

2021

Analysis of assessments on secondary students' development and interpretation of models

Alaina L. Appley
University of Northern Iowa

Let us know how access to this document benefits you

Copyright ©2021 Alaina L. Appley

Follow this and additional works at: <https://scholarworks.uni.edu/grp>



Part of the [Science and Mathematics Education Commons](#), and the [Secondary Education Commons](#)

Recommended Citation

Appley, Alaina L., "Analysis of assessments on secondary students' development and interpretation of models" (2021). *Graduate Research Papers*. 1758.

<https://scholarworks.uni.edu/grp/1758>

This Open Access Graduate Research Paper is brought to you for free and open access by the Student Work at UNI ScholarWorks. It has been accepted for inclusion in Graduate Research Papers by an authorized administrator of UNI ScholarWorks. For more information, please contact scholarworks@uni.edu.

Analysis of assessments on secondary students' development and interpretation of models

Abstract

As districts are making the shift to three-dimensional learning the development of a coherent set of high-quality task-based assessments has been a challenge. For this research I collected and analyzed twelve of my district's assessments over the scientific skill of modeling from grades seven through 12. The analysis involved two tools developed by NGSS and Achieve.org to determine the extent to which the assessments ask students to perform tasks that are driven by phenomena and use the three-dimensional in service of sense-making, the *Task Screener* and the *Framework to Evaluate Cognitive Complexity in Science Assessments*. The findings support what researchers have said about the shift to three-dimensional task-based assessments: Choosing appropriate engaging phenomena is key to developing high-quality rigorous assessments. While most of my district's modeling assessments were found to be three-dimensional they are not rigorous because the phenomena guiding the tasks are too general and not puzzling.

**Analysis of Assessments on Secondary Students' Development and Interpretation of
Models**

Alaina L. Appley

University of Northern Iowa

This creative component was completed in partial fulfillment for a Master of Arts in Science
Education
May, 2021

This Paper by: Alaina L. Appley

Entitled: Analysis of Assessments on Secondary Students' Development and Interpretation of Models

has been approved as meeting the non-thesis requirement for the Degree of Master of Arts

Date: April 14, 2021 Dr. Dawn Del Carlo., Advisor

Date: April 14, 2021 Dr. Jody Stone., Outside Reader

Abstract

As districts are making the shift to three-dimensional learning the development of a coherent set of high-quality task-based assessments has been a challenge. For this research I collected and analyzed twelve of my district's assessments over the scientific skill of modeling from grades seven through 12. The analysis involved two tools developed by NGSS and Achieve.org to determine the extent to which the assessments ask students to perform tasks that are driven by phenomena and use the three-dimensional in service of sense-making, the *Task Screener* and the *Framework to Evaluate Cognitive Complexity in Science Assessments*. The findings support what researchers have said about the shift to three-dimensional task-based assessments: Choosing appropriate engaging phenomena is key to developing high-quality rigorous assessments. While most of my district's modeling assessments were found to be three-dimensional they are not rigorous because the phenomena guiding the tasks are too general and not puzzling.

Table of Contents

| | |
|---------------------------------------------------------------|-----------|
| | 4 |
| Abstract | 3 |
| Chapter 1. Introduction and Framework | 5 |
| Research Questions | 7 |
| Chapter 2. Literature Review | 9 |
| Next Generation Science Standards | 9 |
| The Science and Engineering Practices | 10 |
| Modeling | 11 |
| Assessments | 13 |
| Three-dimensional Assessments | 14 |
| Rigorous and Relevant Phenomena | 15 |
| Coherence | 17 |
| Theoretical Framework | 18 |
| Chapter 3. Project | 20 |
| Achieve.org Task Screener | 21 |
| Rigor and Cognitive Complexity | 23 |
| Survey | 26 |
| Chapter 4. Results and Conclusions | 27 |
| Results | 27 |
| Three-dimensional Alignment | 27 |
| Changing Rigor Through the Grades | 30 |
| Challenges in Developing Assessments | 33 |
| Discussion/Conclusion | 34 |
| Future Work | 36 |
| References | 38 |
| Appendix A- IRB Approval | 42 |
| Appendix B- NGSS Task Screener (Modified to fit space) | 43 |
| Appendix C- Task Screener Full Data Set | 45 |
| Appendix D- Survey Responses | 47 |

Chapter 1. Introduction and Framework

The development of high-quality assessments that meet the Next Generation Science Standards (NGSS) definition of three-dimensions is a challenge that all science teachers are facing. The shift from traditional tests to phenomena-based tasks that address the Science and Engineering Practices (SEPs), Crosscutting Concepts (CCCs), and Disciplinary Core Ideas (DCIs) is a monumental undertaking for many. There is a need for high-quality assessments that focus on these three-dimensions for teachers to use as a guide. The research and development of these tasks is going to take the nation time, and currently assessments for seven through 12th grades that focus on students' development and interpretation of scientific models are lacking. Even more exigent for a district is to develop these assessments to fit the scope and sequence for their courses as students progress through middle and high school. This new reform in science education is about designing learning and assessment environments for students that are grounded in evidence, coherent, and well coordinated at a larger scale than ever before (Pea & Collins, 2008).

The hope for the development of NGSS is to educate an American workforce with the knowledge and skills to meet “many of humanity’s most pressing current and future challenges,” (NRC, 2012). Its goals are reflected in its coherent K–12 structure and the development of the three-dimensions. Students develop the skills (like scientific modeling), the knowledge, and the appreciation for science to guide them into their adult lives. A strong education can teach students how scientists solve problems and how they can develop similar critical thinking skills to apply to their lives and decisions (Sadhu & Laksono, 2018). Teachers must weave these skills,

concepts, and crosscutting concepts into every aspect of their lesson planning, from instruction to assessment (Morrell et al., 2020).

My district is unique in that we adopted the Science and Engineering Practices (SEPs) from the NGSS as our district's course standards. There are seven science practices from which each course instructor must select three to five to focus on in their assessments. *Developing and Using Models* is the only scientific practice selected in all courses. I am interested in improving the quality of our assessments that ask students to engage in this skill. The purpose of focusing on just scientific modeling is that my science department has agreed to develop best-practice assessments on this topic for fall 2021. We have students memorizing models and being asked to replicate them on our assessments all too often. We need to move away from fact-driven assessments to skill-based tasks (Debarger et al., 2017). This qualitative research project will involve collecting data to determine what high-quality assessments are being used in my district and to provide a foundation for improving our future work.

Furtak (2017) analyzed many articles from this current wave of reform and found that as teachers learn about what the new assessments should look like there are many dilemmas that must be overcome. First, teachers lack training about the interrelationships among the three dimensions of NGSS, especially since most of their career they have assessed inquiry practices separately from content understanding or did not assess it at all. Additionally, teachers need help learning and developing the types of tasks that can truly help surface student thinking, and researchers are struggling to do this themselves (Debarger et al., 2017; Furtak, 2017; Pea & Collins, 2008; Shepard, 2000). To aid researchers in developing a pool of vetted assessments this research will help researchers to see what is currently being done in the classrooms. New

tasked-based assessments developed by teachers generally are adaptations of old tests, but they often fail to be three-dimensional if teachers are not reflective and supported (Debarger et al., 2017). The purpose of this research is to present and improve my district's progress towards the goals set forth by the NGSS. It will also provide model assessments over the SEP "developing and using models" for other teachers to use as a guide. The work of my district can help improve three-dimensional assessments nationwide.

Research Questions

1. How do my district's assessments, which ask students to develop and use models, align with the three-dimensional learning qualities of the NGSS?
2. How do the assessments on modeling change in rigor through the grades 7-12?
3. What challenges do teachers face when developing three-dimensional tasked-based assessments for the science classroom?

To answer question one, I used the instrument by NGSS and Achieve.org called the *Task Screener*, that is available for teachers to analyze their assessments to determine if they meet the definition of high quality (NGSS Lead States, 2013d). I examined twelve currently developed assessments from our district's science teachers that focus solely on the practice of scientific modeling, while asking students to engage in all three dimensions. I then used the *Framework to Evaluate Cognitive Complexity in Science Assessments*, also by Achieve.org to answer question two and help develop a model for our teachers of what complexity can look like for students when developing and using models (Achieve, Inc., 2019). The shift in our district has been very stressful for us as we adopt the NGSS and change our grading practices. For the last question, I

surveyed the teachers to determine what challenges they face and reflect on how this research can aid in improving my district's three-dimensional science assessments.

Chapter 2. Literature Review

Next Generation Science Standards

In 2011, the National Research Council published a research paper entitled *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas* that is perceived by many as having laid the groundwork for the shift to three-dimensional learning in the classroom that we are seeing in this current wave of education reform (Ames, 2014). Teachers and researchers in the decade since have been reworking and redeveloping their curriculum and assessments to meet the goals put forth by this council. These goals included encouraging an appreciation for science, development of a knowledge base needed to be educated voters and consumers, and development of skills needed to enter the workforce and solve future world problems.

