Using Learning Performances to Design Three-Dimensional Assessments of Science Proficiency

Paper presented at the NARST Annual International Conference
Baltimore, MD: April 14-17, 2016

Kevin W. McElhaney, Center for Technology in Learning, SRI International
Brian D. Gane, Learning Sciences Research Institute, University of Illinois at Chicago
Christopher J. Harris, Center for Technology in Learning, SRI International
James W. Pellegrino, Learning Sciences Research Institute, University of Illinois at Chicago
Louis V. DiBello, Learning Sciences Research Institute, University of Illinois at Chicago
Joseph S. Krajcik, CREATE for STEM Institute, Michigan State University

Abstract: We describe how principles of evidence-centered design inform the development of classroom-based science assessments that integrate three dimensions of science proficiency—disciplinary core ideas, science and engineering practices, and crosscutting concepts. Beginning with three-dimensional performance expectations (PEs) in Next Generation Science Standards, our design process entails unpacking the assessable aspects of the three dimensions of the PE. We then create an integrated dimension map that describes the relationships among these aspects of the PEs in a way that informs the design of assessments. Based on the map, we specify a set of claims called *learning performances* that collectively represent the proficiencies of the target PEs. Next, for each learning performance we specify an assessment design pattern that describes the target proficiencies, student evidence for those proficiencies, and task features that elicit the desired evidence. We use the design patterns to design tasks that are aligned to learning performances. This paper describes the design approach and includes accompanying examples of assessment design artifacts. We also consider assessment design challenges, next steps, and implications of this work for next generation science assessment.

Suggested citation:

McElhaney, K. W., Gane, B. D., Harris, C. J., Pellegrino, J. W., DiBello, L. V., & Krajcik, J. S. (2016, April). *Using Learning Performances to Design Three-Dimensional Assessments of Science Proficiency.* Paper presented at the NARST Annual International Conference, Baltimore, MD.

Introduction

A significant challenge facing science educators who are shifting instruction to align with the *Next Generation Science Standards* (NGSS) (NGSS Lead States, 2013) is how to assess students' progress toward achieving the new standards. The NGSS reflects an ambitious vision for science education presented in the National Research Council's (NRC) *Framework for K-12 Science Education* (NRC, 2012). This vision emphasizes the integration of disciplinary core ideas, crosscutting concepts, and science and engineering practices as essential for improving students' understanding and providing foundational knowledge for participating in science as a career professional or citizen. The *Framework* emphasizes rich science learning as requiring a tight coupling of what students know and what they can do. This paper addresses the new challenge of developing classroom-based science assessments aligned with performance standards that integrate the three performance dimensions–disciplinary core ideas, science and engineering practices, and crosscutting concepts. To address the challenge, we describe and illustrate a principled design process for classroom-based assessments that can help teachers guide their students' progress toward complex, end-of-grade band standards.

The shift to integrating science practices with content knowledge is based on studies of professional scientific practice (e.g., Latour & Woolgar, 1996) and on empirical research conducted in the years since the publication of the *National Science Education Standards* (NRC, 1996) and Benchmarks (American Association for the Advancement of Science [AAAS], 1993). Selected practices, such as argumentation and modeling, received little attention in these earlier standards documents but are now more prominent (Duschl & Osborne, 2002; Lehrer & Schauble, 2006). Much of this contemporary research is found in such recent synthesis reports as *How People Learn* (NRC, 2000), *Taking Science to School* (NRC, 2007), and *Learning Science in Informal Environments* (NRC, 2009). These reports, as well as the *Framework*, convey the perspective that proficiency in science requires using and applying knowledge in the context of science practice. When students have opportunities to use science practices to develop and apply their ideas, they deepen their conceptual understanding of content as well as their understanding of how to do science. This knowledge-in-use perspective (Duschl, 2008; Harris & Salinas, 2009; Lehrer & Schauble, 2006, Pellegrino & Hilton, 2012) as instantiated in the NGSS holds that disciplinary core ideas, science and engineering practices, and crosscutting concepts together enable learners to make sense of phenomena and design solutions to problems. Consequently, each of the standards integrates these three dimensions into a single knowledge-in-use statement called a *performance expectation* (PE).

High-quality assessments for student learning that are aligned with standards are critical to developing a coherent and consistent approach to K-12 science education (DeBoer, Lee, & Husic, 2008; Kali, Koppal, Linn, & Roseman, 2008; NRC, 2012; Wiliam, 2010). What is tested can constrain what teachers decide to teach and what they hold their students accountable for

(Marx & Harris, 2006). Even if standards are closely aligned with a desired vision for teaching and learning, when assessments are not aligned with the vision, the achievement of a coherent system can be undermined (Fuhrman, Resnick, & Shepard, 2009; Pellegrino, Chudowsky & Glaser, 2001; Pellegrino & Hilton, 2012). Conversely, well-designed assessments can present a positive vision and target for guiding systemic improvements in education (Frederiksen & Collins, 1989). Attempts to build aligned assessments of learning related to selected disciplinary core ideas, crosscutting concepts, and practices can inform assessment design efforts for states and other stakeholders, by identifying likely decisions that teams will need to make in design and potential challenges they will need to address in developing assessments (see e.g., Pellegrino, Wilson, Koenig, & Beatty, 2014). Currently, existing curriculum-embedded classroom assessments or large-scale assessments do not explicitly elicit students' understanding of integrated content (i.e., disciplinary core ideas and crosscutting concepts) and practices (Pellegrino, 2013; Pellegrino et al., 2014). Teachers in a growing number of states need NGSS-aligned classroom assessments to help them understand how their students are progressing toward achieving the new PEs.

The complexity of the NGSS PEs present formidable challenges for assessment design. In particular, the three dimensional view of science proficiency calls for assessment tasks that integrate the three dimensions (Pellegrino et al., 2014). Therefore, traditional approaches to assessment design that only target disciplinary content will be inadequate for three-dimensional assessment. In addition, PEs represent end-of-grade-band performance targets and therefore often incorporate a wide range of proficiencies that may be difficult (if not impossible) to assess in a single task. Designing assessment tasks to be used formatively during the course of instruction will require an approach that can decompose PEs in a systematic way while retaining their three dimensional nature. These challenges suggest the need for a principled design process that allows assessment developers to align classroom-based assessment tasks to PEs.

Assessment Design Framework and Learning Performances

Evidence-centered design (Mislevy & Haertel, 2006) provides a conceptual framing for analyzing content for assessment design, and it can be used to specify the essential and assessable components of NGSS PEs. ECD-based design patterns explicate an argument about what inferences about student attainment of performance targets can be made based on evidence of student proficiency. Design patterns structure the linkages among the targeted student proficiencies, assessment task design features, and observable evidence. Design patterns describe the kinds of assessment tasks that elicit target constructs and demonstrate how particular performances provide evidence for students' knowledge and abilities (Mislevy & Haertel, 2006; Songer, Kelcey, & Gotwals, 2009). The up-front specification of the design framework via ECD promotes systematicity in assessment task design.

To address the goal of formative assessment during the course of instruction, ECD can be used to systematically decompose PEs into multiple *learning performances* that can guide formative assessment opportunities (DeBarger et al., 2014; Harris et al., 2006; Krajcik, McNeill, & Reiser, 2008). Learning performances are knowledge-in-use statements that incorporate aspects of disciplinary core ideas, science and engineering practices, and crosscutting concepts that students need to attain as they progress toward achieving a PE. Learning performances are akin to learning goals that take on the three-dimensional structure of the PEs—they articulate and integrate assessable aspects of performance that build toward the more comprehensive PE. For classroom purposes, the learning performances also help identify important formative assessment opportunities for teachers. Our design process enables us to derive a set of learning performances from a PE in a principled way (described below) that ensures the learning performances meet these requirements.

