Designing From Outer Space: Tensions in the Development of a Task to Assess a Crosscutting Concept

Erin Marie Furtak, University of Colorado Boulder, erin.furtak@colorado.edu
Torsten Binder, University of Duisburg-Essen, torsten.binder@uni-due.de
Kate Henson, University of Colorado Boulder, kate.henson@colorado.edu

Abstract: New visions of science learning that integrate disciplinary core ideas, scientific practices, and crosscutting concepts necessitate new approaches to assessment design. This paper documents the iterative design of an assessment task intended to trace the crosscutting concept of energy across three different disciplinary contexts in high school science. We review and synthesize literature on performance tasks, three-dimensional task design, and research into student thinking about energy. Then, working with examples from a research-practice partnership, we identify three tensions that have emerged in the process of design: tension between practice-as-embodied in the task and the current state of practice in our partner district, tensions in creating scorable outcome space as students model and explain energy across systems, and tension in asking students to represent energy in ways that pose challenges for disciplinary experts. We close by summarizing ongoing challenges for assessment designers engaged in designing assessments for crosscutting concepts.

Introduction

The introduction of the *Framework for the Next Generation Science Standards* [NGSS] changed the US vision of science teaching and learning from a two-dimensional, topics and skill-based approach to a three-dimensional approach. The new science teaching and learning standards foster thinking in the three dimensions: *science and engineering practices*, *core ideas* and *crosscutting concepts* (National Research Council [NRC], 2012). The *science and engineering practices* subsume cognitive, social and physical practices that are required to investigate and build theories and models about the natural world. The *core ideas* are key organizing concepts of single disciplines. These two dimensions of the NGSS are similar to those that have previously been documented as part of international inquiry-oriented science teaching reforms (e.g. OECD, 2017).

However, the *crosscutting concepts* in the NGSS are a new way of thinking about ideas that span all of the sciences, as they are of broad importance and have applications in all science domains (NRC, 2012). These concepts, which include systems and systems thinking, patterns, and cause and effect, can be applied to a wide range of phenomena, and are included in every one of the individual NGSS performance expectations. The crosscutting concepts necessitate not only new approaches to science teaching that help students to foreground and connect these overarching themes, but also new, multicomponent three-dimensional assessments which will be able to evaluate the teaching and learning of all three dimensions in the NGSS (NRC, 2014).

In this paper, we identify tensions that surfaced as we designed a three-dimensional performance task that foregrounded the crosscutting concept of energy in the context of a long-term, mutualistic collaboration between our research team and a large culturally and linguistically diverse school district. The task focuses on the *disciplinary core ideas* of "cycles and energy transfer in ecosystems", "chemical reactions" and "forces and motion", the *scientific practice* of "developing and using models" and "constructing explanations", and the *crosscutting concept* of "energy" (NRC, 2012). Drawing on data from our design of this task, iterative rounds of feedback with teachers, science curriculum coordinators, scientists, and students, we identify challenges facing curriculum and assessment designers as they move into the space of three-dimensional assessment.

Theoretical and conceptual foundations

Consistent with what Ford & Forman (2006) called the 'practice turn' in sociocultural theory, and the following focus on engagement in practice as a goal for disciplinary learning, the *Framework* for the Next Generation Science Standards (NRC, 2012) foregrounds engagement in practice for students as they learn disciplinary core ideas and apply crosscutting concepts. This shift in focus seeks to move the field of science assessment away from a focus on knowledge alone, as has been the traditionally privileged outcome of educational contexts for decades, and toward a definition of learning as changes in participation in disciplinary practices over time (e.g. Wenger, 1998). This shift in the way we theorize about what students do in assessment contexts also repositions the way we think about knowledge. From a sociocultural perspective, assessment designers are no longer considering knowledge as the only outcome (e.g. Shepard, 2000), but rather focus on the ways that students engage in practices as they demonstrate their disciplinary knowledge.

The consequences for the design of assessments are clear: assessments can no longer focus only on eliciting the different types of knowledge that students bring to different contexts, but must also create opportunities for students to engage in scientific practices (NRC, 2014). While this perspective is not new, its primary emphasis on multiple dimensions of science learning is. Many performance assessments developed in the 1990's prioritized students' engagement with concrete materials as they solved contextualized problems (Solano-Flores & Shavelson, 1997) and the ways in which students engaged in processes of inquiry as an outcome of their science learning. However, the extent to which these tasks actually engaged students in higher-level cognitive processes has been questioned (Baxter & Glaser, 1998). For example, the 'Paper Towels' task assessed students' experimental design of which brand of paper towels absorbed the most water (Baxter and Shavelson, 1994), without a focus on underlying scientific principles.