The Lead States (2013) developed performance expectations (PEs) that change the approach to classroom instruction and assessment for all K–12 science teachers. They focus across what they refer to as the three-dimensions: Science and Engineering Practices (SEPs), Crosscutting Concepts (CCCs), and Disciplinary Core Ideas (DCIs). Then they laid out what three-dimensional classroom instruction should look like. Teachers need to be engaging students in investigating and designing experiments from their experiences to increase engagement and learning (NRC, 2012). The result of their work is a new set of science standards called the Next Generation Science Standards (NGSS) (Lead States, 2013). The NRC (2012) also concluded that, “instructional resources are key to facilitating the careful sequencing of phenomena and design challenges across units and grade levels in order to increase coherence as students become

increasingly sophisticated science and engineering learners” (p. S- 3). However, these resources are still missing for many units and teachers are struggling to create assessments that align with this goal.

The Science and Engineering Practices

The Science and Engineering Practices (SEPs) are one of the three-dimensions presented in the NGSS and are the skills or practices that all scientists engage in. These eight components of scientific inquiry are practices such as Asking Questions, Developing and Using Models, Planning and Carrying out Investigations, Analyzing and Interpreting Data, Using Mathematical Thinking, Constructing Explanations, Communicating Information, and Engaging in Argumentation. These should not be taught in isolation. For example, the practice of argumentation is interconnected to students’ development of an investigation procedure and their analysis of the data. Students can even engage in argumentation about the use and development of models.

The NGSS lays out specific capabilities associated with each science practice. Willard (2020) organizes what students should be able to do at each level of learning by tracking each skill through a logical progression flow chart. For example, K-2 students can analyze data from tests to determine if it works as intended. Then, by Grades 3-5 students can use the data to refine the testing process and by Grades 6-8 begin to define operational ranges. Finally, by grades 9-12 they should be able to analyze data to optimize it relative to a certain criteria for success (Willard, 2020). These capabilities are outlined to encourage student growth and coherence of the practices and to give teachers structure to their curriculum development.

Teachers have a strong role in the development of student skills, from demonstrating the need for the skills to developing authentic situations where students must use these skills. These opportunities can give students the conceptual tools to show their current understanding of phenomena (Lesh & Lehrer, 2003). For example, when students problem-solve through the revision process of developing and reorganizing their models, they communicate their internal understanding better for teachers to react to. To become competent thinkers, students must not only engage in the SEPs, but also know when to apply them (Herman, 1992). Applying these practices to science phenomena is a critical step in their development. My district developed our standards around these practices. We wrote out learning targets and success criteria for each of the practices for students as they progress through grades 5-12.

Modeling

A major skill and concept in the NGSS is scientific modeling. It is categorized as a crosscutting concept (CCC): “defining the system under study—specifying its boundaries and making explicit a model of that system—provides tools for understanding and testing ideas that are applicable throughout science and engineering” (NGSS Lead States, 2013b, p. 1). It is also a SEP that asks students to develop their own models and use models to develop explanations (NGSS Lead States, 2013a). This research will focus solely on the second use of modeling as a verb as my district has chosen to coordinate and improve our assessment over this skill in our professional development time.

Models will be defined here as “a means for representing aspects of the world” (Giere, 2009, p. 7). Models serve many purposes to the scientist, and students should be exposed to many types and opportunities to work with models. Models can show the relationship between

systems, act as a tool for finding patterns, or serve as an explanation for a phenomenon. Teachers need to give students opportunities to analyze and revise their models. A common model in chemistry is the Bohr Model of the atom. This model explains the spectrum of hydrogen, but what other power does it have? Can a student also see why other models may better serve to explain the bonding nature of the atom? Modeling becomes a language that allows scientists to communicate ideas, and teachers must open this window for students (Bent, 1984; Coll & Lajium, 2011). When students develop models and are guided in their reflection on the process, they are better able to develop higher-level explanations of phenomena (Cheng & Brown, 2015).

It is important for teachers to give students the opportunities to use models just as scientists do. This looks different across the age groups. Young students should be able to compare models, distinguish between models and the actual object or process, and develop their own models to represent something about the natural world. As their brains develop they should be able to revise models based on evidence and identify limitations in models. By middle school students can develop models to describe phenomena and generate data becoming more aware of how the model is different from the target event. Working through this process will allow students by high school to develop more complex models and use them as a tool to think more abstractly and interweave concepts and phenomena (Willard, 2020). Like a scientist, they can use models as a device that can be manipulated and improved upon to better explain and predict natural events (Cheng & Lin, 2015).

Modeling looks different across the science disciplines and content. The mathematical model is a good example of a type of model used heavily in physics, some in chemistry, but is more rare in biology, environmental, and earth science. On the other hand, biology and earth

sciences make use of cycle models to analyze the flow of energy and matter through systems. Chemists use symbolic and particle models to visualize and predict the unseen nature of matter. Engineers often use models to plan, test, and improve solutions to problems. Exposing students to modeling across curriculum is a big step in getting students to understand the nature of science as a field and as a practice.

Assessments

The purpose of assessments is to measure a student's understanding of content or application of a skill. My district's changes in summative assessments are driven by a desire to improve instruction to ensure all students can achieve at a high standard. We want our assessments to give us the data we need to support and teach our students. Assessments can support quality changes in school when the focus is on skill development (Herman, 1992). The skill development my department is focusing on is developing and using scientific models. To develop quality assessments, we must consider how students use the skill of modeling and ensure we move beyond just memorizing important information. For example, our summative assessments should go beyond asking students to memorize and label the model of a cell. Good assessments help us become better teachers by giving us the data we need to address our students' gaps in content and skill.

There are two types of assessments: formative and summative. Formative assessments are those collected either informally or early on in the learning process. My district has spent the last five years developing a system of formative assessments and reflection process. Therefore, for the purpose of this project my focus is instead on the quality of our summative assessments. These summative assessments are formal opportunities to show student learning that are

collected by teachers as pieces of evidence that will be used to determine a student's course grade. They include classic paper/pencil tests, laboratory reports, and projects, but all should encompass the ideals of the three-dimensions.

Much of learning is constructed from students' experiences, and this must be considered while developing assessments. High quality classroom assessments provide opportunities for students to think like a scientist and construct and demonstrate their own understanding. Cheng and Lin (2015) asked students to create their own models to explain everyday phenomena, and only 12% were able to describe more than just observable features. The authors' work developed a five-level definition for depth of scientific models (Cheng & Lin, 2015). Most students in their study were only able to use models at level one, which means they only used a model to describe observable phenomena. This aligns with what I have seen when I ask students to generate their own models; they struggle to use models to visualize the unseen and are unable to explain observed phenomena at a higher level. A similar study examined students' ability to construct scientific explanations, another Science and Engineering Practice. It focused on the effectiveness of different methods teachers use to scaffold assessments. The authors concluded that it is necessary for teachers to provide students with effective scaffolding on assessments, especially the use of contextualized phenomena, to promote rigor (Kang et al., 2014). This will have implications on my district's assessments. If we do not give our students the resources to engage with the material, they will not be able to demonstrate their skills.

Three-dimensional Assessments

Task assessments designed for the NGSS must target a full range of content and skills. "They must test students' understanding of science as a content domain and their understanding

of science as an approach. And they must provide evidence that students can apply their knowledge appropriately and are building on their existing knowledge and skills in ways that lead to deeper understanding of the scientific and engineering practices, crosscutting concepts, and disciplinary core ideas,” (NRC, 2012, p. 263). Task-based assessments will be defined as those that are focused around a phenomenon with students using language and critical thinking skills to engage with the questions. In science, students may have to solve a problem by collecting or analyzing data. They may need to develop a model to answer questions or organize an idea. For example, students in chemistry may need to interpret PES (photoelectron spectroscopy) data to predict and explain the bonding behavior between two atoms.

Achieve.org created a task screener designed to be used by stakeholders to 1) “determine whether classroom assessment tasks are high quality, designed to elicit evidence of three-dimensional performances, and designed to support the purpose for which they will be used, and 2) to provide a group of reviewers with a common set of features to ground conversations about what it “looks like” for students to demonstrate the kinds of performances expected by three-dimensional standards” (NGSS Lead States, 2013d, p. 1). Nationwide teachers and test developers can have a common language and goal for high quality three-dimensional assessments.

Rigorous and Relevant Phenomena

“To ensure that students are capable of taking on the challenges of tomorrow, investing in a rigorous and focused science education is critical” (Loney, 2014, p. 1). Rigor is defined as the presence of cognitively complex tasks that ask students to engage with phenomena using the three-dimensions. Science teachers are in a unique position to help students develop the

analytical and problem-solving skills needed in all careers. My district values a rigorous set of courses for students and wants students engaging in sensory experiences in which they participate in the processes of scientific discovery. New rigorous assessments created under this goal must draw from all three dimensions and be grounded in phenomena that students engage with by actively trying to figure out the way the world works (sense-making). High cognitive complexity comes from non-routine use of the DCI while authentically using multiple SEPs in service of sense-making (Achieve, Inc., 2019).