Design process

This design process can address either a single PE or a coherent bundle of PEs that is the target of the desired assessments.[1] In this paper, we will use examples from our design of tasks for the PE **MS-PS3-4: Develop a model that predicts and describes changes in particle motion, temperature, and state of a pure substance when thermal energy is added or removed.**

The design process involves the application of ECD, systematic reviews of literature, the description of the performance dimensions in the *Framework*, and the target NGSS PEs to conduct the domain analysis and document the assessment task design. Figure 1 illustrates the overall design and development scheme. It involves both an explicit domain analysis phase where we "unpack" the dimensions of the performance expectations as well as an assessment argument development phase (which we refer to as domain modeling) that provides the blueprint for creating assessment tasks that can be used *during* instruction to provide teachers and students with information regarding progress towards meeting the performance expectations. We have iteratively refined this process in our current work on developing NGSS-aligned middle school physical science assessments. We illustrate each of the design process steps with an example design artifact from our work in the domain of middle school physical science.

Domain Analysis

In ECD, domain analysis entails gathering substantive information about how knowledge and skills are acquired and used in the domain for the purpose of designing assessments. The domain analysis guides the articulation of learning performances. This domain analysis involves (1) unpacking the disciplinary core ideas, science and engineering practices, and crosscutting concepts that are related to the target performance expectations and (2) constructing an integrated dimension map that describe the essential disciplinary relationships and link them to

---

[1] The NGSS recommend addressing bundles of PEs with instruction to achieve instructional coherence and help students relate the standards to each other.

aspects of the targeted crosscutting concepts and science and engineering practices. These steps are elaborated below and illustrated using examples shown in Appendices A, B, and C and Figure 2.

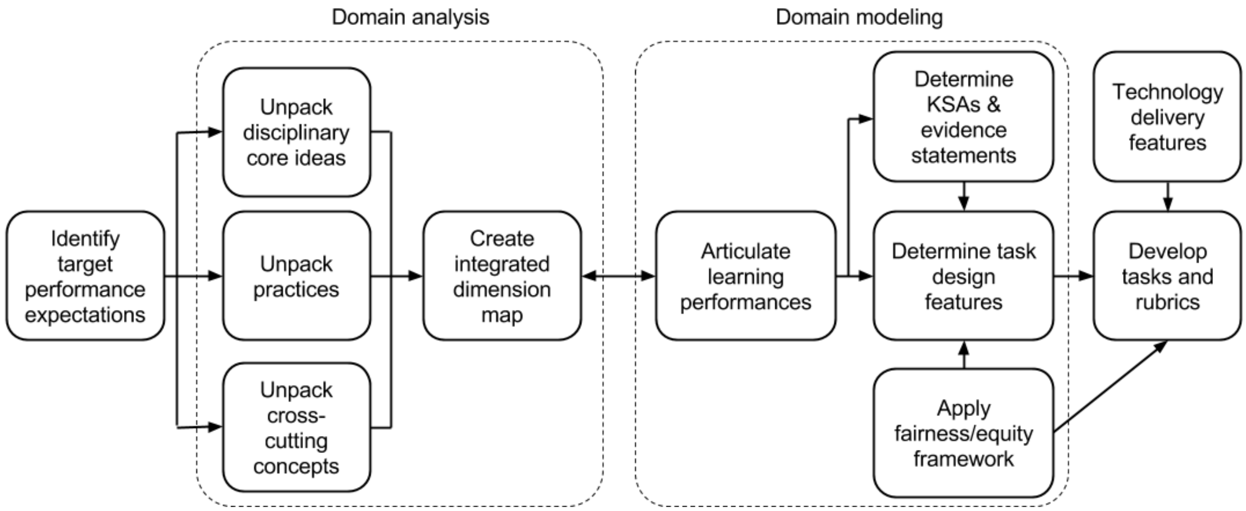


Figure 1. Assessment task design process schematic.

Unpack the science and engineering disciplinary core ideas by elaborating on and documenting the meaning of key terms, determining assessment boundaries for content knowledge, and identifying the background knowledge expected of students to develop grade-level-appropriate understanding of a disciplinary core idea. This elaboration extends what is in the *Framework* and NGSS by identifying research-based problematic student ideas. An example of unpacking of disciplinary core ideas around energy appears in Appendix A.

Unpack the science and engineering practices, which involves defining their core aspects, identifying intersections with other practices, articulating the KSAs associated with the practices, and articulating the evidence required to demonstrate those KSAs (e.g. what constitutes a high level of performance). An example of unpacking of the science practice of Developing and Using Models around energy appears in Appendix B.

Unpack the crosscutting concepts, which involves defining their core aspects, identifying intersections with practices and other crosscutting concepts, articulating the KSAs associated with the crosscutting concepts, and articulating the evidence required to demonstrate those KSAs (e.g. what constitutes a high level of performance). An example of unpacking of the crosscutting concept Cause and Effect appears in Appendix C.

Develop integrated dimension maps that use the elaborations in the unpacking process to lay out the conceptual terrain for achieving each science-engineering PE bundle. These maps describe the essential disciplinary relationships and link them to aspects of the targeted crosscutting concepts and science practices. The maps also illustrate how teachers can support students over time to meet the targeted PEs. These maps are essential to the principled articulation of three-dimensional learning performances that coherently represent the target PEs.

To develop the integrated dimension maps, we begin by creating a traditional concept map (e.g., Schwendimann, 2015) of the essential aspects of disciplinary core ideas included in the target PEs. We begin with the disciplinary core ideas because typically this is the dimension that incurs most of the PEs breadth and complexity. We identify the key concepts and express their disciplinary relationships using arrows linking the concepts. Next, we add to the map the aspects of the practices and crosscutting concepts that are best aligned with the disciplinary relationships for the purpose of creating learning performances and assessment tasks. Usually, the practices and crosscutting concepts that are part of the target PEs are included, but we also include aspects of any additional, related practices and crosscutting concepts that could be used to create learning performances. The resulting dimension map expresses a range of ways that the three NGSS dimensions may be coherently integrated into learning performance statements that collectively represent the target PEs.

Early in our design efforts, we attempted to articulate learning performances based directly on the unpacking documents. We found that for complex domains, we experienced difficulties ensuring the learning performances (1) comprehensively represented and aligned with the PE and (2) integrated dimensions in a principled way. We introduced the dimension maps to facilitate the principled articulation of learning performances. The integrated dimension maps serve to unite the information in the unpacking documents, and they enable the development of a coherent set of learning performances that are connected to the performance expectations. The addition of the dimension maps between the unpacking and writing and refining of a set of learning performances is a significant advancement as it allows task designers to represent the full conceptual space that is possible for integrating the dimensions of highly complex learning goals, such as PE bundles. We found that once the full conceptual space and possible combinations of the three dimensions are expressed in the integrated dimension map, the Learning Performances are more easily articulated and refined. In addition, using the maps to represent the integration of all three NGSS dimensions helps ensure that the Learning Performances derived from the dimension map are also three dimensional, adhering to the vision of NGSS to assess three-dimensional proficiency.

Figure 2 shows an example of an integrated dimension map addressing the PE MS-PS1-4. The structure of the map is based on relationships among disciplinary concepts in the PE, specifically changes in particle motion, particle spacing, temperature, state, and the transfer of thermal energy. The arrows include annotations that (1) link the concepts with text describing the relationship between concepts (e.g., "States of matter *differ in their* particle spacing and changes in particle location relative to other particles") and (2) identify practices and crosscutting concepts (and their specific aspects) that could best elicit evidence of a student's proficiency with the PE. For example, in the lower left area of the map, we deemed the patterns in motion and spacing of particles across different states of matter is aligned not only with using and developing models, but also with evaluating models. We also found this disciplinary relationship to be better aligned with the crosscutting concept of patterns than the Cause and Effect (the crosscutting concept of the PE). In the lower right area of the map, we aligned our data-based characterization of the practice of constructing explanations to the concept that was most likely to yield authentic data (temperature).