In the current reform context, science assessment design frameworks draw deeply upon these previous efforts (e.g. Pellegrino, Chudowsky & Glaser, 2001), but are also informed by new perspectives on student engagement in scientific practices, such as modeling (e.g. Schwarz et al., 2009; Windschitl, Thompson & Braaten, 2008), explanation (e.g. McNeill et al., 2006), and argumentation (e.g. Bricker & Bell, 2008). In addition, three-dimensional science assessments seek to create opportunities for students to demonstrate their learning in the context of compelling, real-world phenomena (NRC, 2014), as is the case with some of the sample tasks that have been generated recently (Achieve, Inc., 2014). In this sense, the design of assessment tasks is repositioned from the ways we have previously thought about transfer of learning (NRC, 2007) to instead focus on contextualized phenomena aligned with students' interest (NRC, 2014).

Multicomponent assessment design

The new vision for assessment tasks based on the *Framework* challenges assessment designers to use multiple inter-related questions or components to fully assess the performance expectations included in the NGSS (NRC, 2014). These tasks will be developed following evidence-centered or construct-centered design processes (NRC, 2014), and will involve an iterative process composed of multiple steps: analyzing and detailing the cognitive domain to be assessed (others have called this 'unpacking,' see Stevens, Delgado & Krajcik, 2010), identifying the inferences that the assessment is designed to support about student learning and determining the types of evidence necessary to support those inferences (Pellegrino et al., 2001), designing tasks that will collect that evidence, and determining how to model evidence to support valid conclusions (NRC, 2014).

Given that the *Framework* (NRC, 2012) vision is still new, and the design processes for *Framework*-aligned tasks even newer (NRC, 2014), examples of what these assessment tasks look like in practice are only beginning to emerge. Achieve, Inc. has released sets of sample tasks (Achieve, Inc., 2014), all of which may take days, even weeks to complete. Pages of single-spaced task prompts are followed by multiple diagrams, data, and images for students to analyze, leading to long tasks that would require significant tailoring for use in teachers' school contexts, as well as scaffolding to support students in responding to the tasks. Clearly, the field is still developing images of what form this type of assessment will take.

Criteria and constraints for the design of a three-dimensional task

Our efforts in this domain have taken place in the context of a larger research-practice partnership (Penuel et al., 2011) intended to develop a system of three-dimensional classroom assessments. This partnership, which dates to 2014, began at the initiation of the school district, which reached out to researchers at our University for support around NGSS-aligned formative assessment design. This mutualistic collaboration (Coburn & Penuel, 2013) involves long-term commitments from researchers with deep support from district administration in our partner district, located outside a large city in the Western US.

Since that time, three externally funded grants have supported our partnership as we have developed a series of multicomponent, pre-post assessments to model student learning within and across school years, creating opportunities for longitudinal tracking of cohorts of students as they move through high school physics, chemistry, and biology. While our initial assessment design efforts spanned multiple disciplinary core ideas and crosscutting concepts, we have simplified our work by focusing on the scientific practice of modeling (Passmore & Svoboda, 2012; Passmore, Schwarz & Mankowski, 2017), and then tracing energy as both a disciplinary core idea unifying instruction across high school physics, chemistry, and biology, as well as a crosscutting concept across these disciplines (c.f. Park & Liu, 2016). We describe this in greater detail in the following section.

Energy: Crosscutting concept and disciplinary core idea

Energy occupies a unique position in the Next Generation Science Standards, as it is both a core idea across the different science disciplines, as well as a cross-cutting concept. Studies in different scientific disciplines have investigated both implicit and explicit learning of the concept (e.g. Park & Liu, 2016; Opitz et al., 2015; Opitz, Blankenstein & Harms, 2016; Neumann et al,, 2013). To date, four main characteristics of energy have been identified: energy is present in different forms; energy can be transformed from one form to another or transferred without changing its form; energy is degraded, whenever it is transformed; and the overall quantity of energy is conserved (Duit, 1984). These characteristics are more or less prominent in disciplinary topics. For example, in life sciences, teaching of the energy concept mostly focuses on energy transfer and transformation processes in open systems (Opitz et al., 2016), whereas in physics all four characteristics are introduced in middle school and are revised with quantitative considerations in high school (Neumann et al., 2013).

