Constructing Assessment Tasks that Blend Disciplinary Core Ideas, Crosscutting Concepts, and Science Practices for Classroom Formative Applications

Christopher J. Harris | Joseph S. Krajcik | James W. Pellegrino | Kevin W. McElhaney

With: Angela Haydel DeBarger | Chanyah Dahsah | Daniel Damelin | Cynthia M. D'Angelo | Louis V. DiBello | Brian Gane | Jane Lee

May 2016

SRI Education™

A DIVISION OF SRI INTERNATIONAL

Authors

Christopher J. Harris, Center for Technology in Learning, SRI Education
Joseph S. Krajcik, CREATE for STEM Institute, Michigan State University
James W. Pellegrino, Learning Sciences Research Institute, University of Illinois at Chicago
Kevin W. McElhaney, Center for Technology in Learning, SRI Education

Collaborators

Center for Technology in Learning, SRI Education
Learning Sciences Research Institute, University of Illinois at Chicago
CREATE for STEM Institute, Michigan State University
Concord Consortium

Acknowledgments

We gratefully acknowledge Deborah Peek-Bown and Phyllis Pennock from Michigan State University; Gauri Vaishampayan and Krystal Madden from University of Illinois at Chicago; Frieda Reichsman from Concord Consortium; and Tiffany Leones, Reina Fujii, and Nonye Alozie from SRI International for their contributions to this research and development effort.

Suggested Citation

Harris, C. J., Krajcik, J. S., Pellegrino, J. W., & McElhaney, K.W. (2016). *Constructing assessment tasks that blend disciplinary core Ideas, crosscutting concepts, and science practices for classroom formative applications.* Menlo Park, CA: SRI International.

This material is based upon work supported by the National Science Foundation under Grant Numbers DRL – 1316903, 1316908, 1316874. 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.

Abstract

How do we measure knowledge in use? In this paper we describe how we use principles of evidence-centered design to develop classroom-based science assessments that integrate three dimensions of science proficiency—disciplinary core ideas, science practices, and crosscutting concepts. In our design process, we first elaborate on, or "unpack", the assessable components of the three dimensions. We then use these elaborations to specify a set of claims called *learning performances* that describe what students need to be able to know and do in order to meet knowledge-in-use learning goals, such as the performance expectations articulated in the U.S. Next Generation Science Standards. Learning performances are crafted as knowledge-in-use statements that integrate aspects of the three dimensions, but are smaller in scope than end-of-grade-band performance expectations. Next, we define task features to elicit from students the desired evidence of proficiency. Our final step entails using design patterns derived from specifying learning performances, specifying evidence, and defining task features to construct tasks that measure science proficiency. We present our design approach, provide examples of tasks, and consider implications of this work for next generation science assessment.

A key challenge in shaping science learning for the 21st century will be to develop new measures of learning that take into account what it means to be proficient in science (Pellegrino, 2013). The emergent view of proficiency, grounded in learning sciences research, emphasizes using and applying knowledge in the context of disciplinary practice. Referred to as knowledge-in-use, this perspective on science proficiency is a centerpiece of the National Research Council's (NRC) Framework for K-12 Science Education (NRC, 2012), embodied in the new U.S. national standards (NGSS Lead States, 2013) and emphasized in the NRC report on developing assessments to measure science proficiency (Pellegrino, Wilson, Koenig, & Beatty, 2014). Central to this view is that disciplinary core ideas, crosscutting concepts that span across science domains, and science practices should be integrated so that science instruction engages students in applying knowledge to make sense of phenomena and solve problems. Accordingly, as students actively do science, they deepen both their conceptual understanding of content as well as their ability to engage in the authentic practices of science. In this paper, we describe our principled and scalable approach for designing assessment tasks that measure student proficiency with new science learning goals that blend disciplinary core ideas and crosscutting concepts with practices. These assessment tasks are intended for formative use within classroom instruction. Drawing on prior research from assessment and curriculum design (e.g., DeBarger, Krajcik, & Harris, 2014; DeBarger, Penuel, & Harris, 2015; Krajcik, McNeill, & Reiser, 2008; Harris, McNeill, Lizotte, Marx, & Krajcik, 2006), we present our design approach, provide examples, and consider implications for classroom-based science assessment.