The assessments created under the new NGSS should be based on relevant and rigorous phenomena. Achieve.org defines high quality scenarios as those that address a rich and puzzling phenomenon or problem with a high-degree of uncertainty (Achieve, Inc., 2019). My colleagues and I have struggled the most with this aspect of assessment development. We are not alone in this. Studies have found that it is not easy to find and develop tasks that are equally challenging across different groups of students (Penuel et al., 2019; Settlage & Jensen, 1996; Shavelson et al., 1991). Choosing appropriate phenomena that is engaging to “students of different ethnicities, genders, and linguistic backgrounds” and “demanding enough to match the expectation of the standard” is rare even with careful consideration (Penuel et al., 2019, p. 1391). As a district, we must consider course phenomena and the corresponding assessment phenomena with care. I expect many of our current assessments are lacking rigor because they fail to get students to use models as scientists do. We want to assess if students can transfer the modeling skill to all types of scientific observations. However, not all content areas are equally amenable to modeling, and collecting assessments from all subject areas will highlight if we are seeing students’ transfer of this skill.

Coherence

There is much value at looking at assessments over the course of several sequential grades (e.g. grades seven through 12). It makes sure all teachers are on the same page and that students are developing their knowledge by building on prior experiences and becoming more sophisticated learners. Penuel et al. (2018) stresses that the teaching of these standards requires a cluster of task items over multiple assessments and years. Development of a horizontally and vertically coherent assessment system contributes directly to the quality of implementation of the NGSS by being a guide by which teachers best prepare students to meet the performance expectations put forth by the state (Penuel et al., 2018; Pruitt, 2014). This research is intended to help my district build an assessment system that reflects both the goals of our district and the goals of NGSS by first building a coherent framework for assessing the practice of modeling in science.

Our district started the process of analyzing the scope of our curriculum two years ago with our shift in grading practices and the adoption of the Iowa Core. *Scope* is defined as the range of learning targets that need to be prioritized within each unit of study (Ediger, 1999). What we choose to focus on as a district must be significant, useful, current, and backed up by multiple sources. Ediger (1999) set out guidelines for how a district should undergo a scope study in science. I chose to focus this study on the skill of scientific modeling because our science department grades five through 12 chose to focus on this skill development first. We also all agreed that developing and using models is a unique and important skill for our students to practice and be assessed on in our classrooms. Hopefully over time, this can become a model for

us and other districts hoping to develop a coherent scope for the new NGSS assessment framework.

Theoretical Framework

The shift to three-dimensional assessments presented by the NGSS is grounded in what we are learning about how students learn best. The traditional foundation for assessments was built on a social efficiency model that was driven by need and the attempt to measure achievement. Learning was about trying to lay a foundation and then systematically accumulate the knowledge that was necessary and possible for a child to learn. As a result most curriculum was disconnected from the real-world and assessments were fact driven and standardized (Shepard, 2000).

Learning opportunities must be contextualized and engaging for students to develop science literacy and interest. Vygotsky's research and theory of social constructivism are more relevant than ever. Science learning is guided by social-constructivist views that suggest that learning is something to be participated in (Shepard, 2000; Vygotsky, 1978). The brain does more than just accumulate facts; learning is an active process that is socially and culturally determined. Students should be asked to develop explanations and models and argue from evidence they are invested in. These practices are not separate from the content; they are a means for students to develop an understanding of the world around them. Even without school, learners develop models of the natural world as they engage with it and teachers must be aware of children's learning patterns to best develop curriculum (Inhelder & Piaget, 1958; Karplus, 1977).

Assessments should be more than just measurements of learning, they can be tools to support and enhance the learning itself. For science assessments to be compatible with social-constructivist models the questions must probe important thinking and practices. These assessments must also be embedded through the learning cycle to make them authentic (Shepard, 2000). Measuring students' thinking requires an understanding of the complexity framework for engaging in sense-making (Vygotsky, 1978). This is why grounding assessments in phenomena are so important, the construction of knowledge is tied closely to the students surroundings and interaction with their senses. The three-dimensional task-based assessment not only gives teachers feedback, but can improve learning and mastery.

Chapter 3. Project

Researchers and teachers need quality examples of scientific modeling tasks that are appropriate for each grade level. The data collected from this research highlights what qualities of NGSS three-dimensional learning my district's assessments are meeting and missing. The approach to data collection mirrored the research questions. First, after collecting a variety of modeling assessments from my district's grade 7-12 science teachers they were analyzed using two resources created by Achieve.org to help teachers in the process of developing quality three-dimensional assessments. The first tool, *Task Screener* (NGSS Lead States, 2013d) generated the data needed to determine the alignment of our district's assessments that ask students to develop and use models with the three-dimensional qualities desired by the NGSS (Appendix A). The second tool the *Framework to Evaluate Cognitive Complexity in Science Assessment* (Achieve, Inc., 2019) guided my second research question: How do the assessments on modeling change in rigor through the grades? It looked for how students were asked to make sense of phenomena. Finally, a survey was sent to the district's middle and high school science teachers to answer my last research question: What challenges are teachers facing when developing three-dimensional tasked-based assessments for the science classroom?

To collect the assessments I contacted, by email, my district's science teachers in seventh through 12 grades. I have worked closely on developing all of the ninth grade chemistry assessments and the 11th grade physics assessments; however, I had not seen many examples from other coworkers. I supplied the teachers with the *Pre-Screener* developed by Achieve (NGSS Lead States, 2013c) and then requested two quality assessments for each course that

focus on students' skill in developing and using scientific models. This *Pre-Screener* ensured that I received work that is intended to meet my district's and the NGSS's goals of three-dimensional tasks. All the teachers in my district have been at least informally trained on the NGSS, but none had used the Achieve set of screeners. The *Pre-Screener* was available for them to help them choose assessments from their prior work that have the most characteristics of phenomena-based tasks. It was not possible to keep anonymity from me regarding who wrote each test due to the fact that my district has only one teacher per grade. I received a total of 12 assessments over a wide variety of disciplinary core ideas from the science courses in my district that all students are required to take: 7th grade science, 8th grade science, chemistry 1, physics 1, earth science 1, and biology. I made a copy of each assessment and deidentified the author and assigned each a number. All methods were approved by the University of Northern Iowa Internal Review Board (Appendix A).

Achieve.org Task Screener

The *Task Screener* is a tool designed to help teachers collaborate in creating better assessments that are “driven by high-quality scenarios that focus on phenomena or problems” (NGSS Lead States, 2013d). I used the *Task Screener* to analyze how well our district's assessments are grounded in phenomena and the degree to which they meet the goal of integrating the three-dimensions (Figure 1). This screener (see Appendix B for full screener) walked me through a series of criterion including:

- 1: Tasks are driven by high-quality scenarios that focus on phenomena or problems;
- 2: Tasks require sense-making using the three dimensions;
- 3: Tasks are fair and equitable;

4: Tasks support their intended targets and purpose.

Figure 1:

Four Criterion for Evaluating Tasks Designed for Three-Dimensional Standards.

| | |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>A. Tasks are driven by high-quality scenarios that are grounded in phenomena or problems.</p> <ul style="list-style-type: none"> 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. | <p>B. Tasks require sense-making using the three dimensions.</p> <ul style="list-style-type: none"> 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: <ul style="list-style-type: none"> 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. |
| <p>C. Tasks are fair and equitable.</p> <ul style="list-style-type: none"> 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. | <p>D. Tasks support their intended targets and purpose.</p> <ul style="list-style-type: none"> 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. |

Note: (NGSS Lead States, 2013d)

I created a spreadsheet and numbered each assessment 1-12 (Appendix C). They were randomly ordered for this tool to ensure the authors remained anonymous. From the *Task*

Screener I chose to focus mainly on criterion A and B, but I also collected data on criterion C in case there was some information of interest to the research question. Data was not collected on criterion D because it requires knowledge of how the teacher used the assessments in class and it did not address my research question. I then took each assessment through the screener recording my answers to the prompts. I also considered what improvements could be made to help edit the assessment to be higher quality.

The organization of this data included first highlighting each incidence of adequately meeting the criteria for high quality. I then broke down each criterion to a series of yes/no qualities to clearly provide evidence on why the assessments were or were not adequate. Criterion A focused on whether the scenario presented real-world phenomena, was specific, presented students with a need-to-know puzzle, and if it was relevant to students. Criterion B was focused on how the assessment asked students to engage in the three-dimensions and was therefore broken down into asking students to demonstrate the DCIs, SEPs, CCCs, integrate multiple dimensions, and make their thinking visible.