Learning about the four characteristics of energy is the subject of many studies (e.g. Neumann et al., 2013, Jin & Anderson, 2012; Nordine, Krajcik, & Fortus, 2011). For example, Jin and Anderson (2012) identified a hierarchical structure of energy understanding for biology students, and Opitz and colleagues (2016) described the energy conceptions of biology students in middle school. Although students' understanding of energy forms and transfer/transformation have been shown to increase over time, students also maintain many prior ideas they held before entering school after instruction (Jin & Anderson, 2012; Lancor, 2014).

Students' conceptions of energy in physics or chemistry develop in a similar way and student learning also seems to be hierarchical. In physics, learning about transfer and transformation is associated with degradation (Neumann et al, 2013) and in chemistry energy transfer and transformation is associated with forms (Teichert & Stacey, 2002). The few studies that have assessed energy across all disciplines (e.g., Opitz, et al., 2017; Park & Liu, 2016) have found high latent intercorrelations between the energy understanding in different disciplines. These findings indicate, that there is little variance in student learning between disciplines while maintaining a large variance within each discipline (Park & Liu, 2016).

Research Questions

The studies reviewed above set a key challenge for three-dimensional science assessment: to determine the types of tasks that will be able to capture the development of student engagement in scientific practices and disciplinary core ideas as they span across crosscutting concepts that students encounter in multiple years of study, such as energy. Specifically, we have proceeded with the process of designing an assessment task in the context of our research-practice partnership, and seek in this paper to respond to the following research questions: *How can we develop a three-dimensional task to assess the crosscutting concept of Energy? What tensions and challenges emerge in this process?* In responding to these questions, we seek not only to identify tensions and the ways we addressed them in our study, but also to inform future assessment design in this area.

Method

Our paper, and the larger project in which it is embedded, uses a Design-Based Implementation Research (DBIR) approach in the context of a research-practice partnership (Coburn & Penuel, 2016) to develop and test innovations fostering alignment and coordination to improve classroom practices (Penuel et al., 2011). We conduct rapid cycles of design that allow us to negotiate means and goals across multiple stakeholders in real time (Cobb et. al, 2013); this paper provides a case analysis of several cycles of rapid prototyping a single task.

Task design procedure

Our task design followed a multi-step, iterative approach (NRC, 2014). We started by identifying the NGSS performance expectations associated with energy as disciplinary core ideas and crosscutting concepts. Next, we 'unpacked' the ways energy was discussed in these performance expectations by building on and expanding frameworks for energy in physics, following Neumann et al. (2014)'s learning progression for energy forms, transfer/transformation, conservation, and degradation/dissipation. Next, working from similar assessments tracing energy across multiple frames of reference (e.g. Ambitious Science Teaching, 2017; Neumann, Fortus & Nordine, 2017), we developed separate versions of the task for every science discipline. This first task draft was piloted with high school environmental science students in our partner district (N = 26).

Based on students' response patterns, as well as the desire to move closer toward a task that could be used at any grade level or disciplinary focus in high schools, we used the students' pilot data to guide our revision of the three separate tasks and combine them into a single crosscutting format. This second version of the task was piloted with disciplinary experts, scientists in the domains of physics, chemistry and biology (N=5). All of the scientists were asked to solve the task and two of the scientists were interviewed about their responses. Additional feedback about the clarity and accuracy of the task was used to inform the next iteration. At this phase we also developed pilot versions of a rubric to score the task, using scientist expert responses to populate the top levels of the rubric, and information from the SOLO Taxonomy (Biggs, 1979), Park and Liu's (2016) study of energy as a crosscutting concept, and Neumann et al.'s (2013) five 'big ideas' about energy. At this same phase, we shared the task with communities of science teachers in the partnership (4 teacher groups, 16 teachers) and collected information about their responses and reactions in detailed fieldnotes. Teachers discussed the task in the context of their own understanding of the content as well as in the context of their perception of their students' understandings.