Rationale

The prior generation of U.S. science standards (e.g., NRC, 1996, 2000) treated disciplinary content and inquiry essentially as separate strands of science learning, and assessments followed suit. In some respects, the form the standards took contributed to this separation: content standards stated what students should know, and inquiry standards stated what they should be able to do. Consequently, assessments measured the knowledge and practice components separately. The shift to integrating science practices with disciplinary core ideas and crosscutting concepts, as emphasized in the U.S. *Next Generation Science Standards* (NGSS; NGSS Lead States, 2013), is based upon studies of authentic, professional scientific practice and a wealth of research studies about student learning, especially in the learning sciences (c.f., recent synthesis reports such as *Taking Science to School* [NRC, 2007] and *A Framework for K-12 Science Education* [NRC, 2012]). This research corpus points to the importance of integrating content (i.e., disciplinary core ideas and crosscutting concepts) and practice by emphasizing that rich science learning requires tight coupling of what students know and what they can do. This idea of science performance (NGSS Lead States, 2013) presents a different way of thinking about science proficiency by emphasizing the knowledge and skills students need to engage in real-world science, such as solving problems, reasoning with evidence, and explaining phenomena (NRC, 2012). It also signifies that measuring science proficiency solely as acquisition of core content knowledge is no longer sufficient (see e.g., Pellegrino, 2013).

Knowledge-in-use learning goals comprise the standards in the NGSS and are articulated as *performance expectations*. Each NGSS performance expectation combines a science or engineering practice, disciplinary core idea, and crosscutting concept into a single statement of what is to be assessed at the end of grade level or grade band. A performance expectation incorporates all three dimensions of knowledge in use by asking students to apply disciplinary and crosscutting knowledge while engaging in a science or engineering practice.

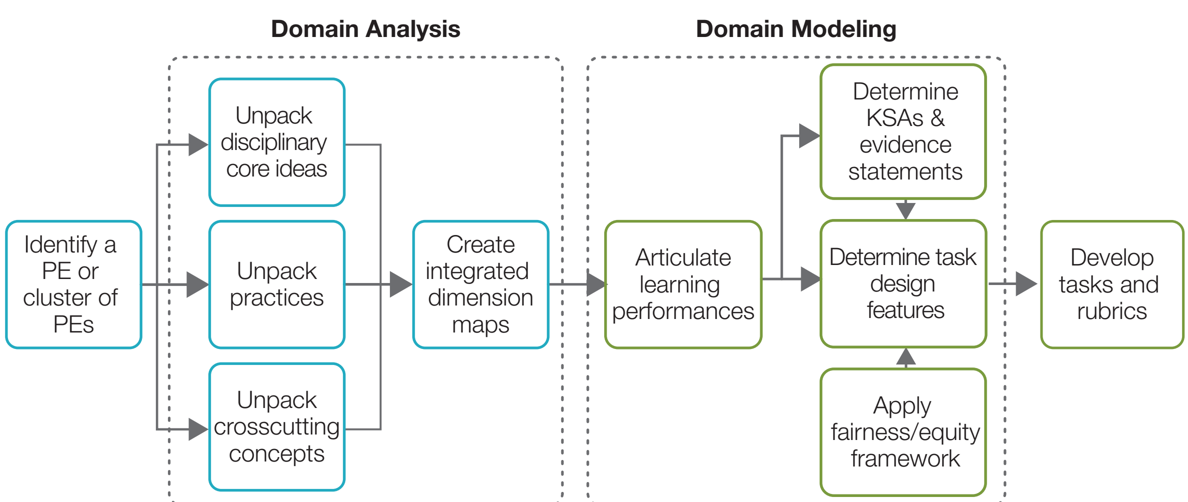
This integrated, knowledge-in-use perspective poses challenges for classroom-focused assessment design. Currently, there are very few examples of assessments that integrate science content and practices in a manner consistent with a knowledge-in-use perspective. Importantly, performance expectations represent summative (i.e., *end-of-grade-band*) performance targets. Although providing some degree of specification, performance expectations do not on their own provide sufficient detail to create assessments that could be used formatively by teachers to gain insight into their students' progress toward meeting them. Thus, a major design challenge is how to create instructionally supportive assessment tasks that integrate the three NGSS dimensions and align to NGSS performance expectations. There is tremendous need for this assessment design work, as assessment will play a central role in supporting implementation of the new directions in science education both in the U.S. and internationally.