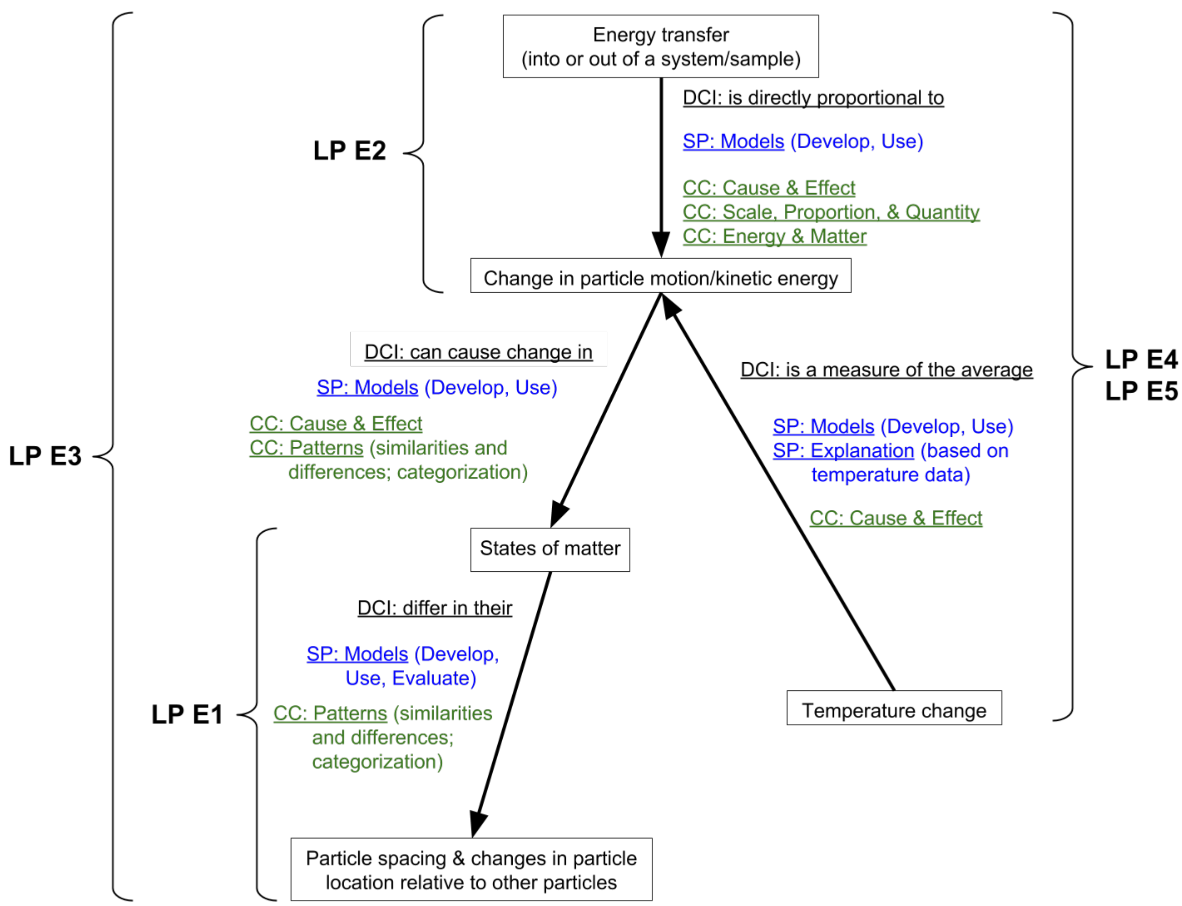


Figure 2. Integrated dimension map of the PE MS-PS1-4. Blue text refers to science practices and green text refers to crosscutting concepts. The three dimensional map clarifies how the PE can be coherently decomposed into a set of learning performances (LP) having various levels of complexity.

Domain modeling

In ECD, domain modeling entails organizing information and relationships from the Domain Analysis for the purpose of linking assessment design features to evidence of students' proficiency in the domain. For the domain modeling we first articulate a set of learning performances based on the integrated dimension maps. These learning performances constitute the claims we wish to make about what students know and can do. We then create design patterns that describe (1) the specific knowledge, skills, and abilities (KSAs) that are necessary to demonstrate proficiency with the learning performance, (2) the features of student responses that could be used as evidence for the claim articulated in the learning performance, and (3) features of assessment tasks that would effectively elicit student responses as defined in (2).

Articulate learning performances. Using the integrated dimension maps, we articulate sets of learning performances that represent target PEs. The articulation of learning performances and the integrated dimension maps is a cyclical, iterative process, with each informing the refinement of the other. Multiple learning performances are designed to build toward a PE in a way that could inform a teacher about a student's progress toward becoming proficient with the target PEs. Learning performances represent a keystone in the evidence-based argument that our assessment tasks represent the NGSS PEs for formative assessment purposes. Learning performances not only represent the content of the PEs, but also address practical, intermediate targets for instruction aligned with the PEs. In this way, learning performances are informed by performance targets, the nature of instruction, and theories of learning in the discipline.

Table 1 lists five learning performances derived from the integrated dimension map shown in Figure 2. This set illustrates how learning performances can increase in complexity, and how they identify specific, intermediate performance targets for instruction that guides students toward the PE. For example, in order to develop a model relating thermal energy, particle motion, and state of matter, students should first be able to demonstrate less sophisticated proficiencies, such as distinguishing among the states based on their particle characteristics (LP E1) and relating thermal energy to particle motion within a particular matter state (LP E2). This particular set of learning performances also separates the concept of temperature from the concept of states, emphasizing the relationship between temperature and particle motion as articulated in the Framework.

We do not claim that these particular learning performances are the only ones that could be derived for PE MS-PS1-4. A different set of learning performance could be articulated linking temperature, particle motion, and states of matter. Specific decisions about learning performances should address the specific needs of teachers, learners, and alignment with curriculum materials, among other considerations.

Table 1: Five Learning Performances derived for NGSS PE MS-PS1-4.

LP E1	Students **evaluate a model** that uses a particle view of matter to explain how states of matter are similar and/or different from each other.
LP E2	Students **develop a model** that explains how particle motion changes when thermal energy is transferred to or from a substance without changing state.
LP E3	Students **develop a model** to explain the change in the state of a substance caused by transferring thermal energy to or from a sample.
LP E4	Students use evidence from a model to **construct a scientific explanation** about how the average kinetic energy and the temperature of a substance change when thermal energy is transferred from or to a sample.
LP E5	Students **develop a model** that includes a particle view of matter **to predict** how the average kinetic energy and the temperature of a substance change when thermal energy is transferred from or to a sample.

Specify design patterns. For each learning performance, we then specify a design pattern that guides the design of assessment tasks that are aligned to each learning performance. Our design patterns have the following components. Table 2 illustrates an example design pattern for learning performance E-02 (which is listed in Table 1).