Rigor and Cognitive Complexity

Next, I analyzed each of the assessment tasks submitted using the *Framework to Evaluate Cognitive Complexity of Science Assessments*; a tool designed for teachers to determine the rigor and degree to which students are engaged in science practices and sense-making (Achieve, Inc., 2019). This rubric (see Appendix B for full rubric) is focused around two essential questions:

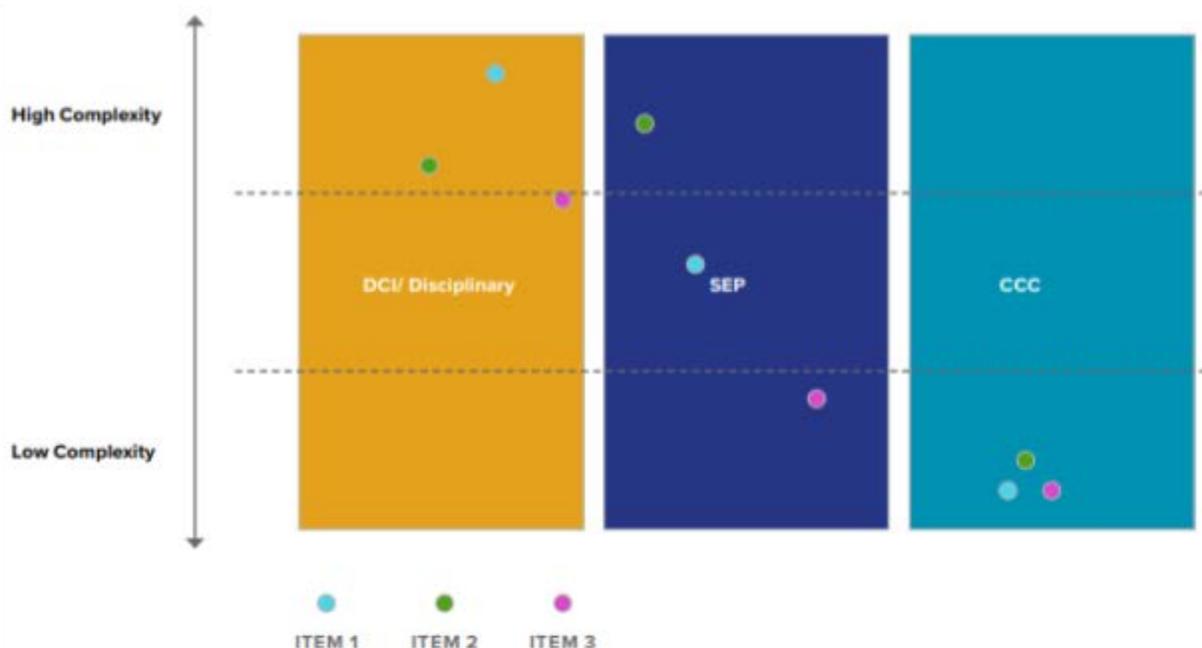
1. To what degree does the task ask students to engage in sense-making?
2. In what ways does the task ask students to use each dimension in service of sense-making?

This science cognitive complexity framework tool is divided into two phases. First, each task is scored on a five point scale with a score of five or four meaning it contains high use of students' understanding needed to contribute to sense-making and a one or two being indicative of a familiar or simple task for students to complete (see Table 1). All three-dimensions and the phenomena were carefully considered. The framework provides a rubric to complete this task. The second phase takes a more holistic look at the task the students are being asked to do. Its guidance allows for a judgment on if the students were at the top (being asked to do science with very little scaffolding) or at the bottom (with a scripted task that does not require very much reasoning) in terms of cognitive complexity. However, there is no value judgment as assessments should vary in complexity throughout the year and low scores should not be viewed as bad. The final analysis created a distribution of items as seen in Figure 2. I then looked at the spread of assessment's complexity level across the grades.

Table 1*Individual Item Analysis*

| | Scenario | SEP | DCI [Conceptual - disciplinary] | CCC [Conceptual - crosscutting] |
|--------|-------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------|
| High | Addressing a rich and puzzling phenomenon or problem presented with high-degree of uncertainty. | Figuring out a phenomenon or problem using multiple SEPs in service of authentic sense-making. | Non-routine use of domain specific science ideas as part of sense-making. | Selection and use of conceptual understanding of crosscutting ideas is necessary and expands students' thinking. |
| Medium | Addressing a phenomenon or problem with some level of uncertainty. | Representation of ideas; use of skills that are relatively complex; some close application. | Supported application of science ideas in typical contexts. | Specific crosscutting concept understanding is needed and is used to focus students' thinking. |
| Low | Addressing routinely encountered or highly simplified scenarios. | Using the mechanics, skills, and specific knowledge associated with practices isolated from sense-making. | Producing previously-learned ideas and conceptual procedures in routine, well-practiced ways. | Crosscutting concepts are implicitly part of the task, but they are not required in service of sense-making. |

Note. (NGSS Lead States, 2019)

Figure 2*Distribution of Items Example*

Note. (NGSS Lead States, 2019)

Survey

For the final phase, I emailed the teachers three open-ended survey questions to gain insight into what challenges they face when trying to develop three-dimensional task-based assessments.

1. When developing assessments that include the scientific practice of modeling, what difficulties do you face in making them three-dimensional?
2. How do you consider what phenomena or problems to include in the assessment? Are they similar to scenarios students see in class or are they new for the assessment?
3. Do you find it difficult to make modeling assessments rigorous? Explain.

I color-coded the answers to look for trends in struggles that can guide future professional development and help make sense of the results from the Achieve resources. I used deductive reasoning and coded using specific keywords including CCC and rigor, but I was open to other patterns of concerns that teachers had about the process.

Chapter 4. Results and Conclusions

Results

The project was designed to answer questions about my district's assessments that I can share with others struggling with the same challenges. The data collected was plentiful and helpful in showing the strengths and weaknesses of the current assessments that focus on the topic of scientific modeling. It also provided a helpful starting point to improve these assessments going forward. The two tools, the *Task Screener* and the *Framework to Evaluate Cognitive Complexity in Science Assessments*, worked well together to build a clear picture of the quality of my districts modeling assessments and the tasks that guide these assessments. There were very few surprises during the analysis, but overall there is evidence for a lot of consistency in how we are having students engage in scientific modeling.

Three-dimensional Alignment

To answer the question regarding how well our assessments on the skill of modeling align with the three-dimensional qualities of NGSS I first created a spreadsheet from the main questions and big ideas I found in the *Task Screener*. There were four main sections or criteria: A. Was the assessment task driven by high-quality scenarios that focus on phenomena? B. Does the task require sense-making using the three dimensions? C. Were the tasks fair and equitable? D. Does the task support their intended targets and purpose? For each criteria I chose a few guiding points that focused on collecting evidence towards identifying if the task met the success of the criteria. I chose to eliminate criteria D. because I was unsure of each assessment's target

and purpose. For most questions I answered with a yes or no or some qualifying statement. I also offered suggestions for improvement in the moment as I was considering the quality of the tasks.

The first analysis looked at the overall number of assessments that met the criteria of adequate (Table 2). Only one assessment task was driven by a high-quality scenario and this assessment's focus was actually on modeling in an engineering rather than scientific context. Nine of the twelve assessments met the criteria for sense-making using all three-dimensions. Only three were adequately fair and equitable; mostly because they didn't offer any student choice. This meant only one assessment met all three qualities.

Table 2

Summary of Task Screener Results

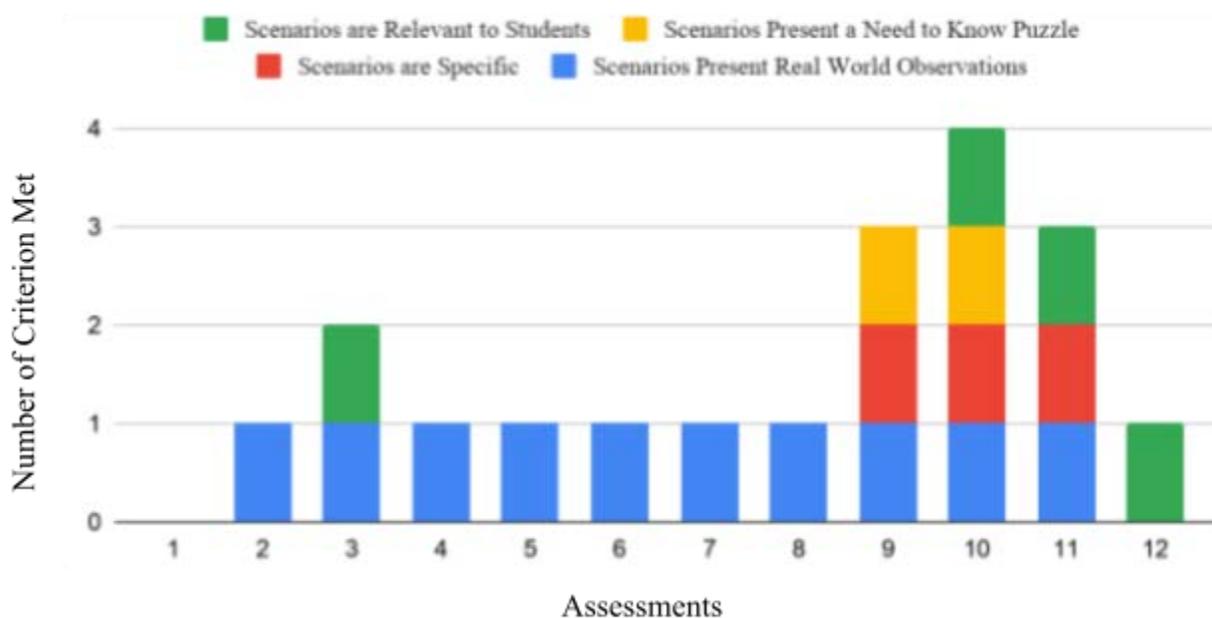
| Quality | Assessments meeting criteria of adequate (12) |
|------------------------------------------------------------------|-----------------------------------------------|
| 1: Task driven by high-quality scenarios that focus on phenomena | 1 |
| 2: Task requires sense-making using the three dimensions | 9 |
| 3: Tasks are fair and equitable | 3 |

I wanted to look deeper into the data to see what qualities the assessments were missing. I first focused on how the tasks in the assessments were or were not driven by high-quality scenarios. The criteria included having scenarios that were relevant to students, present a need-to-know puzzle, be specific, and they must present a real world observation (Figure 3). While all but one of the assessments were based on the real world observation, there was much missing. Only four scenarios were relevant to the students and three of the twelve were specific. The remaining were too abstract, for example the discussion of bonding was about finding atoms their 'best date'. Or they were too general; giving a DNA sequence to develop a model to explain

protein synthesis with no specific connection or storyline. High-quality scenarios should contain a ‘need to know puzzle’. This was present in only two tasks and is an easy way we can improve our assessments.