Our approach to meeting this challenge uses principles of evidence-centered design (ECD) (Almond, Steinberg, & Mislevy, 2002; Mislevy & Haertel, 2006), which has gained increasing attention as a comprehensive approach for principled assessment design and validation. ECD has been used in wide-ranging assessment design contexts, from the development of large scale, high stakes assessments to the design of classroom-based assessments and other proximal or close measurement instruments. ECD emphasizes the evidentiary base for specifying coherent, logical relationships among the (a) learning goals that comprise the constructs to be measured (i.e., the claims articulating what students know and can do); (b) evidence in the form of observations, behaviors, or performances that should reveal the target constructs; and (c) features of tasks or situations that should elicit those behaviors or performances. The need for a principled approach to assessment design, such as ECD, was explicitly discussed in the NRC's report on developing assessments aligned to the NGSS (Pellegrino et al., 2014).

Our Evidence-Centered Design Process

To address the goal of formative assessment aligned with the NGSS, we use ECD to systematically unpack NGSS performance expectations and synthesize the unpacking into multiple components that we call learning performances, which can guide assessment task development for formative use. Our learning performances constitute knowledge-in-use statements that incorporate aspects of disciplinary core ideas, science practices, and crosscutting concepts that students need to be able to integrate as they progress toward achieving larger end-of grade-band performance expectations. Our design process, summarized in Figure 1 and described below, enables us to derive a set of learning performances from a performance expectation or clustered set of performance expectations in a principled way. This process involves three distinct phases – 1) *domain analysis*, which involves unpacking of the three NGSS dimensions in the performance expectations to understand the assessable components, 2) *domain modeling*, constructing learning performances and specifying design patterns for tasks associated with them, and 3) *task construction,* using design patterns to create tasks and accompanying rubrics. Although Figure 1 illustrates what appears to be a fairly linear process that begins with selecting one or more performance expectations to unpack and then proceeds forward in a step-by-step fashion, it is important to realize that the process is very iterative. The step of articulating learning performances, for example, might lead a designer to revisit and refine integrated dimension maps. Alternatively, a designer might decide to conduct the unpacking and mapping in tandem, thus gradually building out the dimension maps as the unpacking unfolds. Below, we illustrate the process using an NGSS performance expectation from middle school physical science, MS-PS1-2: *Analyze and interpret data on the properties of substances before and after the substances interact to determine if a chemical reaction has occurred.*

Figure 1. Design process for developing formative assessment tasks aligned to NGSS



Domain Analysis – Unpacking the Dimensions of Performance Expectations

In ECD, domain analysis typically entails gathering substantive information about how knowledge is acquired and used in a given domain, such as physical science or life science. The Framework and the NGSS specify meaningful ways to integrate the core ideas, crosscutting concepts, and practices to promote assessment of learning in the domain. The domain analysis informs the construction of learning performances that describe the knowledge-in-use that students need to demonstrate as they progress toward achieving the target performance expectations. The process for articulating learning performances begins with a purposive domain analysis of the three NGSS dimensions that comprise the performance expectations. The resources that we use for unpacking include the *Framework*, NGSS and NGSS appendices, and research literature on the dimensions and their components. Unpacking the dimensions of the target performance expectation(s) is the foundational step in our design approach, as it provides the anchors constituting each dimension and provides a clear focus for what should be elicited in assessment tasks. We use the elaborations from the unpacking to create comprehensive integrated dimension maps that provide a visual representation of the target performance expectations.

Unpack disciplinary core ideas. In the domain analysis phase of evidence-centered design, we first unpack disciplinary core ideas associated with an NGSS performance expectation or a cluster of performance expectations at a given grade level or grade band. Unpacking disciplinary core ideas entails thoughtful consideration of ideas in relation to students' grade level, or expected level of expertise. It requires that when focusing on an aspect of a disciplinary core idea, we elaborate the meaning of key sub-ideas, define clear expectations for what ideas students would be expected to use, demarcate boundaries for what students are or are not expected to know, identify background knowledge that is expected of students in order to develop a grade-level-appropriate understanding of a disciplinary core idea, and identify research-based problematic student ideas. We also identify phenomena that provide compelling examples of the disciplinary core idea. In Table 1 we provide excerpts from our unpacking of aspects of the disciplinary core idea of Matter and Its Interactions. The aspects we unpack relate to the component ideas of chemical reactions from the NGSS performance expectation MS-PS1-2.