The third iteration of the task was piloted with university undergraduates in freshman-level chemistry courses, as they are representative of students who have completed three years of high school science (N=23). Following this administration, the research team conducted focus sessions (Briggs & Peck, 2015) with representative samples of student responses to identify further revisions to the task, as well as to the scoring rubric. Based on these experiences, we developed a revised, simplified format for the task.

Sources of data and analytic approach

We draw on several sources of data collected across multiple settings during the spring and fall of 2017, including in-depth running design notes from weekly university-based research team meetings, handwritten fieldnotes made during bi-monthly design meetings with district partners, meeting agendas, artifacts, and fieldnotes from school-based teacher learning community meetings, facilitation guides, fieldnotes, and artifacts created at all-district professional development meetings, student responses to pilot versions of the task, and fieldnotes and artifacts from administering the task to scientists.

The authors of this paper met multiple times to review and discuss the data, and identified initial tensions. These tensions were shared with members of the research team, who then interrogated our developing ideas. We then tested the initial tensions we identified against other forms of data, refining them as we developed the case study. Our identification of tensions emerging from across these multiple settings and sources of data occurred during these conversations, as well as in the course of our regular research team activities. After we wrote up our initial emergent tensions and claims to support them within the research team, we shared those claims with district science coordinators, scientists, other members of our research team, as well as other learning scientist colleagues. We integrated their feedback and reflections in the final draft of this paper.

Emergent tensions in process of design

Our analysis of the preceding sources of data have led us to identify three emergent tensions surfaced in our attempts to design a three-dimensional assessment task intended for use in tracing the development of students' understanding of how to model energy in systems across multiple school years. We describe each of these tensions below, with illustrative examples of our iterative cycles of design, piloting, and revision.

Tension between practice-as-embodied in the task and current classroom practice

The original idea for the energy task came from Sabrina and Liz, the two district science coordinators in our partner district, while examining published examples of assessment in Nordine (2017). Noting that using contexts related to sustainability was a priority in their district, Sabrina seized upon the idea of extending the example of biofuels into a context that might be used across high school physics, chemistry, and biology. We began developing a modeling and explanation task that would allow students to trace energy as a crosscutting concept across systems that represented different disciplinary core ideas in the three science domains.

As we mocked up versions of the task, Sabrina and Liz reflected on whether or not it was the kind of assessment that would align with current science teaching practice in the district, which largely involve traditional instruction representative of the majority of US science classrooms (e.g. Banilower et al., 2012). District leadership had recently asked her to justify her department's focus on the three-dimensional vision of learning, even though she was working in a non-NGSS state. Noting that the classroom practice of most teachers in the district was nowhere near the three-dimensional approach we were aiming for, Sabrina exclaimed, "I feel like I'm on Pluto." When asked to explain more about what she meant, she elaborated that she felt like the vision we were going for in our partnership was so distant from what was happening in classrooms in the district that the design for the assessment task was starting to feel like outer space. Nevertheless, both Sabrina and Liz committed to the task as 'aspirational' to inform vision for science teaching and learning in the district, and they intend to begin using this task as a way of compelling changes in classroom practice. However, in the meantime, this also means that piloting and initial use of the assessment is taking place in classrooms in which students have little experience with or opportunity to learn through modeling and explanation; that is, practice-as-embodied in the task feels, at times, billions of miles away from what students experience on a daily basis.

Tensions between modeling, explanation, and scorable student responses

Long-standing lines of research in science education have examined the ways in which students' abilities to create and use models support their learning of important science concepts (e.g. Schwarz et al., 2009), and studies of the ways that students create and revise models have identified multiple types of scaffolds, such as checklists, that help students engage in this scientific practice (Kang et al., 2014). In parallel, researchers have also established frameworks and scaffolds for engaging students in the scientific practice of constructing explanations (e.g. McNeill et al., 2006; Songer & Gotwals, 2012). In our work, we have sought to create a task prompt that builds on lines of research from both of these traditions to move toward engaging students in using their models to creare explanations, in which students create a model for a given phenomenon, and then use that model to develop an explanation of a specific phenomenon.

The left side of Figure 1 shows the first version of the task that had a large, blank space for students to draw a model of how algae produces oil that could be used to fuel a bus (lower-left frame), as well as an unstructured space for students to write an explanation for how energy flows in this system. We provided images at the upper-right and lower-left to scaffold students' responses about energy transfer from the sun to the algae, as well as energy helping the bus move. We created similar versions that alternated the focus at the center, one with a larger frame for biology students (focusing on how algae capture energy from the sun through photosynthesis) and another for physics students (focusing on how energy helps a bus move up a hill). Our initial pilot with high school students quickly indicated that students were unsure what to make of the different reference frames, as well as the large blank box at the center, and wrote little to no explanation. Students were unfamiliar with the phenomenon of using algae to produce biofuels and were unsure how to approach the task.