- **Focal KSAs and evidence statements.** Though we specify learning performances to be assessable in a single task, learning performances are sufficiently complex so as to integrate multiple and distinct aspects of the target science proficiencies. For each learning performance we identify several focal KSAs. The focal KSAs constitute the basis for evidence statements, which describe observable features of student performance that could provide evidence of proficiency. Evidence statements inform the specification of assessment task features and the development of assessment tasks and rubrics.
- **Additional KSAs**. Additional KSAs represent proficiencies that are not targeted by the learning performance, but that may be required for students to demonstrate proficiency with the learning performance, such as prerequisite disciplinary knowledge or familiarity with specific scientific representations. Additional KSAs are important to articulate in design patterns because they increase task designers' awareness of task features that may be construct irrelevant. Additional KSAs also inform when teachers may appropriately use tasks during the course of classroom instruction.
- **Assessment task design features**. Design patterns describe features of assessment tasks that will elicit the focal KSAs. Design patterns distinguish two general types of task design features: (1) *characteristic features* of tasks, which *must* be present to provide the desired evidence of proficiency, and (2) *variable features* of tasks, which may be varied in order to vary task difficulty, focus, or context or to address the needs of students with specific instructional requirements or abilities. Characteristic and variable task features

are important for communicating to task developers what features are needed to design tasks that can provide evidence for the claim articulated in the learning performance.

- **Equity and fairness framework**. Our design patterns include task features derived from an equity and fairness framework that we developed to help ensure that our tasks are fair to students of diverse social and cultural groups. This framework is based on Universal Design for Learning (UDL) (Rose & Meyer, 2002, 2006; Rose, Meyer, & Hitchcock, 2005), which was originally conceived to address the needs of students with disabilities. We have broadened UDL to encompass aspects of culture, gender, class, and other social attributes.

For large-scale, high-stakes assessment, design patterns often incorporate features of the technology-based delivery system intended for the tasks, ensuring alignment between specific technology affordances, task design features, and desired evidence of student proficiency. For classroom-based formative assessment, we intend for our design patterns to be generalizable to a range of delivery methods. The design patterns are therefore useful to a range of assessment designers with unique technology considerations appropriate for diverse classroom settings. For this reason, we consider the technology affordances during the task design phase.

Assessment task development

We use the design patterns to guide the development of assessment tasks aligned with the learning performances. The design patterns for each learning performance guide task development, ensuring that the tasks align with a clear specification of the evidence to be derived from each student response. In our approach, a *task* is not synonymous with a single response prompt and its associated answer, but rather comprises a *scenario* composed of multiple elements. A task or scenario might include various types of student responses including selections, written elements, drawings, or interactions with simulations. We intend for teachers to use a handful of tasks at appropriate points during instruction to gauge their students' progress toward achieving the PEs.

Each assessment task is designed to be completed in 5-10 minutes. This task length balances the desire to engage students in authentic science practices with the need for teachers use the tasks flexibly during instruction and get timely information from the tasks for formative purposes.

We deliver the tasks using the Next Generation Science Assessment (NGSA) task portal hosted by the Concord Consortium and available on our project website (http://nextgenscienceassessment.org). During the development of specific tasks, we consider the alignment of specific technology-based affordances of the NGSA task portal for engaging students with all three dimensions of the learning performance. We also review tasks using our equity/fairness framework and revise the tasks accordingly.

Table 2. Design pattern for learning performance E2.

Learning Performance	**LP E2:** Students **develop a model** that explains how particle motion changes when thermal energy is transferred to or from a substance without changing state
Focal Knowledge, Skills, and Abilities	• Ability to develop a model that explains the change in motion of particles resulting from the transfer of thermal energy to or from a substance • Ability to support the model with a statement describing the causal relationship between thermal energy change and particle motion change • Ability to represent the correspondence between model features and real-world entities (e.g. matter particles) and their attributes (e.g. speed)
Evidence Statements	Student's response includes • A drawing that depicts particles in a substance moving faster with the transfer of thermal energy to the system and/or slower with the transfer of thermal away from the system • A statement describing the causal relationship between thermal energy transfer and the change in particle motion • A legend or labels on the drawing to indicate how model features represent types of particles and particle speed
Additional Knowledge, Skills, and Abilities	• Knowledge that a model is a representation (e.g., a drawing) that explains why or how a phenomenon occurs • Ability to use a computer-based drawing tool
Characteristic Task Features	• Task presents a scenario where the transfer of thermal energy causes an observable phenomenon that can be explained by a change in particle motion without changing the state of a substance • Task prompts students to develop a model to explain what is observed in the scenario • Task prompts students to describe (in text) what the model shows • Task provides students with a computer-based drawing tool to develop a model • Tasks provide a scientifically authentic investigation context that is accessible to students with diverse cultural backgrounds and experiences. • Tasks use straightforward language that is accessible to students with diverse linguistic abilities
Variable Task Features	• The type of scenario/phenomenon students are asked to model • How the phenomenon is represented in the task (e.g. video, verbal description, static image) • State of matter of the substance to or from which thermal energy is transferred • Whether thermal energy is transferred to or from the object, or both • The features of the modeling/drawing tools that the task provides for students • Specific prompt for students to include a legend or labels for their model • Task scaffolding features that can help elicit relevant model features and mechanistic elements • Visual aids to support comprehension by students with diverse linguistic and visual processing abilities

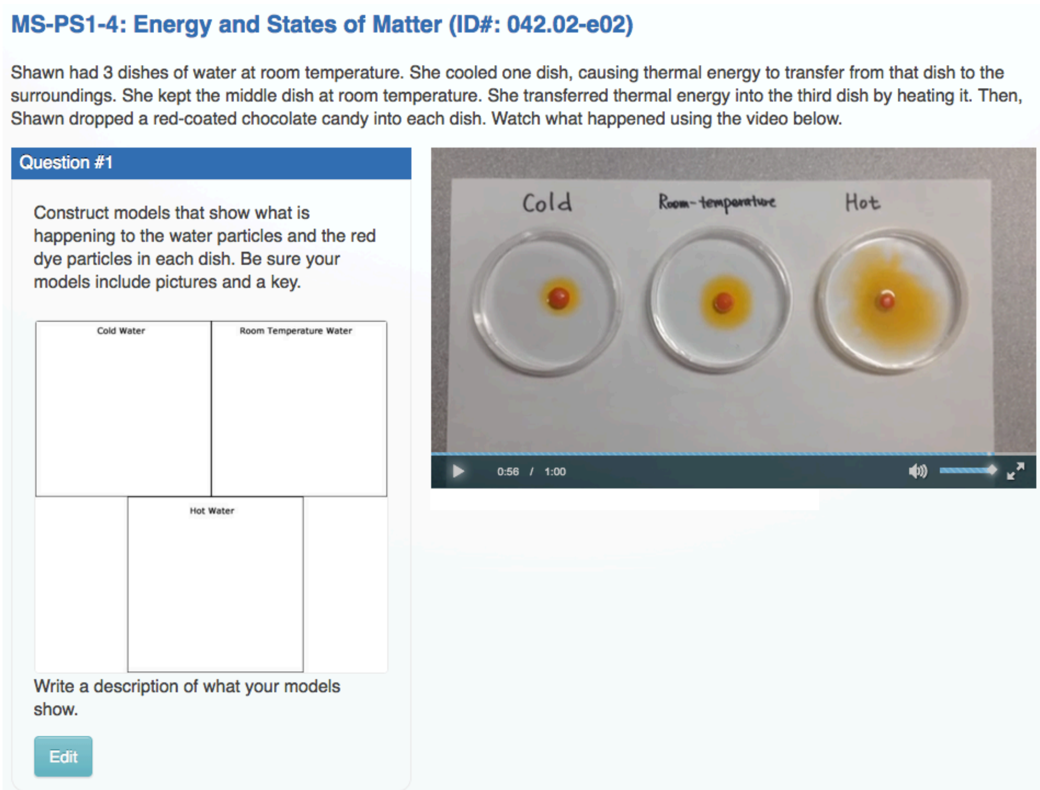
Figure 3 illustrates an example assessment task aligned with learning performance E2. The task aligns with the design pattern in Table 2, affording students the opportunity to produce the evidence of proficiency described in the evidence statement. The task includes all of the characteristic features and select variable features, including a prompt for students to include a legend or labels for their model. This task presents students with a video of dye spreading at different rates through water at different temperatures. The task elicits evidence that students can integrate the relationship between thermal energy transfer and particle speed (disciplinary core idea), developing a model (science practice), and the underlying mechanism for how thermal energy causes the dye to spread at different rates (crosscutting concept).