Unpack the science practices. Our unpacking of the science practices involves clearly articulating the essential grade-band appropriate performance for each practice. We articulate specific aspects of practices students are to perform, specify the evidence required for students to demonstrate a high level of proficiency with a practice, identify prior knowledge that is required of students to demonstrate the practice, and identify common challenges that students may encounter as they are developing sophistication with the practice. We also identify productive intersections between the practice and other science practices. To accomplish this unpacking, we reference Appendix F from NGSS (NGSS Lead States, 2013) as well as the research literature on science practices. Below, in Table 2, we provide a brief example of unpacking the science practice of analyzing and interpreting data.

Unpack Crosscutting Concepts. Crosscutting concepts, such as *Patterns* and *Cause and Effect*, are ideas that apply across science disciplines. Unpacking crosscutting concepts involves identifying the important aspects of each, as well as how the crosscutting concepts intersect with targeted science practices and within a particular set of disciplinary core ideas.

Table 1: Excerpts from Unpacking Aspects of a Disciplinary Core Idea related to Chemical Reactions

Aspect of a Disciplinary Core Idea (aspect shown in italics)	• Substances react chemically in characteristic ways. In a chemical process, the atoms that make up the original substances are regrouped into different molecules, and these *new substances have different properties from those of the reactants.*
Elaborating the meaning of key sub-ideas	• Properties of substances are characteristics [quality or condition] of substances that can be observed or measured • *Characteristic properties* are properties that are independent of the amount of a sample and that can be used to identify substances
Defining expectations for understanding (within the target grade band)	• At the middle school level, students should learn (1) that each pure substance has characteristic properties that can be used to identify it and that (2) characteristic properties can be measured and used to determine that new substances produced from a chemical reaction are different from the original substances
Identifying assessment boundaries (for the target grade band)	• At the middle school level, students are not expected to know the term bond or how chemical bonds are formed or broken during chemical reactions
Prerequisite knowledge	• Knowledge of how to make observations and measurements to identify substances based on their properties
Student challenges	• Students often believe that the total mass decreases during a chemical reaction when a gas is produced (e.g., Nussbaum, 1985)
Relevant phenomena	• Everyday examples of reactions include combustion (e.g., burning of wood, sugar, steel wool), decomposition reactions (e.g., rotting of bananas and electrolysis of water into oxygen), and mixing (e.g., acid-base reactions) • Pure substances are made from a single type of atom or molecule and include sugar (sucrose), sodium chloride, carbon dioxide, oxygen, ammonia, and water.

Table 2: Unpacking the Science Practice of Analyzing and Interpreting Data

Aspects of the practice	• Organize data to highlight patterns, such as in a visual display (e.g., table, graph, flowchart) • Summarize data using descriptive statistics • Identify patterns (e.g. similarities and differences, causal and correlational, linear and nonlinear) • Identify sources of measurement variation or outlying data and determine how to address them
Intersections with other practices	• Patterns uncovered by data analysis and interpretation may constitute evidence for explanations • Models should be consistent with available real world data. Models can produce data for interpretation and analysis. • Scientific arguments evaluate the appropriateness/completeness of data analyses, the consistency of data analysis with a hypothesis, theory, or model, or the strength of a conclusion that can be inferred from data. • Methods of data analysis and interpretation are appropriate to specific scientific questions • Scientists communicate scientific information using descriptions and visual displays of analyzed data • Scientists use mathematical and computational approaches to interpret and analyze data
Evidence Required to Demonstrate Practice	• Student organizes data in a clear way that highlights patterns that are relevant or meaningful to a scientific question • Student uses appropriate descriptive statistics to summarize data in a way that addresses a scientific question • Student identifies relevant or meaningful patterns that address a scientific question • Student identifies relevant sources of measurement variation or outlying data and address them appropriately in the analysis
Prerequisite Knowledge	• Knowledge of types of patterns and relationships among variables (e.g., causation, correlation, linearity, nonlinearity) • Knowledge about statistical methods used to summarize data • Knowledge that data collected from the real world involve measurement variation and outliers
Student Challenges	• Students struggle to interpret data from experiments (e.g., Zimmerman, 2000) • Students struggle to develop informative representations of data (e.g. Lehrer, 2007)

Similar to our unpacking of practices, we also specify the evidence required for a student to demonstrate a high level of proficiency with the crosscutting concept. To guide our unpacking, we reference Appendix G from NGSS, the *Framework*, and the research literature on crosscutting concepts. Table 3 illustrates unpacking the crosscutting concept of *Patterns*.

