advantage students cultural and linguistic backgrounds
Value student ideas
Build confidence and interest in science
strongly agree or agree
disagree or strongly disagree

"In the end, tasks measure what a student knows and can do. Having rich task that keeps the student engaged and motivated and empowered is nice, but accurately assessing the student is the main goal of the assessment. DCIs and SEPs should be the focus. DCIs are obviously very important because they are reflective of the science content knowledge we expect students to have."

"It is not the responsibility of assessments to play a social justice role. Science facts are inherently unbiased--focusing on empirical science ideas is the best way to support students."

"All students won't be interested in all subjects in school, including science--it is unfair to expect assessment tasks to build confidence or attend to students' cultural backgrounds."

"If we want to support all students, it is critical that assessments actually provide meaningful feedback about student learning. To do so, they must ask all students to show what they do know and can do, and value a wide range of ways of knowing."

"If we don't design assessments that support all students, what's the point?"

Points of Divergence

Nature of Phenomena

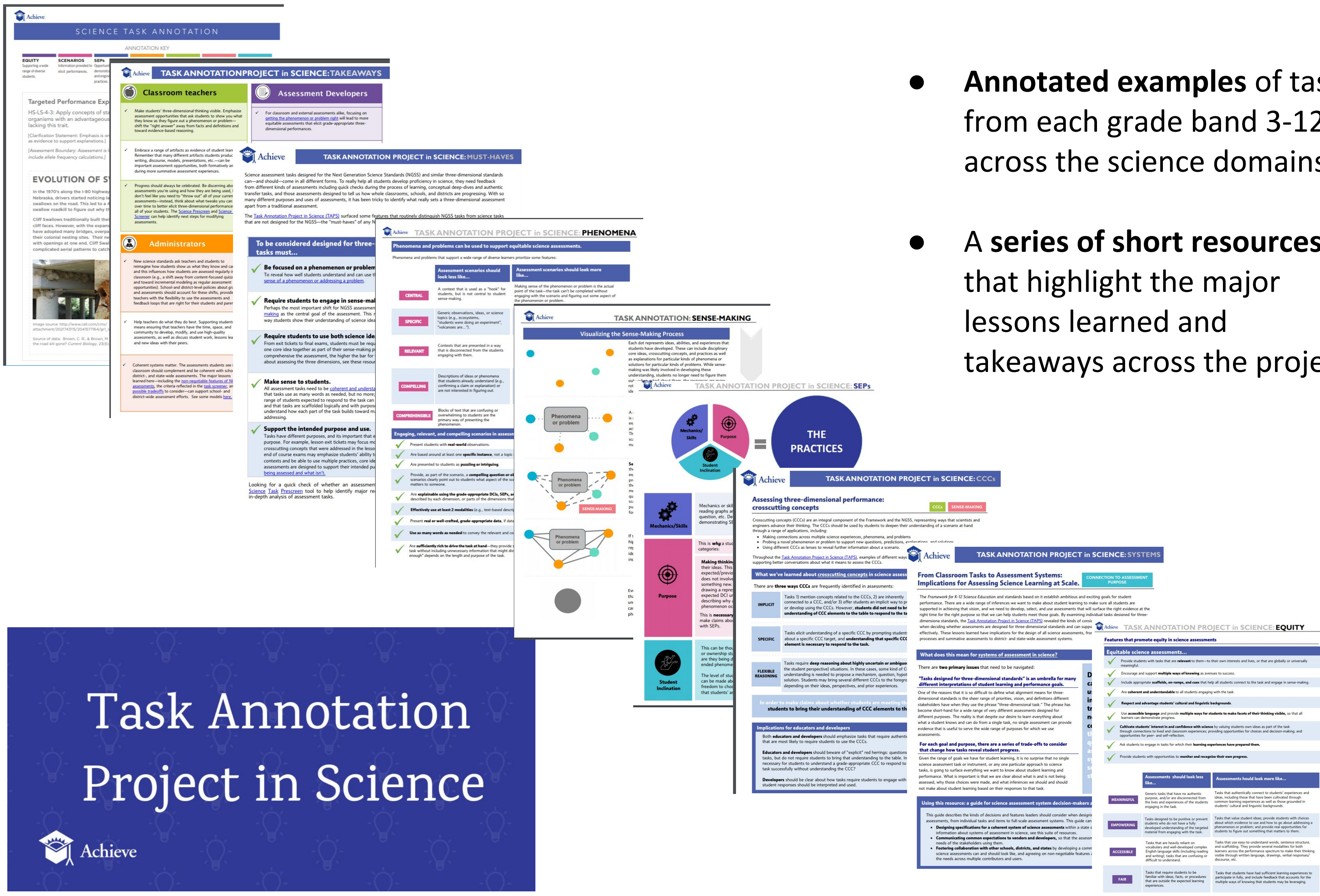
Scenarios and task prompts should support students in coherently and progressively making sense of a targeted phenomenon/problem.	Scenarios are tools to elicit the targeted dimensions and show how the dimensions can be used to make sense of phenomena. Coherence from the student perspective is not necessary.
Phenomena and problems should be specific such that they require students to address that specific instance.	Phenomena and problems must be based on real science but they can be general as long as students must use the targeted science principles to engage with it.

Points of Divergence

Equity

Agree: Tasks should be equitable and fair!

Tasks should present information in multiple modalities	Tasks need to use accessible language
Tasks should use as few words as possible while still being comprehensible	Tasks need to use accessible language
Tasks should allow students to respond in formats other than writing	Tasks may include formats other than writing as part of the response but should also require written responses.



- **Annotated examples** of tasks from each grade band 3-12 across the science domains

- A **series of short resources** that highlight the major lessons learned and takeaways across the project.

Thank you!

Symposium paper: bit.ly/narsttaps

TAPS resources: bit.ly/taskannotationprojectinscience

This project was funded through the generous support from Chevron and the Hewlett Foundation.