Figure 3

Task Screener Criterion A: Scenario Guiding the Task

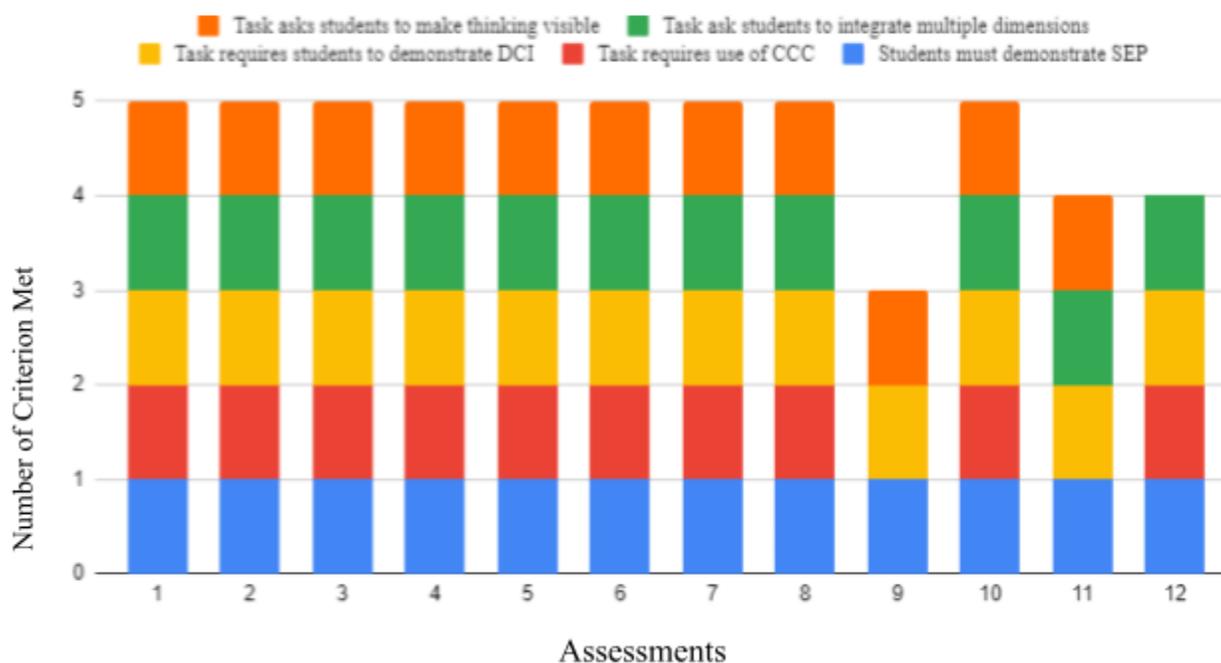


A big part of this research is seeing how well my district’s assessments are utilizing the three-dimensions. To meet this criterion the tasks must ask students to make their thinking visible, integrate multiple dimensions and demonstrate each of the three dimensions: DCI, CCC, and SEP (Figure 4). All of the tasks contained quality use of the science and engineering practices and disciplinary core ideas, which was not a surprise since I specifically asked for assessments that focused on tasks that asked students to develop and interpret scientific models. Nine of the twelve assessments had tasks that met all the criteria. While only two of the

assessments were missing a CCC, I felt that this area could be improved by changing some vocabulary in the tasks to include specific references to the CCCs.

Figure 4

Task Screen Criterion B: Utilizing the Three-Dimensions



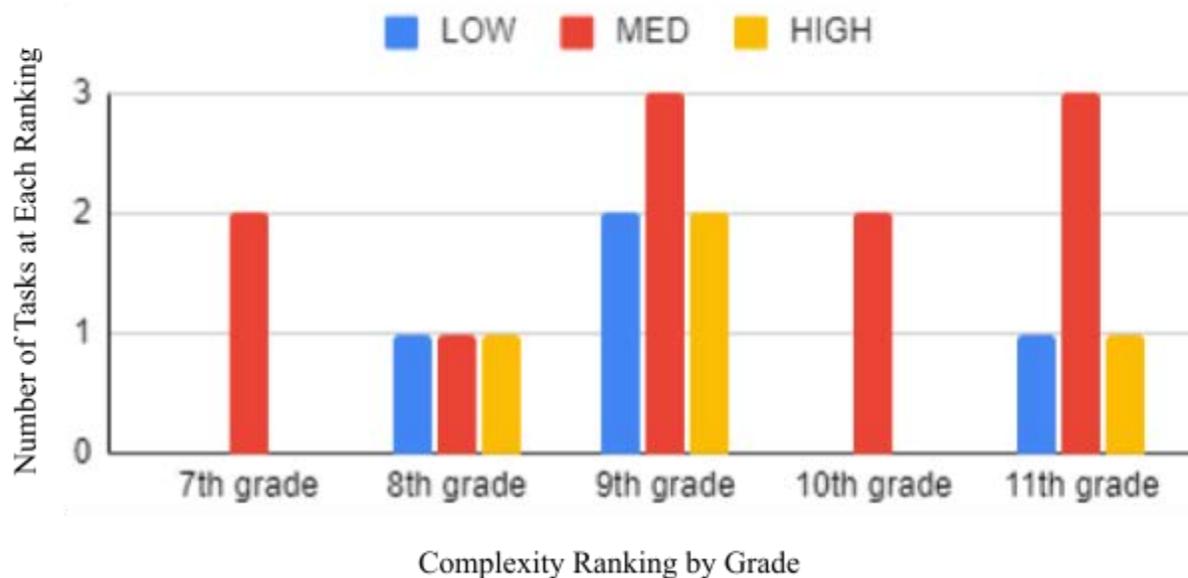
Changing Rigor Through the Grades

The second tool I used to analyze the same twelve assessments was the *Framework to Evaluate Cognitive Complexity in Science Assessment*. I took each assessment through the tool by ranking them on a scale of one to five on the level of complexity for the scenario, SEPs, DCIs, and CCCs. I kept Table 1 visible to me so that I could see the definition of each category and best determine the degree to which students were engaging in sense-making. Half of the assessments had multiple parts to them so that I was able to analyze 19 different task items.

For my first analysis I looked just at the complexity of the modeling that students were engaging in by grade. Fifteen of the nineteen task items asked students to use relatively complex modeling skills with a few close applications. This scores as a medium complexity or lower. For example, one task had students create a particle model (ticker-tape motion) for a biker, but never had them use that model to explain anything. Only four asked students to figure out a phenomenon by using multiple SEP's in service of sense-making (a level four or higher). A high-quality example task had students use their model of a balloon car they created to explain how Newton's laws applied to their design and its success. There was little to no pattern across the grades, as each teacher created assessments ranked on average a medium level of complexity (Figure 5).

Figure 5

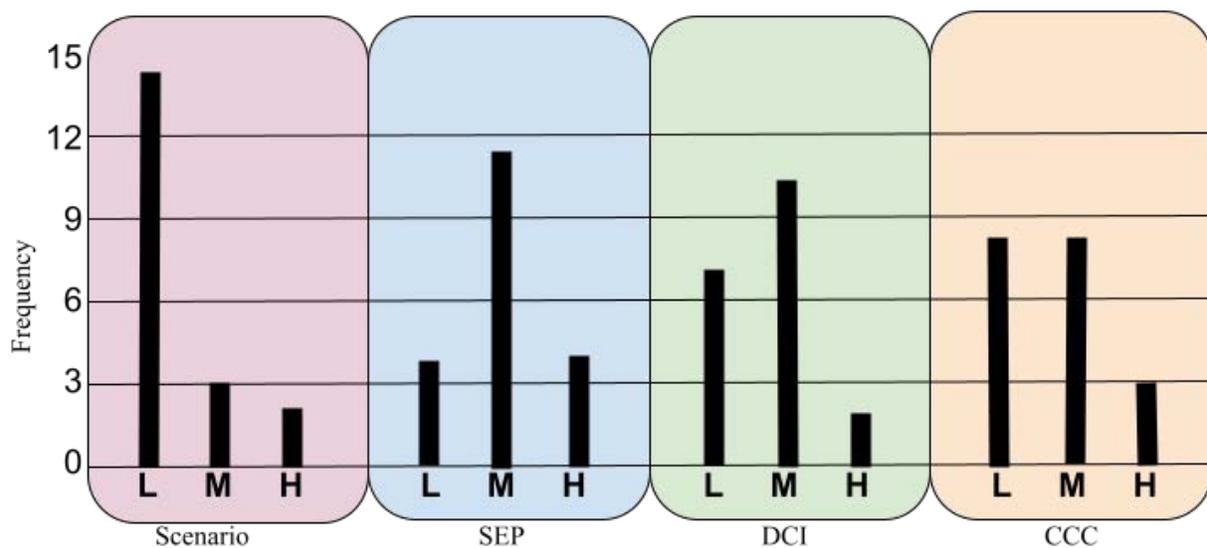
Complexity of Modeling (SEP) Task



Since the complexity did not increase throughout the grades I decided to look at our overall complexity in each category by tabulating the number of low, medium, and high scores for each category (Figure 6). This made it clear that the biggest weakness we have in our rigor actually comes from the scenario that is driving the task. This is similar to what I saw in the previous analysis. This really highlights that quality three-dimensional integration is based on engaging students in rich and puzzling phenomena that have a high degree of uncertainty. Our scenarios/phenomena are rarely rigorous and therefore we are failing to develop consistently high quality assessments.