At the same time, as a research team, we engaged in conversations about difficulties we would encounter in modeling student understanding of a crosscutting concept if the task was so intimidating as to provide no place for students to begin. We were also concerned that students would not explain the entire process in a large outcome space, and reflected that breaking the explanation into smaller pieces linked to the different pieces of the model might help us generate more scorable information. We also had concerns that the three versions of the task might be non-comparable across grades, creating difficulties in tracking students across multiple years of science courses.

Our solution to these design challenges is shown in the right side of Figure 1, a portrait-oriented version of the task that repositions the question in the context of corn, a familiar crop to the students in our partner district, and ethanol, a substance students commonly see or hear about at local gas stations. It breaks the modeling of energy transfer and transformation into four, equally-sized boxes (corn plants capturing energy from the sun; distilling ethanol from fermented glucose; combusting ethanol in a piston; a bus moving on a road). Each box is then matched with a specific question about energy in that part of the model, with separate outcome spaces. This version of the task, then, was intended to strike a balance between having students create a model and use that model to create an explanation, as well as having smaller pieces of the task with more accessible outcome spaces for students.

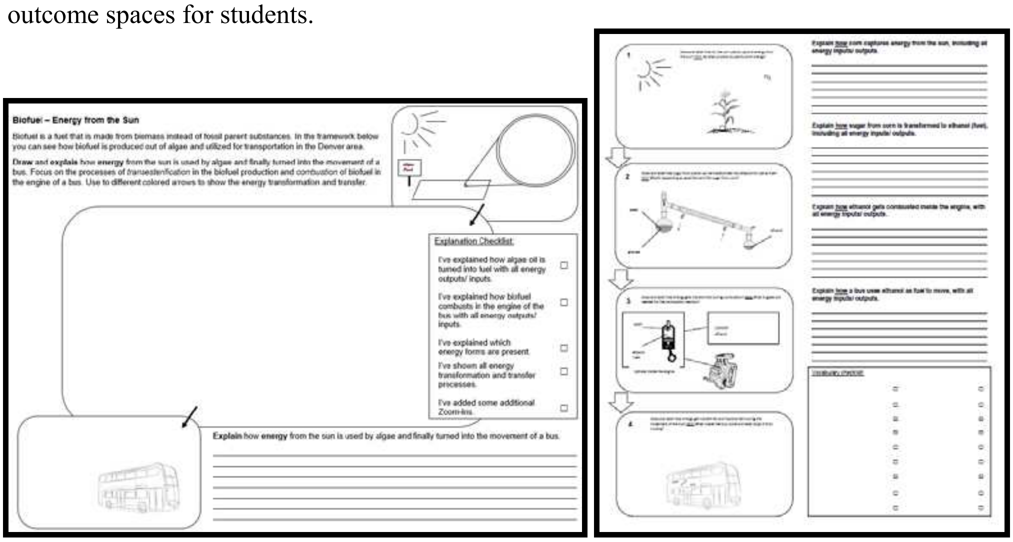

Figure 1. Versions of task including a disciplinary focus on modeling energy transfer in chemistry (left), and a later iteration breaking the explanation into frame-specific sections (right).

In this process of designing the task, as Figure 1 shows, we also experimented with different checklists for vocabulary, modeling and explanation. Following Kang and colleagues' (2014) findings, we knew that providing some level of scaffolding would create more opportunities for students to make their thinking visible in the task, and would increase the quality of their models (e.g. focusing on both visible and invisible processes) and explanations (explaining transfer and transformation in the different parts of the model). We shared this version of the task with our teacher partners, collecting both their responses and their impressions of the task.

When performing initial modeling work to explore the ways students were likely to think about this crosscutting concept, our team developed a task, based on Eisenkraft (2017), in which we built an inefficient calorimeter and then created initial and revised models of how energy would flow through the system when we burned a piece of Pirate Booty. The first time we completed the task, members of the research team with biology, environmental science, and physics backgrounds all included different types of energy in their models. The physics major wrote about kinetic and potential energy, common ways of talking about energy in his discipline, whereas the biologist and environmental scientists focused on energy transfer and transformation within the system, consistent with their experiences modeling energy and matter flow in ecosystems. Our experiences engaging teachers at our partner schools in this calorimeter activity yielded similar results.