Challenges and Next Steps

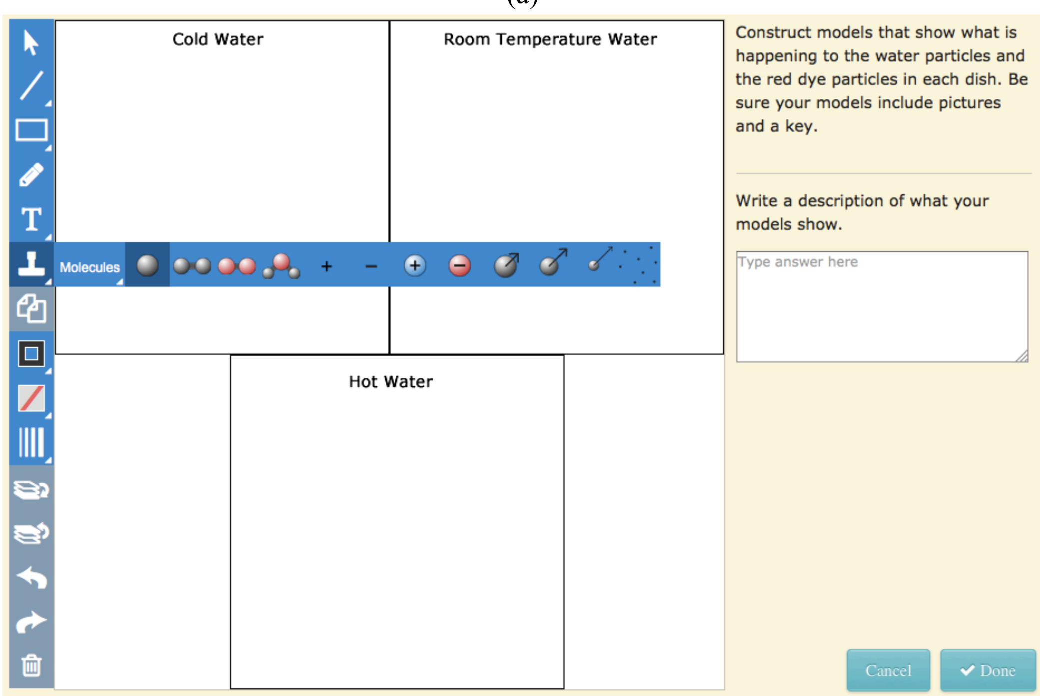
Our three-dimensional assessment design work has also entailed an effort to explore the kinds of rubrics that could accompany the tasks. The design of rubrics that could be used to measure three-dimensional learning presents particular challenges. We have explored several approaches to rubric design that exhibit trade-offs between informativeness and practical utility. For example, a single rubric integrating all three dimensions of the learning performance may be relatively simple to apply. However, because the complexity of our tasks requires students to demonstrate several distinct proficiencies, integrated rubrics may conflate these proficiencies with each other.

One solution to this problem is to “foreground” a particular dimension of interest relative to the other two (McElhaney, D’Angelo, Harris, Seeratan, Stanford, & Debarger, 2015). This foregrounding approach could address three-dimensional proficiency in a way that is anchored to a particular dimension and might be useful for specific applications. For example, foregrounding a crosscutting concept could help characterize student learning occurring longitudinally across science disciplines. The approach could not, however, be used to fully measure three-dimensional proficiency, unless multiple rubrics were applied to the same responses.

A second solution is to design multiple rubrics based on the evidence statements associated with each focal KSA. Because each assessment task targets multiple focal KSAs, each task response could be scored using multiple rubrics to distinguish the different types of proficiencies required for a task. This approach has the additional advantage of focusing scorers’ attention on specific features of student responses (potentially improving reliability) and allows the scoring decisions to be tightly linked to the focal KSAs (facilitating the separation of proficiencies in a way that informs formative reporting). Moreover, recent research illustrates benefits of constructing composite summaries of multiple rubrics that capture separable aspects of performance on a task to inform unidimensional scaling (Bernhardt & Crockett, 2014). Using this approach, separate scores on a task might be combined into a holistic score that could be modeled as a unidimensional construct.



(a)



(b)

Figure 3. (a) Example task aligned with LP E-02. (b) Drawing tool including features to support particulate representations of matter.

Rubric design approaches will also depend on how they will be used; the same rubrics used for empirical validation of tasks might not be designed in the same way as rubrics to be used by teachers to formatively assess student work. Our future work involving statistical modeling of students' responses and classroom use of tasks by teachers will explore the utility of different types of rubrics for these diverse purposes.

Ensuring the generalizability of the design process to other disciplines and grade bands is a second important challenge. Though the design process at a broad level contains no discipline or grade-band specific steps, our current efforts to apply the process to create assessment tasks in a a second domain (middle school life science) raises key questions. For example, we aimed to make our practice and crosscutting concept unpacking documents as general as possible while keeping a specific discipline (physical science) and grade band (middle school) in the sharpest focus. To what extent are these documents generalizable to other disciplines and grade bands? What is the most appropriate scope of these unpacking documents? Our work in life science also raises questions about generating the integrated dimension maps. Does the process we used to generate these maps for physical science PEs (beginning with disciplinary concepts, then integrating the other dimensions) also work for PEs in other disciplines? Our current efforts toward developing assessment tasks in life science will help us refine the design process so it is more readily generalizable across disciplines and grade-bands.

Implications

Though we have fully applied our design process only in the domain of middle school physical science, we believe it is sufficiently flexible to be adapted for the full range of disciplines and grade-bands of NGSS. A particularly intriguing application of our design process is the NGSS Engineering Design PEs. The *Framework* identifies engineering design as a discipline distinct from the earth, life, and physical sciences. Accordingly, engineering design has a unique set of disciplinary core ideas. The *Framework* also clearly articulates a vision of *integrating* science and engineering, enabling students to "explore the practical use of science" and providing "a context in which students can test their own developing scientific knowledge and apply it to practical problems." (NRC, 2012, p. 12) The complexity of assessing the integration of science and engineering clearly calls for a principled design process such as ECD. Furthermore, applications of our assessment design process are not necessarily limited specifically to the NGSS. The approach could potentially be used for any multidimensional performance construct. For example, the forthcoming Framework for K-12 Computer Science Education (www.k12cs.org) will also inform the development of standards integrating content and practices in the computer science discipline.

Finally, we believe our design process could be extended to inform the development of standards-aligned curriculum materials as well as assessment. Learning performances have the potential to serve as anchors for the design of curriculum units aligned to specific standards. Using learning performances to guide the design of both curriculum and assessments could help ensure that curriculum and assessments are well aligned to each other. This application of our design process would help curriculum developers use a principled approach to align new curriculum materials to NGSS or other complex performance standards, rather than retrofitting previously unaligned materials to the standards.

Acknowledgments

We gratefully acknowledge Angela DeBarger from Lucas Education Research; Chanyah Dahsah, Jane Lee, and Deborah Peek-Brown from Michigan State University; Gauri Vaishampayan from University of Illinois at Chicago; Dan Damelin and Frieda Reichsman from Concord Consortium; and Cynthia D'Angelo, Nonye Alozie, Reina Fujii, and Tiffany Leones from SRI International for their contributions to this research and development effort.