Figure 6

Complexity Ranking by Dimension



Number of assessment items (19 total) at each complexity level (low, medium, high)

Challenges in Developing Assessments

The teachers involved in the assessment design process gave insights into the challenges in developing assessments that include the scientific process of modeling. I emailed each of the four teachers that supplied the tests some questions. I collected these in a table to code (Appendix D). I first read through everyone's responses, then I started coding any mention of CCCs the color teal. Two teachers mentioned that it was challenging to include the CCCs authentically as a connection to the modeling. See Table 3 for full coding totals.

One focus of this research was to analyze the rigor of the assessments and I asked a specific question about rigor and coded every mention of it red. Teachers seem to struggle with this balance. They want it to be authentic, but fear it will make the questions too challenging or too unfocused for students (coded in blue). Some topics seem to be harder than others, but everyone is aware and seems to want more rigorous assessments. They just don't know how to go about developing them.

The second question focused on how the teachers choose phenomena to guide their assessments. All four teachers mention not straying far from the scenarios they use in lessons; usually asking students to take it a step further or changing some portion of it. One mentioned that they use examples from well-known experts. I know that we often use phenomena that we were familiar with in the past, prior to the release of the NGSS. After the analysis of the twelve assessments, this is an area where we can devote more training. The *Task Screener* showed that much of what guides are tasks are too general and not very puzzling.

Table 3*Coding of Survey Responses*

| Code | Times mentioned as a challenge |
|-----------------------------------------------------|--------------------------------|
| CCCs- Cross-cutting concepts (teal) | 2 |
| Rigorous (red) | 4 |
| Questions written that are focused and clear (blue) | 2 |

Discussion/Conclusion

After the analysis of twelve assessments I was able to see a wide variety of modeling tasks. The data collected supports that we are doing a good job at getting students to engage in scientific modeling in conjunction with the disciplinary core ideas. We have made a strong shift towards including the SEPs in all of our assessments and are engaging students in a variety of modeling experiences. Students are analyzing real-world models and developing models following scientific principles. We also give them the opportunity to make their thinking visible and discuss the CCCs. We have included so many aspects of three-dimensional assessments and have students completing a variety of modeling tasks that mirror classwork. This supports what Herman (1992) says about the shift to more contextualized assessments improving assessments.

The biggest weakness of our assessments is the use of phenomena to guide the tasks. Both the *Task Screener* and the science cognitive complexity data support this claim. This research will form a strong foundation for improving our assessments. As students progress through our courses over the next few years we need to make adjustments to these tasks to get them engaged in modeling like scientists do. Many of the guiding phenomena are too general and do not present a need-to-know puzzle. I think this is because they are too much like our old

assessments in that they use just the classic scenarios we have seen and used our whole careers. This supports what Furtak (2017) says: there will be cascading challenges as we shift to three-dimensional assessments. We need to start building off of our students' interest and be open and flexible to new ideas. Keeping in mind the qualities of good phenomena we can build tasks that are more conducive to sense-making.

Before this research I had a misconception about rigor. I was equating rigor with complexity. However, rigor can be achieved at all grades. It's about choosing phenomena that gets students to engage in grade-appropriate sense-making. The relationship between good phenomena and rigor was made clear to me during this research. The use of the modeling to understand real-world puzzles is rigor. This supports Penuel's findings that even with careful consideration finding good phenomena is challenging (2019).

I and my colleagues have been concerned about the open-ended nature of task-based assessments, but the data suggests that this is not what we need. High quality assessments ask students to apply modeling to phenomena that is *specific* and relevant. This will improve both the rigor and quality of our assessments, while addressing concerns over giving students too much freedom. We need to switch our focus from the questions we are asking to the driving phenomena that guides the tasks. This is the key. I suggest starting with work from Penuel and the University of Colorado Boulder, where new resources are being released.

Teachers need more examples of high quality science assessments. The process of adopting new standards and changing assessments to reflect the three-dimensional nature of science has been challenging for everyone. This research shows teachers what type of challenges they may face when writing task-based assessments over the skill of modeling. Modeling is not

the biggest challenge; instead it is making assessments rigorous and based in puzzling phenomena. I suggest more collaborative work between teams to select specific and relevant phenomena. Brainstorming between colleagues and resource building can help switch the focus to careful initial selection of guiding phenomena. I hope they also can see the value in using the set of Achieve.org tools to effectively analyze their current assessments.

Future Work

This research provided a lot of data about my district's modeling assessment that can be used to improve those assessments and help create stronger assessments in the future. If other districts repeat this research it would be interesting to see if they are experiencing similar challenges. A broader look across districts could also be reflective of the method of change that is occurring. Do most districts focus on teaching the SEPs or have more made the switch to being strongly phenomena-based but are struggling with engaging students in the scientific practices?

Additional research over assessments on other scientific skills would be helpful. For example, our department has been discussing the situations where we encourage argumentation in the classroom and specifically the types of tasks students do. The presentation of strong phenomena can improve our assessments that focus on modeling as well as those that focus on developing explanations and argumentation. I would like to collect all of the assessments that ask students to develop arguments and run them through the *Framework to Evaluate Cognitive Complexity in Science Assessments* and map the items phenomena in chronological order to look for consistency and flow. Then research what pattern of complexity is most supported by data and then rewrite our assessments to better reflect what the research says.

Looking at the work of teachers from seventh through 12th grade has given me insight into how students are going to develop over the coming years. Coherent assessment systems contribute directly to high-quality adoption of new curriculum, especially when teams work together (Penuel et al., 2019). When I see the younger students in four years they will have had more experience modeling and I need to adjust my assessments to reflect that. The work by Willard (2020) needs to be a reference document that we reflect on often as students gain more and more experience in the three-dimensional system of task-based assessments. As the teacher of the oldest students in our district my expectations need to grow with them.

Much of the value of the *Task Screener* is in the meaningful discussions it can lead to with colleagues about the quality of the tasks. I would like to lead a group of science teachers through this process, specifically with regard to engaging students in scientific argumentation. The pairing with the *Framework to Evaluate Cognitive Complexity in Science Assessment* worked well to see the role the phenomena plays in the quality of the tasks. I think many teachers could develop better tasks with less stress if they better understood what it means to engage students in sense-making and how the phenomena could guide that.

References

- Achieve, Inc. (2019, September 26). *A Framework to Evaluate Cognitive Complexity in Science Assessments*. Achieve. <https://www.achieve.org/cognitive-complexity-science>
- Ames, R. T. (2014). A review of science standard history culminating with Next Generation Science Standards. *Journal of Education and Training, 1*(2), 48–57.
<https://doi.org/10.5296/jet.v1i2.5292>
- Bent, H. A. (1984). Uses (and abuses) of models in teaching chemistry. *Journal of Chemical Education, 61*(9), 774–777. <https://doi.org/10.1021/ed061p774>
- Cheng, M. F., & Brown, D. E. (2015). The role of scientific modeling criteria in advancing students' explanatory ideas of magnetism. *Journal of Research in Science Teaching, 52*(8), 1053–1081. <https://doi.org/10.1002/tea.21234>
- Cheng, M. F., & Lin, J. L. (2015). Investigating the relationship between students' views of scientific models and their development of models. *International Journal of Science Education, 37*(15), 2453–2475. <https://doi.org/10.1080/09500693.2015.1082671>
- Coll, R. K., & Lajium, D. (2011). Modeling and the future of science learning. In M. S. Khine & I. M. Saleh (Eds.), *Models and Modeling: Cognitive Tools for Scientific Enquiry* (pp. 3–21). Springer Netherlands. https://doi.org/10.1007/978-94-007-0449-7_1
- Debarger, A. H., Penuel, W. R., Moorthy, S., Beauvineau, Y., Kennedy, C. A., & Boscardin, C. K. (2017). Investigating purposeful science curriculum adaptation as a strategy to improve teaching and learning. *Science Education, 101*(1), 66–98.
<https://doi.org/10.1002/sce.21249>

Ediger, D. M. (1999). *Scope and sequences in science*. (ED436393). ERIC.