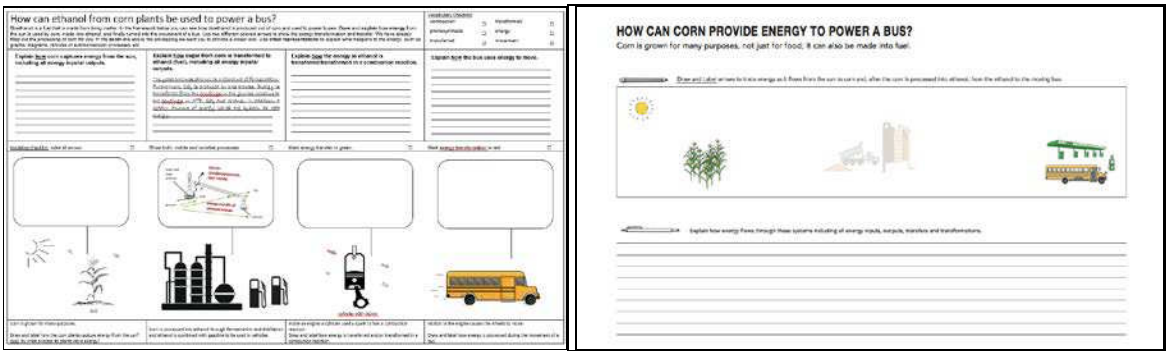

Figure 2. Task with zoom-ins (left) and simplified version with single modeling outcome and explanation (right).

When we gave an updated version of the task shown in Figure 2 (left) to disciplinary experts, we were seeking to prompt them to not only focus on how energy flowed through the system, but also to make macro- and micro-level connections (e.g., not only noting that energy transfer and transformation are occurring in photosynthesis, but also writing out equations for carbon fixing in the process of photosynthesis). We were surprised to find that most of the experts took on the sections associated with their field first and later apologized for their lack of familiarity with other sections. For example, the physicist stated that "... the complexity of the photosynthetic process is outside of my specialty," and the biologist noted on her task, "I found this difficult because I don't teach this topic." Counter to our expectations, these experts - all of whom had extensive knowledge of contexts inside and outside of their fields - were also feeling limited by the same disciplinary boundaries uncovered by members of the research team. This led us to wonder about the ambition of the crosscutting concepts themselves, since they were seeking to represent larger ideas in science that even pushed the boundaries of the ways scientists think on a daily basis.

As we piloted the task with high school and college-level students, we also became increasingly aware that the boxes around the different elements of the model - vestiges of the original three-version task - actually might be reinforcing these disciplinary boundaries that were challenging for the scientists. As such, the later versions of the task removed both the boxes around the different parts of the model with specific disciplinary foci, as well as the suggested 'zoom-out' boxes intended to prompts students to draw micro-level processes.

These experiences prompted us to simplify the task to allow a broader aperture of responses, where disciplinary experts might be able to 'go deep' at the micro- or nano-scale, while also making the task accessible to students responding on the macro-scale on the basis of their everyday experiences. We also hoped that we could find a task format that would allow students to work with the ideas of the crosscutting concept of energy without necessarily being turned off by checklists of vocabulary words with which they might not be familiar. Thus the revised version of the task, shown on the right side of Figure 2, included fewer scaffolds for both the model and the explanation outcome space. At the time of publication of these proceedings, this version was being administered to physics, chemistry, and biology students in our partner district.

Discussion

As the field moves toward new ways of thinking about science learning, new methods for developing classroom-based assessments of this learning are necessitated (NRC, 2014). Our experiences developing the Energy Assessment Task has illustrated that a design in 'outer space' may seem that way not only to the students and teachers in our partner district, but also to scientists whose daily work takes place within disciplinary constraints that reinforce the very boundaries that new ways of thinking about science learning - in particular, crosscutting concepts - are intended to diminish. The tensions we have identified are likely only the beginning of those that assessment designers and those working in partnership with districts, schools and teachers are likely to uncover. However, we emphasize that such aspirational assessments are an important component in the new systems of curriculum materials and professional learning experiences currently being developed to support *Framework*-aligned learning experiences for students (e.g. Reiser et al., 2017).