Creating integrated dimension maps. The unpacking process reveals the essential elements of each of the three dimensions encompassed in the target performance expectations. We use these elaborations to develop integrated dimension maps that lay out the dimensional "terrain" for fully achieving each performance expectation. The maps are visual representations that describe the essential disciplinary core idea relationships and link them to aspects of the targeted crosscutting concepts and science practices (or to closely related crosscutting concepts and practices as identified by the unpacking process). Each map illustrates how the three dimensions work together to define proficiency with a performance expectation and, importantly, shows a range of possible ways to combine aspects of the three dimensions in an assessment. These maps are essential to the principled articulation of three-dimensional learning performances that coherently represent the target performance expectations.

To create an integrated dimension map, we first use the unpacking of one or more disciplinary core ideas to develop a concept map that illustrates the relationships between core sub-ideas. The relationships between sub-ideas are then linked via appropriate crosscutting concepts and science practices. In this way, a concept map provides an organizing structure for considering how crosscutting concepts and science practices will work together with core sub-ideas to represent the breadth of the performance expectation. Once all three dimensions are brought together in a visual representation (i.e., integrated dimension map), we use the map to help specify the learning performances.

Table 3: Unpacking the Crosscutting Concept of Patterns

Key Aspects	• Ability to identify the presence of patterns in phenomena or data • Ability to characterize the strength, direction, or nature of patterns in phenomena or data • Ability to classify objects or relationships into types according to similarities or differences • Ability to describe why patterns exist and exhibit specific characteristics
Intersections with Practices	• Explanations address how and why particular patterns occur • Models describe observed patterns or predict patterns • Data analysis serves to identify and characterize patterns
Evidence Required to Demonstrate Application	• Students must demonstrate that they can identify, characterize, classify, and describe the reason for the occurrence of three types of patterns: • Repeated occurrences, such as spatially or temporally repeating objects or entities (e.g., extended atomic structures; phase changes) • Similarities, differences, and comparisons of 1) amount or degree across quantities or properties and 2) categories/types of entities: (e.g., comparing physical properties before and after substances interact; distinguishing states and types of matter) • Correlations and trends, such as positive and negative, linear and nonlinear, strong and weak (e.g., relating particle motion, temperature, kinetic energy, changes in thermal energy, and amount of substance)
Prerequisite Knowledge	• Knowledge that patterns are regularly occurring shapes or structures and repeated events, or relationships that can be used to classify objects or attributes • Knowledge about the characteristics of specific types of patterns, such as the frequency of a repeating event or the strength of a correlation between two variables • Relevant disciplinary knowledge needed to identify, characterize, and explain observed patterns

Domain Modeling – Specifying a Knowledge-in-Use Design Pattern

Leveraging the unpacking of science practices, crosscutting concepts, and disciplinary core ideas described above, we then move toward specifying a knowledge-in-use assessment argument. In this phase, we consider relationships among the claims we want to make about what students know and can do, evidence that would demonstrate competency with respect to these claims, and features of tasks to elicit the desired evidence. Our claims, evidence, and task features reflect a knowledge-in-use perspective in that we emphasize the application of core ideas and crosscutting concepts through engagement in a science practice. Each claim takes the form of what we refer to as a *learning performance* that clearly describes what we expect students to demonstrate to provide evidence that they have achieved an aspect of a performance expectation (McElhaney et al., 2016). Learning performances represent a keystone in the evidence-based argument that our assessment tasks represent the NGSS performance expectations for formative assessment purposes. As described below, our design process enables us to derive a set of learning performances from a performance expectation in a principled way that ensures the learning performances meet these requirements.

Articulating learning performances. We use the integrated dimension map to articulate and/or refine a set of knowledge-in-use claims called *learning performances* that collectively describe the proficiencies that students need to demonstrate in order to meet a performance expectation. A single learning performance is crafted as a knowledge-in-use statement that is smaller in scope and partially represents a performance expectation. Each learning performance describes an essential part of a performance expectation that students would need to achieve at some point during instruction to ensure that they are progressing toward achieving the more comprehensive performance expectation. Together, a set of learning performances provides the detail needed to create a coherent and bundled set of assessment tasks that would provide evidence that students can use and apply the knowledge aligned to a performance expectation or cluster of performance expectations. In this way, learning performances are akin to learning goals that take on the three-dimensional structure of the performance expectations—they articulate and integrate assessable aspects of performance that build toward the more comprehensive performance expectation. Table 4 shows two learning performances for the performance expectation MS-PS1-2.