<https://files.eric.ed.gov/fulltext/ED436393.pdf>

Furtak, E. M. (2017). Confronting dilemmas posed by three-dimensional classroom assessment:

Introduction to a virtual issue of Science Education. *Science Education*, 101(5), 854–867.

<https://doi.org/10.1002/sce.21283>

Giere, R. N. (2009). Why scientific models should not be regarded as works of fiction. In

Fictions in science (pp. 248–258). Routledge. New York.

Herman, J. L. (1992). What research tells us about good assessment. *Educational Leadership*,

49(8), 74–78.

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence:*

An essay on the construction of formal operation structures. (Vol. 22). Psychology Press.

Kang, H., Thompson, J., & Windschitl, M. (2014). Creating opportunities for students to show

what they know: The role of scaffolding in assessment tasks. *Science Education*, 98(4),

674–704. <https://doi.org/10.1002/sce.21123>

Karplus, R. (1977). Science teaching and the development of reasoning. *Journal of Research in*

Science Teaching, 14(2), 169–175.

Lesh, R., & Lehrer, R. (2003). Models and modeling perspectives on the development of

students and teachers. *Mathematical Thinking and Learning*, 5(2–3), 109–129.

<https://doi.org/10.1080/10986065.2003.9679996>

Loney, E. (2014). Making the case: The importance of a rigorous science education. *The Hunt*

Institute, No. 05 (ED559394). ERIC. <https://files.eric.ed.gov/fulltext/ED559394.pdf>

Morrell, P. D., Park Rogers, M. A., Pyle, E. J., Roehrig, G., & Veal, W. R. (2020). Preparing

- teachers of science for 2020 and beyond: Highlighting changes to the NSTA/ASTE standards for science teacher preparation. *Journal of Science Teacher Education*, 31(1), 1–7. <https://doi.org/10.1080/1046560X.2019.1705536>
- NGSS Lead States. (2013a). *Next Generation Science Standards: For states, by states (Appendix F)*. Washington DC: The National Academies Press. <https://doi.org/10.17226/18290>.
- NGSS Lead States. (2013b). *Next Generation Science Standards: For states, by states (Appendix G)*. Washington DC: The National Academies Press. <https://doi.org/10.17226/18290>
- NGSS Lead States. (2013c). *Next Generation Science Standards: For states, by states (Task Pre-Screener)*. Washington DC: The National Academies Press. <https://doi.org/10.17226/18290>
- NGSS Lead States. (2013d). *Next Generation Science Standards: For states, by states (Task Screener)*. Washington DC: The National Academies Press. <https://doi.org/10.17226/18290>
- NRC. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. National Academies Press.
- Pea, R., & Collins, A. (2008). Learning how to do science education: Four waves of reform. In *Designing Coherent Science Education* (Vol. 3, pp. 1–12).
- Penuel, W. R., Frumin, K., Van Horne, K., & Jacobs, J. K. (2018, April 13-17). *A phenomenon-based assessment system for three-dimensional science standards: Why do we need it and what can it look like in practice*. [Paper presentation]. Annual Meeting of the American Educational Research Association 2018, New York, NY.
- Penuel, W. R., Turner, M. L., Jacobs, J. K., Van Horne, K., & Sumner, T. (2019). Developing

- tasks to assess phenomenon-based science learning: Challenges and lessons learned from building proximal transfer tasks. *Science Education*, 103(6), 1367–1395.
<https://doi.org/10.1002/sce.21544>
- Pruitt, S. L. (2014). The Next Generation Science Standards: The Features and Challenges. *Journal of Science Teacher Education*, 25(2), 145–156.
<https://doi.org/10.1007/s10972-014-9385-0>
- Sadhu, S., & Laksono, E. W. (2018). Development and validation of an integrated assessment for measuring critical thinking and chemical literacy in chemical equilibrium. *International Journal of Instruction*, 11(3), 557–572.
- Settlage, J., & Jensen, M. (1996). Investigating the Inconsistencies in College Student Responses to Natural Selection Test Questions. *The Electronic Journal for Research in Science & Mathematics Education*, 1(1). <https://ejse.southwestern.edu/article/view/7553>
- Shavelson, R. J., Baxter, G. P., & Pine, J. (1991). Performance Assessment in Science. *Applied Measurement in Education*, 4(4), 347-362. https://doi.org/10.1207/s15324818ame0404_7
- Shepard, L. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4–14.
- Vygotsky, L. (1978). Interactions between learning and development. In *Readings on the Development of Children* (pp. 29–36). Worth Publishers.
- Willard, T. (2020). *The NSTA atlas of the three dimensions*. National Science Teaching Association.

Appendix A- IRB Approval

Dear Investigator(s):

Your study, Analysis of Assessments on Secondary Students' Development and Interpretation of Models, has been approved by the UNI IRB through the review procedures authorized by 45 CFR 46.104, effective September 27, 2020. You may begin recruitment, data collection, and/or analysis for your project. You are required to adhere to the procedures and study materials approved during this review, as well as to follow IRB policies and procedures for human subject research posted on the IRB website.

If you need to make changes to your study design, samples, procedures, or study materials, please email todd.evans@uni.edu request approval of the changes before they are implemented, and attach any revised study materials with edits highlighted. You may expect a response within a couple of days.

Your study will not require annual review and approval by the IRB. However, you will receive an annual study update request, which will ask if the study is still active and if any problems have arisen. Advisors: If your student has graduated, please reply to the annual update request on the student's behalf.

If at any time you observe any problems or incidents that are serious and unexpected (e.g. you did not include them in your IRB materials as a potential risk), you must report this to the IRB within 10 days. Examples include unexpected injury or emotional stress for study participants, missteps in the consent process, or breaches of confidentiality. The IRB will advise on any next steps that might be necessary.

If you need a signed approval letter, contact the IRB office and one will be provided for your records.

Best wishes for your project success.

Todd Evans
IRB Chair

--

--

Todd A. Evans, PhD, ATC, LAT
Associate Professor, Athletic Training
University of Northern Iowa
Human Performance Center
2351 Hudson Road
Cedar Falls, IA 50614-0244
(319) 273-6152
todd.evans@uni.edu
Fax number: 273-7023

Appendix B- NGSS Task Screener (Modified to fit space)

Criterion A.

Tasks are driven by high-quality scenarios that are grounded in phenomena or problems.

| Tasks designed for the NGSS include clear and compelling evidence that: | What was in the task, where was it, and why is this evidence? | | | | |
|----------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------|--------------------------|--------------------------|-----------|
| i. Making sense of a phenomenon or addressing a problem is necessary to accomplish the task. | 1) Is a phenomenon and/or problem present? <div style="border: 1px solid #ccc; height: 20px; margin: 5px 0;"></div> 2) Is information from the scenario necessary to respond successfully to the task? <div style="border: 1px solid #ccc; height: 20px; margin: 5px 0;"></div> | | | | |
| 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 | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | |
| | Scenarios are based around at least one specific instance, not a topic or generally observed occurrence (e.g., observations related to a specific hurricane rather than "hurricanes" in general) | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | |
| | Scenarios are presented as puzzling/intriguing | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | |
| | Scenarios create a "need to know" | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | |
| | Scenarios are explainable using grade-appropriate SEPs, CCCs, DCIs | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | |
| | Features of scenarios | Yes | Somewhat | No | Rationale |
| | Scenarios effectively use at least 2 modalities (e.g., images, diagrams, video, simulations, textual descriptions) | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | |
| | If data are used, scenarios present real/well-crafted data | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | |
| | The local, global, or universal relevance of the scenario is made clear to students' | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | |
| | Scenarios are comprehensible to a wide range of students at grade-level | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | |
| | Scenarios use as many words as needed, no more | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | |
| | Scenarios are sufficiently rich to drive the task | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | |

Across all indicators, there is _____ evidence of quality of this criterion (choose one).

- No
 Inadequate
 Adequate
 Extensive

1. When considering whether the scenario creates a need to know for students, consider whether the scenario makes the uncertainty associated with explaining a phenomenon or solving a problem central, in ways that are likely to 1) connect with students' own experiences or knowledge, and 2) connect to disciplinary core ideas (regardless of whether those ideas are explicitly named or required by the task).

2. Consider whether an authentic stakeholder group is interested in the outcome of the scenario, and/or whether students are given enough information to answer the question "why should I care?"