Returning to the situated perspective that we bring to this work (Greeno, 2006), we acknowledge the critical role that tools such as tasks like these might play not only in reorganizing participation structures in the classrooms we support, but also to create opportunities to discuss the ways in which the task embodies a shared vision for the district as it moves toward different ways of thinking about science learning outcomes (Wenger, 1998). We also acknowledge the number of critical questions that we direct to those developing assessments in this domain, including: What does it mean to move toward a vision of assessment that is so 'out there' that even scientists are challenged by thinking in that way? How can we scaffold student participation in assessments like these when students' opportunities to learn through instruction aligned with the task are still limited? What does this mean for the vision for learning and teaching in the NGSS?

References

Achieve, Inc. (2014). *NGSS Classroom Sample Tasks.* Downloaded November 17, 2017 from https://www.nextgenscience.org/classroom-sample-assessment-tasks.

Banilower, E. R., Smith, P. S., Weiss, I. R., Malzahn, K. A., Campbell, K. M., & Weis, A. M. (2013). *Report of the 2012 National Survey of Science and Mathematics Education*. Chapel Hill, N.C.

Baxter, G. P. and Glaser, R. (1998), Investigating the cognitive complexity of science assessments. *Educational Measurement: Issues and Practice*, *17*, 37–45.

Baxter, G. P.& Shavelson, R.J. (1994) Science performance assessments: Benchmarks and surrogates. *International Journal of Educational Research, 21,* 279-298

Biggs, J. (1979). Individual difference in study processes and the quality of learning outcomes. *Higher Education*, 8(4), 381–394.

Bricker, L. A., & Bell, P. (2008). Conceptualizations of argumentation from science studies and the learning sciences and their implications for the practices of science education. *Science Education*, *92*(3), 473–498.

Briggs, D. C. & Peck, F. A. (2015). Using learning progressions to design vertical scales that support coherent inferences about student growth. *Measurement: Interdisciplinary Research & Perspectives*, 13, 75-99.

Cobb, P., Jackson, K., Smith, T., Sorum, M., & Henrick, E. (2013). Design research with educational systems: Investigating and supporting improvements in the quality of mathematics teaching and learning at scale. *The National Society for the Study of Education Yearbook*, *112*(2), 320–349.

Coburn, C. E., & Penuel, W. R. (2016). Research-practice partnerships in education: Outcomes, dynamics, and open questions. *Educational Researcher*, *45*(1).

Duit, R. (1984). Learning the energy concept in school—empirical results from the Philippines and West Germany. *Physics Education*, *19*(2), 59–66.

Eisenkraft, A., (2016). Teaching about energy as a crosscutting concept. In J. Nordine, Ed., *Teaching About Energy Across the Sciences,* pp. 39-60. Arlington, VA: National Science Teachers Association Press.

Ford, M., & Forman, E. (2006). Redefining disciplinary learning in classroom contexts. *Review of Research in Education, 30*, 1-32.

Jin, H., & Anderson, C. W. (2012). Developing assessments for a learning progression on carbon-transforming processes in socio-ecological systems. In A. C. Alonzo & A. W. Gotwals (Eds.), *Learning progressions in science: Current challenges and future directions* (pp. 151–182). Rotterdam: Sense.

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.

Lancor, R.A. (2014). Using student-generated analogies to investigate conceptions of energy: A multidisciplinary study. *International Journal of Science Education, 36*(1), 1–23.

McNeill, K. L., Lizotte, D. J., Krajcik, J., & Marx, R. W. (2006). Supporting students' construction of scientific explanations by fading scaffolds in instructional materials. *The Journal of the Learning Sciences*, 15(2), 153–191.

National Research Council. (2007). *Taking science to school: Learning and teaching science in grades K-8.* Washington, D.C.: National Academies Press.