This material is based upon work supported by the National Science Foundation under Grant Numbers DRL – 1316903, 1316908, 1316874. Any opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

American Association for the Advancement of Science. (1993). *Benchmarks for science literacy*. New York, NY: Oxford University Press.

Bernhardt, A., & Crockett, T. (2014, June) *Transitioning to NGSS: One state's early work offers a glimpse into the future*. Paper presented at the National Conference on Student Assessment, New Orleans, LA.

DeBarger, A. H., Harris, C. J., D'Angelo, C., Krajcik, J., Dahsah, C., Lee, J., & Beauvineau, Y. (2014). Constructing assessment items that blend core ideas and science practices. In Polman, J. L., Kyza, E. A., O'Neill, D. K., Tabak, I., Penuel, W. R., Jurow, A. S., O'Connor, K., Lee, T., and D'Amico, L. (Eds.). *Learning and becoming in practice: The International Conference of the Learning Sciences (ICLS) 2014, Volume 3*. Boulder, CO: International Society of the Learning Sciences.

DeBoer, G. E., Lee, H.-S., & Husic, F. (2008). Assessing integrated understanding of science. In Y. Kali, M. C. Linn, & J. E. Roseman (Eds.), *Designing coherent science education* (pp. 153–184). New York, NY: Teachers College Press.

Duschl, R. A. (2008). Science education in three-part harmony: Balancing conceptual, epistemic, and social learning goals. *Review of Research in Education*, *32*, 268–291.

Duschl, R. & Osborne, J. (2002). Argumentation and discourse processes in science education. *Studies in Science Education*, *38*, 39-72.

Frederiksen, J. R., & Collins, A. (1989). A systems approach to educational testing. *Educational Researcher*, *18*(9), 27–32.

Fuhrman, S. H., Resnick, L. B., & Shepard, L. (2009). Standards aren't enough. *Education Week*, *29*(7), 28.

Harris, C. J., McNeill, K. L., Lizotte, D. L., Marx, R. W., & Krajcik, J. (2006). Usable assessments for teaching science content and inquiry standards. In M. McMahon, P. Simmons, R. Sommers, D. DeBaets & F. Crowley (Eds.), *Assessment in science: Practical experiences and education research* (pp. 67-88). Arlington, VA: National Science Teachers Association Press.

Harris, C. J., & Salinas, I. (2009). Authentic science learning in primary and secondary classrooms. In M. I. Saleh & M.S. Khine (Eds.), *Fostering scientific habits of mind: Pedagogical knowledge and best practices in science education* (pp. 125–144). Netherlands: Sense.

Kali, Y., Koppal, M., Linn, M. C., & Roseman, J. E. (2008). Preface. In Y. Kali, M. C. Linn, & J. E. Roseman (Eds.), *Designing coherent science education* (pp. xi–xx). New York, NY: Teachers College Press.

Krajcik, J., McNeill, K. L., Reiser, B., (2008). Learning-goals-driven design model: Developing curriculum materials that align with national standards and incorporate project-based pedagogy. *Science Education, 92*(1), 1-32.

Latour, B., & Woolgar, S. (1996). *Laboratory life: The construction of scientific facts.* Princeton, NJ: Princeton University Press.

Lehrer, R., & Schauble, L. (2006). Scientific thinking and science literacy: Supporting development in learning in contexts. In W. Damon, R. M. Lerner, K. A. Renninger & I. E. Sigel (Eds.), *Handbook of child psychology, 6th ed.* (Vol. 4). Hoboken, NJ: John Wiley and Sons.

Marx, R. W., & Harris, C. J. (2006). No Child Left Behind and science education: Opportunities, challenges, and risks. *The Elementary School Journal, 106*(5), 467–477.

McElhaney, K., D'Angelo, C., Harris, C. J., Seeratan, K., Stanford, T., & DeBarger, A. (2015, April). *Integrating crosscutting concepts into 3-dimensional scoring rubrics.* Paper presented at the annual meeting of the National Association for Research in Science Teaching, Chicago, IL.

Mislevy, R., & Haertel, G. (2006). Implications of evidence-centered design for educational testing. *Educational Measurement: Issues and Practice, 25*(4), 6–20.

National Research Council. (2000). *How people learn: Brain, mind, experience*. (Expanded ed.). Washington, DC: The National Academies Press.

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

National Research Council. (2009). *Learning science in informal environments: People, places, and pursuits*. Washington, DC: The National Academies Press.

National Research Council (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: National Academies Press.

NGSS Lead States (2013). *Next generation science standards: For states, by states*. Washington, DC: National Academies Press.

Pellegrino, J. W. (2013). Proficiency in science: Assessment challenges and opportunities. *Science, 340*(6130), 320-323.

Pellegrino, J., Chudowsky, N., & Glaser, R. (Eds.). (2001). *Knowing what students know: The science and design of educational assessment.* Committee on the Foundations of Assessment. Board on Testing and Assessment, Center for Education. Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

Pellegrino, J. W., & Hilton, M. L. (Eds.). (2012). *Education for life and work: Developing transferable knowledge and skills in the 21st century*. Washington, DC: The National Academies Press.

Pellegrino, J. W., Wilson, M., Koenig, J. & Beatty, A. (Eds.) (2014). *Developing assessments for the Next Generation Science Standards.* Washington, DC: The National Academies Press.

Rose, D. H., & Meyer, A. (2002). Teaching every student in the digital age: Universal design for learning. Alexandria, VA: ASCD.

Rose, D. H., & Meyer, A. (Eds.). (2006). *A practical reader in universal design for learning.* Cambridge, MA: Harvard Education Publishing Group.

Rose, D., Meyer, A., & Hitchcock, C. (Eds.). (2005). *The universally designed classroom.* Cambridge, MA: Harvard Education Press.

Schwendimann, B. A. (2015). Concept maps as versatile tools to integrate complex ideas: From kindergarten to higher and professional education. *Knowledge Management & E-Learning: An International Journal (KM&EL)*, *7*(1), 73-99.

Songer, N. B., Kelcey, B., & Gotwals, A. W. (2009). How and when does complex reasoning occur? Empirically driven development of a learning progression focused on complex reasoning about biodiversity. *Journal of Research in Science Teaching*, *46*(6), 610–631.

Wiliam, D. (2010). What counts as evidence of educational achievement? The role of constructs in the pursuit of equity in assessment. *Review of Research in Education*, *34*, 254–284.

Appendix A

Unpacking of the Disciplinary Core Ideas Components
PS1.A: Structure and Properties of Matter, PS3.A: Definitions of Energy, and PS3.B: Conservation of Energy and *Energy Transfer*

Component PS1.A: Structure and Properties of Matter

Elements

- Substances are made from different types of atoms, which combine with one another in various ways. Atoms form molecules that range in size from two to thousands of atoms.
- Gases and liquids are made of molecules or inert atoms that are moving about relative to each other.
- In a liquid, the molecules are constantly in contact with others; in a gas, they are widely spaced except when they happen to collide.
- In a solid, atoms are closely spaced and do not change relative locations.
- The changes of state that occur with variations in temperature or pressure can be described and predicted using these models of matter.

Component PS3.A: Definitions of Energy

Elements

- The term "heat" as used in everyday language refers both to thermal energy (the motion of atoms or molecules within a substance) and the transfer of that thermal energy from one object to another. In science, heat is used only for this second meaning; it refers to the energy transferred due to the temperature difference between two objects.
- The temperature of a system is proportional to the average internal kinetic energy and potential energy per atom or molecule (whichever is the appropriate building block for the system's material). The details of that relationship depend on the type of atom or molecule and the interactions among the atoms in the material.
- Temperature is not a direct measure of a system's total thermal energy. The total thermal energy (sometimes called the total internal energy) of a system depends jointly on the temperature, the total number of atoms in the system, and the state of the material.