Specifying design patterns. Before we develop assessment tasks, we specify a design pattern (Mislevy & Haertel, 2006) for each learning performance. The design patterns serve to complete the documentation of the assessment argument connecting task designs to performance expectations. Design patterns include numerous elements that guide the principled development of tasks that elicit evidence of proficiency with the learning performance. *Focal Knowledge, Skills, and Abilities*

Table 4: An NGSS Performance Expectation and Two Related Learning Performances

MS-PS1-2: *Analyze and interpret data on the properties of substances before and after the substances interact to determine if a chemical reaction has occurred.*	
Learning performances (LPs) for MS-PS1-2	
LP 1:	Students analyze and interpret data to determine whether substances are the same based upon patterns in characteristic properties.
LP 2:	Students construct a scientific explanation about whether a chemical reaction has occurred using patterns in data on properties of substances before and after the substances interact.

(KSAs) refer to the proficiencies to be targeted by the assessment task. We articulate multiple KSAs for a learning performance to capture the range of proficiencies needed to demonstrate that learning performance. Evidence statements articulate the observable features of student performance that can provide evidence of a high level demonstration of the learning performance. Evidence statements inform the development of both tasks and scoring rubrics. *Characteristic task features* describe the attributes that are common across all the tasks for a learning performance. *Variable task features* describe the features that can vary across tasks, such as the level of scaffolding to vary task difficulty. Table 5 illustrates a design pattern for Learning Performance 1 articulated in Table 4 above. The learning performance design patterns inform the design of tasks and rubrics that integrate the three NGSS performance dimensions.

An important consideration for our task design process is the application of an equity/fairness framework to help ensure that our tasks are accessible and fair to students of diverse cultural, linguistic, and socioeconomic backgrounds. Our framework draws from Universal Design for Learning (UDL) (Rose & Meyer, 2006; Rose, Meyer, & Hitchcock, 2005) – which articulates a set of guiding principles for designers to accommodate individual differences – and is informed by research studies on fair and equitable assessment practices in science (e.g., Luykx et al, 2007; Wolf & Leon, 2009). We use the framework to help articulate task design features for our design patterns and to apply these features to the design and refinement of tasks to ensure their accessibility and fairness to students in diverse classroom settings.

Table 5: Knowledge-in-Use Design Pattern for a Learning Performance

Learning Performance (Claim)	• Learning Performance 1: Students analyze and interpret data to determine whether substances are the same based upon patterns in characteristic properties.
Focal Knowledge, Skills, and Abilities (KSAs)	• Ability to apply the scientific principle that substances can be identified by their characteristic properties • Ability to determine whether substances are the same or different using data on properties of substances • Ability to support a statement about the identity of substances based on similarities or differences in data about characteristic properties of substances
Evidence Required to Demonstrate Proficiency	• A statement that two substances are the same or different. • A statement identifying all available characteristic properties of the substances as the same (e.g. density, melting point, boiling point, solubility, flammability and odor), or that none of the characteristic properties of the given substances are the same • A statement that the same substances always have the same set of characteristic properties or that different substances have at least one different characteristic property.
Characteristic Task Features	• Assessment is limited to analysis of the following characteristic properties: density, melting point, boiling point, solubility, flammability, color, and odor. • The term “substance” means a pure substance (not a mixture of substances). • Tasks provide data in a table about characteristic and/or noncharacteristic properties of several substances. • Tasks prompt students to identify whether substances are the same or different and to justify their choice. • 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	• Numbers of substances included in the data table • State of the substances in question (i.e., solid, liquid, or gas state) • Types and numbers of characteristic and non-characteristic properties included as data • Task scaffolding features to help elicit relevant data patterns and scientific principles • Tasks use visual aids to support comprehension by students with diverse linguistic and visual processing abilities

Task Construction – Developing Three-Dimensional Tasks and Rubrics

The final phase of the design process involves using the design patterns to construct assessment tasks aligned with each learning performance. The task designs make use of both characteristic and variable task features, allowing for the development of multiple tasks within a 'family' that vary in difficulty level while maintaining alignment with the learning performance. The task design process also considers the ways student responses will be scored and evaluated for evidence of the focal KSAs.