National Research Council. (2012). *A framework for K-12 science education: Practices,*

crosscutting concepts, and core ideas. Washington, D.C.: National Academies Press.
National Research Council. (2014). *Developing Assessments for the Next Generation Science Standards*. Washington D.C.: National Academies Press.
Neumann, K., Viering, T., Boone, W. J., & Fischer, H. E. (2013). Towards a learning progression of energy. *Journal of Research in Science Teaching*, *50*(2), 162–188.
Neumann, K., Fortus, D. & Nordine, J. (2016). Assessing energy as a crosscutting concept. In J. Nordine, Ed., *Teaching About Energy Across the Sciences,* pp. 139-170. Arlington, VA: National Science Teachers Association Press.
Nordine, J., J. Krajcik, and D. Fortus. 2011. Transforming energy instruction in middle school to support integrated understanding and future learning. *Science Education, 95*(4): 670–699.
OECD. (2017). *PISA Assessment and Analytical Framework: Science, Reading, Mathematics, Financial Literacy and Problem Solving.* Paris: OECD Publishing.
Opitz, S. T., Blankenstein, A., & Harms, U. (2017). Student conceptions about energy in biological contexts. *Journal of Biological Education*, *51*(4), 427–440.
Opitz, S., Harms, U., & Neumann, K., Kowalzik, K., & Frank, A. (2015). Students' Energy Concepts at the Transition between Primary and Secondary School. *Research in Science Education*, 45(5). 691-715.
Park, M., & Liu, X. (2016). Assessing Understanding of the Energy Concept in Different Science Disciplines, *100*(3), 483–516.
Passmore, C., Schwarz, C. V., & Mankowski, J. (2017). Developing and using models. In C. V. Schwarz, C. Passmore, & B. Reiser (Eds.), *Helping students make sense of the world using next generation science and engineering practices* (pp. 109–135). Arlington, VA: National Science Teachers' Association Press.
Passmore, C. M., & Svoboda, J. (2012). Exploring opportunities for argumentation in modelling classrooms. *International Journal of Science Education*, *34*(10), 1535–1554.
Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). *Knowing What Students Know: The Science and Design of Educational Assessment.* Washington D.C.: National Academies Press.
Penuel, W. R., Fishman, B. J., Cheng, B. H., & Sabelli, N. (2011). Organizing research and development at the intersection of learning, implementation, and design. *Educational Researcher, 40*(7), 331-337.
Reiser, B. J., Michaels, S., Moon, J., Bell, T., Dyer, E., Edwards, K. D., McGill, T.A.W., Novak, M. & Park, A. (2017). Scaling Up Three-Dimensional Science Learning Through Teacher-Led Study Groups Across a State. *Journal of Teacher Education,* 68(3), 280-298.
Schwarz, C. V., Reiser, B. J., Davis, E. A., Kenyon, L., Achér, A., Fortus, D., Shwartz, Y., Hug, B. and Krajcik, J. (2009), Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners. *Journal of Research in Science Teaching*, *46*(6), 632–654.
Schwarz, C. V., Reiser, B. J., Davis, E. a., Kenyon, L., Achér, A., Fortus, D., Krajcik, J. (2009). Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners. *Journal of Research in Science Teaching*, *46*(6), 632–654.
Schwarz, C.V., & White, B.Y. (2005). *Metamodeling knowledge: Developing students' understanding of scientific modeling.* Cognition and Instruction, 23(2), 165–205.
Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher, 29*(7), 4–14.
Solano-Flores, G., & Shavelson, R. J. (1997). Development of performance assessments in science: Conceptual, practical, and logistical issues. *Educational Measurement: Issues and Practice, 16*(3), 16–25.
Songer, N. B., & Gotwals, A. W. (2012). Guiding explanation construction by children at the entry points of learning progressions. *Journal of Research in Science Teaching*, *49*(2), 141–165.
Stevens, S. Y., Delgado, C., & Krajcik, J. S. (2010). Developing a hypothetical multi-dimensional learning progression for the nature of matter. *Journal of Research in Science Teaching*, *47*(6), 687–715.
Teichert, M. A., & Stacy, A. M. (2002). Promoting understanding of chemical bonding and spontaneity through student explanation and integration of idea. *Journal of Research in Science Teaching*, 39(6), 464–496.
Wenger, E. (1998) *Communities of practice: learning, meaning, and identity. Cambridge, U.K.; New York, N.Y.: Cambridge University Press,*
Windschitl, M., Thompson, J., & Braaten, M. (2008). Beyond the scientific method: Model-based inquiry as a new paradigm of preference for school science investigations. *Science Education*, 92(5), 941–967.

Acknowledgement
This material is based upon work supported by the National Science Foundation under Grant No. 1561751. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.