Component PS3.B: Conservation of Energy and Energy Transfer

Element

- The amount of energy transfer needed to change the temperature of a matter sample by a given amount depends on the nature of the matter, the size of the sample, and the environment.

Elaboration of Ideas

- The amount of energy transfer needed to change the temperature of a matter sample by a given amount depends on the nature of the matter, the size of the sample, and the environment.
- Matter is made of atoms and molecules. These atoms and molecules are constantly in motion. All particles of matter have kinetic energy because they are in motion.
- The average kinetic energy of the particles in a substance is measured by temperature.
- Thermal energy is the total kinetic energy of the particles that make up a substance.
- The total thermal energy in a system depends on the number of particles, temperature, particle motion, mass, and types of matter.
  - The greater the number of atoms and molecules (mass), the greater the thermal energy
  - The higher temperature, the higher the thermal energy an object has.
  - When thermal energy is transferred to an object, it increases the kinetic energy of substance making its temperature increase.
  - When thermal energy is transferred from an object, its particles slow down making the temperature decrease.
  - The thermal energy of a substance is directly related to the average speed of its atoms/molecules (particle motion)
  - When the average speed of the atoms and molecules of a sample increases, the kinetic energy of the atoms and molecules increases and, therefore, the thermal energy of the sample increases.
  - When thermal energy is transferred to a substance, it makes the particles move faster. The movement of particles slows down if thermal energy is transferred from a substance to the surroundings.
  - The thermal energy of a sample is associated with types of matter that make up the sample.
  - Samples that are made of the same substance, have the same mass and are at the same temperature have the same amount of thermal energy.
  - Samples that are made of different substances may have different amounts of thermal energy even if they have the same mass and temperature.
- The state of a pure substance can be classified as a solid, a liquid, or a gas.
  - For a given substance, a solid has the least kinetic energy, and move in position, so the location of the particles does not change.
  - The particles in a liquid have more kinetic energy, and they move faster than the particles in a solid do.
  - The particles in a gas have the most kinetic energy and can move fast enough to break away from each other.
- If thermal energy is transferred from or to the sample. The state of a substance might change.
  - If enough thermal energy is transferred to a solid, it becomes a liquid and a liquid becomes a gas. The reverse happens when enough thermal energy is transferred from sample to the surroundings.

Domain Boundary

- Students are not expected to know the relationship between heat and temperature.
- Students are not expected to compare situations where both the number of particles and the temperature of the objects vary.
- Students are not expected to compare the thermal energy of objects or samples that combine more than one pure substance (e.g., mixtures).

Prior Knowledge

Students should have proficiency with the following performance expectations at the end of the previous grade band (grades 4-5):

- 4-PS3-3 Ask questions and predict outcomes about the changes in energy that occur when objects collide.
- 5-PS1-1 Develop a model to describe that matter is made of particles too small to be seen.

Student Challenges

Some students may hold the following alternative conceptions:

- Particle sizes of substances are increased when changing states from liquid to gas (Pereira & Pestana, 1991; Griffiths & Preston, 1989).
- Only things that are warm or hot have thermal energy (Hermann-Abell & Deboer, 2010).
- The thermal energy of an object is not related to the temperature of the object (AAAS Project 2061, 2008), the number of molecules that make up an object (Hermann-Abell & Deboer, 2010), the kinetic energy and speed of the molecules that make up an object (AASS Project 2061, n.d.).

References

American Association for the Advancement of Science. (1993). Benchmarks for Science Literacy. New York: Oxford University Press.

Griffiths, A. K., & Preston, K. R. (1989). An Investigation of Grade 12 Students' Misconceptions Relating to Fundamental Characteristics of Molecules and Atoms. Paper presented at the National Association for Research in Science Teaching Annual Conference. San Francisco, CA.

Herrmann-Abell, C. F., & DeBoer, G. E. (2010). Probing middle and high school students’ understanding of energy transformation, energy transfer, and conservation of energy using content-aligned assessment items. Paper presented at the National Association for Research in Science Teaching Annual Conference, Philadelphia, PA.

Pereira, M. P., & Pestana, M. E. M. (1991). Pupils’ representations of models of water. International Journal of Science Education, 13(3), 313-319.

Appendix B

Unpacking of the Science Practice Developing and Using Models

Scientific model
A scientific model is an abstract, simplified, representation of a phenomenon or system of phenomena that makes its central features explicit and visible. Models can be used to describe/illustrate phenomena and generate explanations and predictions. They include diagrams, physical replicas, mathematical representations, analogies, and computer simulations.

We distinguish three aspects of the practice. **Developing models** involves generating a representation having elements and relationships that explain a target phenomenon, representing the correspondence between these elements and the real world, and specifying the limitations of the model in explaining the target phenomenon. **Using models** involves applying a previously developed model to answering a scientific question and can include generating explanations based on the model. **Evaluating models** involves determining whether (and describing why) a model does or does not include the appropriate or necessary information to explain a phenomenon.

Integration with other Practices:

- Output from models can be used as evidence for **explanations** and **arguments**
- **Scientific arguments** critique or defend the quality or appropriateness of models
- Models are developed based on results of **data analysis**
- **Investigations** may inform the development of models or involve the use of models

Aspects of the practice

Aspect	Develop Models	Use Models	Evaluate Models
Model elements	**Specify** elements of the model (and their attributes) and **describe** why these elements are necessary	**Identify** appropriate elements of the model (and their attributes)	**Determine** whether and **describe** why the elements are/are not appropriately included, specified or identified

Relationships among elements	**Represent** the relationships or interactions among model elements and **describe** why these relationships are important	**Describe** the appropriate relationships or interactions among model elements	**Determine** whether and **describe** why the relationships among elements are/are not appropriately represented or described
Correspondence	**Represent** the correspondence between model elements/attributes and the target phenomenon or available data	**Describe** the correspondence between model elements/attributes and the target phenomenon or available data	**Determine** whether and **describe** why the correspondence is/is not appropriately represented or described
Limitations	**Specify** the limitations of the model and **describe** why these limitations exist	**Identify** the limitations of the model	**Determine** whether and **describe** why the limitations are/are not appropriately specified or identified
Explanation/ prediction		**Explain or predict** phenomena using the model	**Determine** whether and **describe** why the model does/does not appropriately explain or predict the phenomenon

Knowledge, skills and abilities for performing the practice

What KSAs do middle school students need to use in order to perform the practice?

Aspect	**Develop Models**	**Use Models**	**Evaluate Models**
Model elements	Knowledge that a model contains elements (observable and unobservable) that represent specific aspects of real world phenomena		
	Ability to **specify** the appropriate elements (and their attributes) of a model Ability to **describe** why specific elements are necessary	Ability to **identify** the elements (and their attributes) of a model	Ability to **judge** the appropriateness of elements included in the model Ability to **articulate** reasoning for making above judgment
Relationships among elements	Ability to **represent** the relationships among elements of a model Ability to **describe** why specific relationships are important	Ability to **describe** the relationships among elements of a model	Ability to **judge** the appropriateness of relationships among elements included in the model Ability to **articulate** reasoning for making above judgment
Correspondence	Knowledge that models should correspond with observations of real-world phenomena or available data		

	Ability to **represent** the correspondence of elements of a model to their real-world counterparts or available data	Ability to **describe** the correspondence of elements of a model to their real-world counterparts or available data	Ability to **judge** the correspondence between model elements and the real world represented in the model Ability to **articulate** reasoning for making above judgment
Limitations		Knowledge that models have limitations	
	Ability to **specify** the limitations of a model Ability to **describe** why these limitations exist	Ability to **identify** the limitations of a model	Ability to **judge** the specified limitation in a model Ability to **articulate** reasoning for making above judgment
Explanation/ prediction		Knowledge that models are used to generate explanations and make predictions Ability to **generate** an explanation **or make** a prediction using a model	Ability to **judge** the explanation or prediction generated based on the model Ability to **articulate** reasoning for making above judgment

Evidence for each component of the practice (rubric)

What is a high level of performance that you would expect to see for each component?