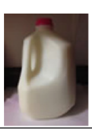
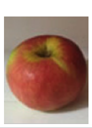
We iteratively refine the design of tasks using several steps, including (1) think-aloud sessions that examine whether tasks are comprehensible to students and whether they elicit three dimensional proficiency, (2) judgments by independent experts on the alignment of tasks with learning performances and of learning performances with performance expectations, (3) an equity/fairness review to ensure tasks reflect fair assessment design principles, and (4) classroom studies with teachers, who provide design feedback on tasks for formative use.

Task Examples. Figure 2 and Figure 3 illustrate physical science assessment tasks designed for classroom use. They were developed to assess a learning performance aligned with the NGSS performance expectation MS-PS1-2. Both are from a set of tasks developed using the design pattern for learning performance 1 that emphasizes uncovering patterns on properties of substances through data analysis and interpretation (see table 5). The first task,

Figure 2. Are all sugars the same? Physical science assessment task for learning performance 1.

Amy wondered whether the sugars found in honey, milk, sugarcane, and apples are the same kind of sugar. To find this out, she chemically removed a sample of sugar from each food and recorded the properties of the sugars in a data table shown below.

Data Table: Characteristic properties of sugars found in 4 different foods

Source of Sugar Sample	Density	Solubility in Water	Melting Point
Honey 	1.69 g/cm^3	Yes	103°C
Milk 	1.53 g/cm^3	Yes	202°C
Sugar Cane 	1.59 g/cm^3	Yes	186°C
Apple 	1.70 g/cm^3	Yes	103°C

Use the information in the data table to help Amy determine whether any of the foods have the same type of sugar. Support your answer with the data and with what you know about the properties of matter.

Figure 3. Are any of the liquids the same? Physical science assessment task for learning performance 1.

Miranda found four different bottles filled with unknown pure liquids. She measured the mass, volume, and boiling point of the liquid samples, and also calculated the density of each. The data are displayed in Table 1.

Table 1. Data on four unknown pure liquids.

Liquid Sample	Mass	Volume	Density	Boiling Point
1	6.10 g	6.10 cm^3	1.00 g/cm^3	100° C
2	5.43 g	6.10 cm^3	0.890 g/cm^3	211° C
3	9.38 g	10.20 cm^3	0.920 g/cm^3	300° C
4	9.08 g	10.20 cm^3	0.890 g/cm^3	211° C

Miranda wondered if any of the liquids are the same substance. Help Miranda by responding to the following two questions:

1. Which information from the data table would you use to determine whether any of the liquids are the same substance? Be sure to tell why.
2. Based on the information in the table, what conclusion can you make about whether any of the liquids are the same? Support your answer with what you know about the properties of matter.

Are all sugars the same? (Figure 2), and the second task, *Are any of the liquids the same?* (Figure 3), share several characteristic features. These features include data presented in a table format, prompts to determine whether substances are the same, and straightforward language to reduce text demand for middle school students. The tasks differ on their variable features, including types of characteristic properties, use of relevant and irrelevant data, scaffolding in the prompts, and use of visual aids.

Implications and Conclusions

At this time, a critical need exists for research and development of high-quality assessments that align with the standards in NGSS that express knowledge-in-use learning goals (Pellegrino, 2016). Moreover, teachers need to be able to use these tasks in classrooms to provide themselves and students with information about their students' progress towards achieving the NGSS performance expectations. Having exemplary assessment tasks that integrate the three NGSS dimensions and that can be used formatively will be important to multiple stakeholders. Teachers, students, parents and school officials are interested in using high quality assessments that support STEM college and career readiness. Assessment researchers need to better understand the design principles and psychometric properties of assessments that integrate disciplinary core ideas, crosscutting concepts and science practices. Science education researchers wish to better understand the implications of widespread adoption of a three-dimensional learning perspective, including the development and evaluation of new science curricula, teacher professional development, and large-scale standardized assessment. Science educators and policy makers want assessments that reliably measure the knowledge and abilities that are needed to engage in and support authentic science inquiry in the classroom.

Our design approach has several important advantages. First, it reflects a broadly accessible vision of how to design NGSS-aligned assessments and provides a systematic approach for documenting principled design decisions. Our ECD-based process allows us to explicitly link task design features to the evidence required to demonstrate proficiency with performance expectations. The process also supports the articulation of task features that promote usability across diverse classroom settings. Another advantage is that our approach is not discipline or grade-band specific—we expect it to generalize from our initial work in middle school physical science to the other science and engineering disciplines and grade-bands—and we are presently using the approach to develop tasks and accompanying rubrics that address performance expectations in middle school life science.