Criterion B.
Tasks require sense-making using the three dimensions.

| Tasks designed for the NGSS include clear and compelling evidence that: | What was in the task, where was it, and why is this evidence? |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| i. Completing the task requires students to use reasoning to sense-make about phenomena or problems. | Consider in what ways the task requires students to use <i>reasoning</i> to engage in sense-making and/or problem solving. |
| ii. The task requires students to demonstrate grade-appropriate: <ul style="list-style-type: none"> • SEP element(s) • CCC element(s) • DCI element(s) | Evidence of SEPs (which element [s], and how does the task require students to demonstrate this element in use?) |
| | Evidence of CCCs (which element [s], and how does the task require students to demonstrate this element in use?) |
| | Evidence of DCIs (which element [s], and how does the task require students to demonstrate this element in use?) |
| iii. The task requires students to integrate multiple dimensions in service of sense-making and/or problem-solving. | Consider in what ways the task requires students to use <i>multiple dimensions together</i> to sense-make and/or problem-solve. |
| iv. The task requires students to make their thinking visible. | Consider in what ways the task <i>explicitly prompts</i> students to make their thinking visible. Look for evidence of how the task surfaces current understanding, abilities, gaps, and problematic ideas. |

Across all indicators, there is _____ evidence of quality of this criterion (choose one).

- No
- Inadequate
- Adequate
- Extensive

Suggestions for improvement of the task for Criterion B:

Appendix C- Task Screener Full Data Set

| Assessment Number | Level | Driving Phenomenon | Is the information in the scenario necessary to respond successfully to the task? | Scenarios Present Real World Observations | Scenarios are specific | and Present a need to know puzzle | Grade Appropriate | Use multi-modal | Relevant to students | Comprehensible to a wide-range of students at grade-level | Use as many words as needed, no more | Sufficiently Rich | Quality of this Criteria | Suggestions for improvement |
|--------------------|-------|----------------------------------|-----------------------------------------------------------------------------------|-------------------------------------------|------------------------|-----------------------------------|-------------------|-----------------|----------------------|-----------------------------------------------------------|--------------------------------------|------------------------|--------------------------|------------------------------------------------------------------------------|
| 1 | 9 | Elements are looking for a date. | Yes | No, Elements don't "date" | Too abstract | Somewhat, the search for a match | yes | Yes | No | Somewhat, Advanced isn't related to scenario | Yes | Parts are disconnected | Inadequate | Consider Real-life scenarios, minerals, sports drinks? |
| 2 | 9 | Nucleosynthesis | yes | kinda | Too abstract | no | yes | yes | no | somewhat, it's pretty hard | Yes | Yes | Inadequate | Move model 1 to top, Consider relating to paint colors? |
| 3 | 9 | River Images | Yes | Yes | too general | no | too easy | no | yes | need clear high level | yes | no | Inadequate | Consider an image of a local river |
| 4 | 9 | Convection Currents | yes | kinda | too general | no | yes | yes | kinda | yes | yes | Yes, but disconnected | inadequate | Maybe drive from last question forward |
| 5 | 10 | Protein Syn | yes | kinda | too general | no | yes | yes | kinda | somewhat, it's pretty hard | yes | yes | inadequate | Driving phenomena, Can you match a sequence to a portion of a common protein |
| 6 | 10 | Connect Photosyn and CR | Yes | kinda | too abstract | no | yes | yes | kinda | somewhat, it's pretty hard | maybe more needed | yes | inadequate | Consider using a specific flower to drive phenom |
| 7 | 11 | Motion | No, (teacher motion) | Yes | too many | no | yes | yes | kinda | yes | yes | yes | inadequate | Connect motion together |
| 8 | 11 | Moving Objects | Yes, computer needed | Yes | too many | no | Yes | yes | kinda | yes | yes | yes | inadequate | Reduce scenarios or have a more practical connection between each |
| 9 | 8 | Reflection and Refraction | yes | yes | yes | yes | yes | yes | some | yes | yes | no | inadequate | Relate scenario 1 to the lake |
| 10 | 8 | Balloon Cars Motino | yes | yes | yes | yes | yes | yes | yes | yes | yes | yes | adequate | good |
| 11 | 7 | Moon phases | No | yes | yes | no | yes | yes | yes | yes | more details needed | yes | inadequate | Clarify goal and design limitations in instructions |
| 12 | 7 | Ecosystem Energy Diagrams | No | unclear | no | no | yes | yes | yes | yes | more details needed | yes | inadequate | Clarify goal and design limitations in instructions |

| Assessment Number | Students must demonstrate SEP | Task requires use of CCC | Task requires students to demonstrate DCI | Task asks students to integrate multiple dimensions | Task asks students to make thinking visible | Quality of this Criteria | Suggestions for improvement |
|--------------------|-----------------------------------------------------------------------------|-------------------------------------|-------------------------------------------|-----------------------------------------------------|-----------------------------------------------------|--------------------------|-----------------------------|
| 1 | Develop and Interpret Models | Structure and function and patterns | Periodic table and bonding | Yes | Questions are well broken down into parts, asks why | adequate | |
| 2 | Develop and Interpret Models | Matter and Energy, | Nucleosynthesis | Somewhat, CCC is written as a conclusion | Yes, question broken down, #4, ask for explanation? | adequate | Give more guidance for ? 4 |
| 3 | Develop and Interpret Models | Unclear | erosion | No | Yes, asks why | inadequate | Cause and Effect or Pattern |
| 4 | Develop and Interpret Models and analyzing data | Good, multiple | earth's energy cycles | yes | yes | adequate | Make CCC more explicit |
| 5 | Develop and Interpreting Models and Develeoping Explanations | Structure and function and patterns | protein syn | yes | yes | adequate | Make CCC more explicit |
| 6 | Developing Models | Structure and Function | cR and PS | yes | yes | adequate | Make CCC more explicit |
| 7 | Develop and Interpret models | Scale Proportion and Quantity | Motion | Yes | Yes | adequate | Make CCC more explicit |
| 8 | Develop and Interpret Models, analyze data, math and computational thinking | System and System Models | Newton's Laws HS-PS2-1 | Yes | Yes | adequate | Make CCC more explicit |
| 9 | Develop models | unclear | Light Reflection/Refraction | no | yes | inadequate | Make CCC more explicit |
| 10 | Develop and interpret models | Scale Proportion and Quantity | Newton's motion | yes | yes | adequate | good |
| 11 | developing models | unclear | moon phases | yes | yes | inadequate | Make CCC more explicit |
| 12 | developing models | matter and Energy, | ecosystems | yes | no | adequate | good |

Appendix D- Survey Responses

| Questions | Teacher 1 | Teacher 2 | Teacher 3 | Teacher 4 |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1. When developing assessments that include the scientific practice of modeling what difficulties do you face in making them three-dimensional? | <p>1. Making sure they can represent the crosscutting concept. Many times I have them write a statement implementing the CCC. Also encouraging the students to find ways to model, not just write out the information in sentences.</p> | <p>The challenges I face when creating modeling assessments in biology is in the design. How do I make the problem challenging yet achievable for my lower level learners.</p> | <p>The challenge with modeling assessments is not having students go too far. If the questions are not focused, students tend to answer with all of the know;edge they have on a subject, whether it is relevant to the question or not.</p> | <p>Choosing rigorous phenomena and connecting a CCC authentically, Task-based is also hard, I struggle to find a puzzle/story to guide the reason for students to either develop a model or to interpret the model.</p> |
| 2. How do you consider what phenomena or problems to include in the assessment? Are they similar to scenarios students see in class or are they new for the assessment? | <p>2. When I am considering phenomena I look at what some of the experts have used (Paul Andersen), I also look at the main theme or final assessment. Sometimes they are similar scenarios and occasionally they tie into the final assessment. I am still growing and learning the 3D process so each year I try to improve my delivery and the assessments.</p> | <p>When looking at phenomena, I try to link it to scenarios we have discussed in class, but the specific examples are new situations. For example, if we discuss photosynthesis ideal conditions and practice those, on the assessment, they may need to extrapolate what would happen in an airtight container. We aspire to link their learning to situations they would encounter in real life.</p> | <p>I often do similar scenarios to what students do in class only with an extra component, like an explanation of how the model works or where we used it before. A lot of times the phenomena are developed weekly and students are aware of the big ideas for that week. Then we get tested about the stuff we do in the labs. I try my best to let students know what those big ideas/phenomena are ahead of time so that they can narrow their focus/</p> | <p>Always very similar to class, changing numbers/elements to create different scenarios. This is one of the hardest parts for me with the modeling standard. I can't think of ideas that are different from class. It's easy for argumentation and investigations.</p> |

| | | | | |
|-------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------|
| <p>3. Do you find it difficult to make modeling assessments rigorous? Explain.</p> | <p>3. There are a few topics, like water cycle, that made it difficult to make it rigorous. Like I said above I keep doing research and reading Science teaching information to improve each year with rigor, 3D, and delivery.</p> | <p>The challenge with modeling assessments is not having students go too far. If the questions are not focused, students tend to answer with all of the know;edge they have on a subject, whether it is relevant to the question or not.</p> | <p>I do. Especially with the Evidence Based Reporting, I just feel like I haven't had enough training or knowledge of the grading system in my collegiate studies. It is hard for me to come up with assessments that are rigorous but also applicable to what we are doing. I don't like to surprise students about what's on their tests, so that is part of the reason why coming up with that rigor can be a challenge for me. I am continuing to improve, especially in my Earth Science course. I do have Chem 1 to help me guide my ideas for assessments, I just have to put them in another language.</p> | <p>Yes, they often are very straightforward and uninteresting. I don't often change much from the formatives.</p> |
|-------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------|