Aspect	Develop Models	Use Models	Evaluate Models
Model elements	**Specifies** only the appropriate and necessary elements (and their attributes) in the model needed to explain the target phenomenon and **describes** why these elements are necessary	**Identifies** the appropriate and necessary elements (and their attributes) in the model	**Determines** whether and **describes** why the elements are/are not appropriately included, specified or identified
Relationships among elements	**Represents** only the appropriate relationships and/or interactions among the elements in the model needed to explain the target phenomenon and **describes** why these relationships are important	**Describes** the appropriate relationships and/or interactions among the elements in the model	**Determine** whether and **describe** why the relationships among elements are/are not appropriately represented or described
Correspondence	**Represents** the correspondence	**Describes** the correspondence	**Determines** whether and **describes** why

	between model elements and the real world phenomenon or available data	between model elements and the real world phenomenon or available data	the correspondence is/is not appropriately represented or described
Limitations	**Specifies** the appropriate limitations of the model with respect to explaining the target phenomenon and **describes** why these limitations exist	**Identifies** the appropriate limitations of the model	**Determines** whether and **describes** why the limitations are/are not appropriately specified or identified
Explanation/ prediction		**Constructs** a correct and complete explanation or prediction phenomena using the model [see explanation unpacking document]	**Determines** whether and **describes** why the model does/does not appropriately explain or predict the phenomenon [see explanation unpacking document]

References

Grosslight, L., Unger, C., Jay, E., & Smith, C. L. (1991). Understanding models and their use in science: Conceptions of middle and high school students and experts. *Journal of Research in Science teaching*, *28*(9), 799-822.

Krajcik, J., & Merritt, J. (2012). Engaging Students in Scientific Practices: What Does Constructing and Revising Models Look Like in the Science Classroom?. *Science Teacher*, *79*(3), 38-41.

Lehrer, R., & Schauble, L. (2000). Developing model-based reasoning in mathematics and science. *Journal of Applied Developmental Psychology*,*21*(1), 39-48.

National Research Council (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: National Academies Press.

NGSS Lead States (2013). *Next generation science standards: For states, by states*. Washington, DC: National Academies Press.

White, B. Y., & Frederiksen, J. R. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and instruction*, *16*(1), 3-118.

Appendix C

Unpacking of the Crosscutting Concept *Cause and Effect*

Cause and effect addresses the identification of mechanisms, causal relationships, and chains of events and interactions that govern scientific phenomena. Knowledge about causal relationships is often necessary to make predictions about new situations and developing engineering solutions. We identify

Intersections with Practices

- **Explanations** articulate conditions, mechanisms, evidence. Causal mechanisms inform predictions and theories.
- **Arguments** evaluate statements about causal mechanisms and evidence.
- Patterns from **data analysis** provide evidence of causal relationships
- **Investigations** are planned and carried out to generate evidence of causal relationships
- **Models** are developed and used to provide evidence for and test hypothesized causal relationship
- Scientists **ask questions** about underlying causal mechanisms underlying phenomena
- Causal mechanisms inform **designed solutions** to problems based on science principles

Aspects and examples

Aspect	Description
Causes	**Identify** or **describe** the cause(s) that lead to the given effect(s) under various conditions
Effects	**Identify** or **describe** the effect(s) that result from the given cause(s) under various conditions • One or more effects directly resulting from specific conditions • Probabilistic description of possible effects
Conditions	**Identify** or **describe** the conditions under which causal relationships occur • Qualitative description or numerical range
Mechanism/ Intermediate events	**Identify** or **describe** the chain of intermediate events that links cause and effect **Identify** or **articulate** scientific principles (e.g. specific disciplinary concepts or underlying models/theories) that justify how/why cause and effect are linked
Evidence	**Describe** or **provide** observations/data that constitute evidence for the causal relationships. Evidence may come from the real-world or a model
Predictions/ theories	**Articulate** predictions that are based on an identified causal mechanism **Describe** how identified causal mechanisms inform established theories
Solutions	**Design** a solution based on the identified causal mechanism

Knowledge, Skills, and Abilities (KSAs) needed to understand instances of cause and effect

Aspect	KSAs
Causes	Ability to **identify** the cause(s) that lead to the given effect(s) under various conditions Knowledge about the directional correspondence between causes and effects
Effects	Ability to **identify** the effect(s) that result from the given cause(s) under various conditions Ability to describe multiple effects probabilistically Knowledge about the directional correspondence between causes and effects
Conditions	Ability to **identify** a numerical range of conditions under which specific causal relationships occur Ability to **describe** (qualitatively) the conditions under which specific causal relationships occur Knowledge that cause and effect relationships can be (are?) conditional
Mechanism/ Intermediate events	Ability to **identify** or **describe** the chain of intermediate events that links cause and effect Ability to **identify** or **articulate** scientific principles (e.g. specific disciplinary concepts or underlying models/theories) that justify how/why cause and effect are linked Knowledge of scientific principles and/or underlying models/theories in the discipline
Evidence	Ability to **identify** observations/data that constitute evidence for the causal relationships Ability to interpret available data
Predictions/ theories	Ability to **articulate** a prediction based on an identified causal mechanism Ability to **describe** how identified causal mechanisms inform established theories Knowledge about the nature of established scientific theories
Solutions	Ability to **design** a solution based on the identified causal mechanism

Evidence for each aspect of the crosscutting concept (rubric)
What is a high level of performance that would be expected for each aspect?

Aspect	Description
Causes	Correctly **identifies** or **describes** the cause(s) that lead to the given effect(s) under various conditions
Effects	Correctly **identifies** the effect(s) that result from the given cause(s) under various conditions Correctly **describes** multiple possible effects probabilistically and specifies their relative probabilities of occurrence
Conditions	Correctly **identifies** or **describes** the conditions under which specific causal relationships occur using an appropriate qualitative description or numerical range
Mechanism/ Intermediate events	Correctly **identifies** or **describes** a sequence of intermediate events that links cause and effect Correctly **identifies** or **articulates** scientific principles (e.g. specific disciplinary concepts or underlying models/theories) that justify how/why cause and effect are linked
Evidence	**Describes** or **provides** appropriate and sufficient observations/data that constitute evidence for the causal relationships
Predictions/ theories	**Articulates** appropriate or correct predictions based on an identified causal mechanism Correctly **describes** how identified causal mechanisms inform established theories
Solutions	**Designs** an appropriate solution that addresses a given problem based on the identified causal mechanism

References

Brewer, W. F., Chinn, C. A., & Samarapungavan, A. (1998). Explanation in scientists and children. *Minds and Machines*, *8*(1), 119-136.

Metz, K. E. (1991). Development of explanation: Incremental and fundamental change in children's physics knowledge. *Journal of Research in Science Teaching*, *28*(9), 785-797.

National Research Council (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: National Academies Press.

NGSS Lead States (2013). *Next generation science standards: For states, by states*. Washington, DC: National Academies Press.

Thagard, P. (1992). *Conceptual revolutions*. Princeton, NJ: Princeton University Press.