While our design approach has important advantages, challenges also exist. One central question is whether rubrics should integrate the NGSS dimensions into a single score or separately evaluate aspects of performance for all the three dimensions. This involves issues related to ease of use and feasibility, including the extent to which each of the three performance components are separable and identifiable. Teachers will also need professional development on how to use these tasks in the classroom. Thus, creating models of how three-dimensional tasks can be used formatively in the classroom will be instrumental for effective classroom use.

Developing a coherent and consistent approach to science education depends upon having high-quality assessments of student learning that align to performance expectations in NGSS. Our ECD-guided assessment design methodology is particularly well suited to assuring that necessary new assessments accurately measure the *integration* of disciplinary core ideas and crosscutting concepts with science practices in a coherent and consistent manner. Our ongoing program of research and development aims to provide answers to critical questions related to the design and use of next generation science assessments.

References

Almond, R. G., Steinberg, L. S., & Mislevy, R. J. (2002). Enhancing the design and delivery of assessment systems: A four-process architecture. *Journal of Technology, Learning, and Assessment, 1 (5).* Retrieved from http://ejournals.bc.edu/ojs/index.php/jtla/article/view/1671

DeBarger, A. H., Krajcik, J. S., & Harris, C. J. (2014, April). *Using the Framework to design assessments of modeling in physical and Earth science.* Paper presented at the annual meeting of the National Association for Research in Science Teaching, Pittsburgh, PA.

DeBarger, A. H., Penuel, W. R., Harris, C. J., & Kennedy, C. A. (2015). Building an assessment argument to design and use next generation science assessments in efficacy studies of curriculum interventions. A*merican Journal of Evaluation.* DOI: 10.1177/1098214015581707

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.

Krajcik, J., McNeill, K. L. , & Reiser, B. J. (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.

Lehrer, R. (2007). Introducing Students to Data Representation and Statistics. In K. Watson & K. Beswick (Eds). Mathematics: Essential Research, Essential Practice. *Proceedings of the 30th annual conference of the Mathematics Education Research Group of Australiasia,* Vol. 1, pp. 22-41. Adelaide: MERGA.

Luykx, A., Lee, O., Mahotiere, M., Lester, B., Hart, J., & Deaktor, R. (2007). Cultural and home language influences on children's responses to science assessments. *The Teachers College Record, 109(4*), 897-926.

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 annual conference of the National Association for Research in Science Teaching, Baltimore, MD.

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. (1996). *National Science Education Standards*. Washington, DC: National Academy Press.

National Research Council. (2000). *Inquiry and the National Science Education Standards.* Washington, DC: National Academy Press.

National Research Council (2007). T*aking science to school: Learning and teaching science in grades K-8*. Washington, DC: 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.

Nussbaum, J. (1985). The particulate nature of matter in the gaseous phase. In R. Driver, E. Guesne, & A. Tiberghien (Eds.), *Children's ideas in science* (pp. 124-144). Milton Keynes: Open University Press.

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

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

Pellegrino, J. W. (2016). *21st Century Science Assessment: The Future Is Now*. (SRI Education White Paper). Menlo Park, CA: SRI International.

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

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

Wolf, M. K., & Leon, S. (2009). An investigation of the language demands in content assessments for English language learners. *Educational Assessment, 14(3-4),* 139-159.

Zimmerman, C. (2000). The development of scientific reasoning skills. *Developmental Review, 20(1),* 99-149.

SRI Education™

SRI Education, a division of SRI International, is tackling the most complex issues in education to identify trends, understand outcomes, and guide policy and practice. We work with federal and state agencies, school districts, foundations, nonprofit organizations, and businesses to provide research-based solutions to challenges posed by rapid social, technological and economic change. SRI International is a nonprofit research institute whose innovations have created new industries, extraordinary marketplace value, and lasting benefits to society.

Silicon Valley
(SRI International headquarters)
333 Ravenswood Avenue
Menlo Park, CA 94025
+1.650.859.2000
education@sri.com

Washington, D.C.
1100 Wilson Boulevard, Suite 2800
Arlington, VA 22209
+1.703.524.2053

www.sri.com/education

STAY CONNECTED