Constructing Assessment Items That Blend Core Ideas and Science Practices

Angela Haydel DeBarger, Christopher J. Harris, Cynthia D'Angelo, SRI International,
333 Ravenswood Ave., Menlo Park, CA 94025
angela.haydel@sri.com, christopher.harris.@sri.com, cynthia.dangelo@sri.com
Joseph Krajcik, Chanyah Dahsah, Jane Lee, CREATE for STEM Institute, Michigan State University,
College of Education, Suite 114, 620 Farm Lane, East Lansing, MI 48824,
krajcik@msu.edu, chan_yah@yahoo.com, jane.jiyoung@gmail.com
Yves Beauvineau, Culturally Responsive Science Pedagogies,
984 Pine Glade Rd, Nederland, CO 80466, beauvineau.yves@gmail.com

Abstract: How do we measure knowledge in use? In this paper we describe how principles of Evidence-Centered Design can be used to construct assessments that meet the ambitious call by the science education community of integrating disciplinary core ideas with practices. In our design process, we first elaborate on, or "unpack", the assessable components of disciplinary core ideas and science practices. Then, we use these elaborations to specify evidence to meet knowledge-in-use learning goals and to define task features to elicit the desired evidence. Our final step entails using a template derived from elaborating learning goals, specifying evidence, and defining task features to construct items that measure science proficiency. We present our design approach, provide examples, and consider implications of our overall work.

Introduction

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 on 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 recently released NRC report on developing assessments to measure science proficiency (NRC, 2013). Central to this view is that disciplinary content and practice should be integrated so that as students apply knowledge in the context of science practice, they deepen their conceptual understanding of content as well as their understanding of how to do science. In this paper, we describe our systematic and scalable approach for designing assessment items that measure student proficiency of new science learning goals that blend disciplinary core ideas with practices. Drawing on prior research (e.g., DeBarger, Krajcik, & Harris, 2014), we present our design approach, provide examples, and consider implications of our overall work.

Theoretical Framework

The prior generation of U.S. science standards (e.g., NRC, 1996, 2000) treated content and inquiry as fairly separate strands of science learning, and assessments followed suit. In some respects, the form the standards took contributed to the problem: content standards stated what students should know, and inquiry standards stated what they should be able to do. Consequently, assessments separately measured the knowledge and practice components. The shift to integrating science practices with core ideas, as emphasized in new U.S. standards documents, is based upon studies of actual scientific practice and what we currently know about student learning (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 uniting content and practice by emphasizing that rich science learning requires tight coupling of what students know and what they can do. This represents a different way of thinking about science content in that core ideas serve as thinking tools to solve problems, make decisions, and explain phenomena (NRC, 2012). It also signifies that measuring proficiency solely as acquisition of core content knowledge is no longer sufficient.

This integrated, knowledge-in-use perspective poses challenges for assessment design. At this time, there are very few examples of assessments that integrate core concepts and science practices in a manner indicative of a knowledge-in-use perspective. There is tremendous need to do this assessment design work well, as assessment will play a central role in supporting implementation of new directions in science education both in the U.S. and internationally. Our approach to taking on this challenge is to use the tenets of Evidence-Centered Design (ECD; e.g., Almond, Steinberg, & Mislevy, 2002), which have been used to guide assessment design and validation. ECD is an approach that has been used in a wide range of assessment design contexts, from the development of large scale, high stakes assessments to the development of classroom-based assessments and other types of proximal or close measurement instruments. ECD emphasizes the evidentiary base for specifying coherent, logical relationships among the (1) learning goals that comprise the constructs to be measured; (2) evidence in the form of observations, behaviours, or performances that should reveal the target constructs; and (3) features of tasks or situations that should elicit those behaviours or performances.

Our Evidence-Centered Design (ECD) Approach

Underlying ECD is an overarching conception of assessment as an argument from imperfect evidence. Messick (1994, p. 16) lays out the basic narrative, saying that we "would begin by asking what complex of knowledge, skills, or other attributes should be assessed, presumably because they are tied to explicit or implicit objectives of instruction or are otherwise valued by society. Next, what behaviors or performances should reveal those constructs, and what tasks or situations should elicit those behaviors?" An enabling framework for organizing ECD is a set of multiple "layers" that distinguish various activities and structures as part of the assessment enterprise. These layers serve to instantiate an assessment argument in explicit, operational processes (Mislevy & Riconscente, 2006).

Below we highlight our efforts for two layers of ECD: (1) *domain analysis*, which involves unpacking science practices and disciplinary core ideas to identify assessable knowledge and skills; and (2) *domain modeling*, specifying a knowledge-in-use assessment argument to articulate linkages among the claims we want to make about what students can do, evidence to demonstrate proficiency, and task features to elicit the desired evidence. To illustrate these layers, we draw examples from the Framework for K-12 Science Education (NRC, 2012), although this approach is applicable to developing assessments for any knowledge-in-use learning goals.

Domain Analysis – Unpacking Science Practices

Unpacking the science practices involves reviewing research on science teaching and learning to develop clarity about the essential grade-band appropriate skills for each practice. We articulate the claims we want to make about students' skills associated with the practice, specify the evidence required to demonstrate the 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 a practice. Below, in Tables 1 and 2, we provide examples of unpacking the practices of constructing explanations and modeling.

Table 1: Unpacking the Science Practice of Constructing Explanations.

Claims	• Ability to construct a scientific explanation based on valid and reliable evidence obtained from multiple sources (including the students' own experiments) consistent with scientific knowledge, principles, and theories • Ability to apply scientific reasoning to show why the data or evidence is adequate for the explanations or conclusions
Evidence Required to Demonstrate Practice	To show this practice, the student will need to analyze and interpret data with regard to a scientific question and respond in a written or an oral format. The explanation framework includes three components: claim, evidence and reasoning. • The claim is a testable statement or conclusion that answers the original question. • Evidence is scientific data that supports the student's claim. At least two pieces of evidence are required. Evidence can come from a range of data sources, such as an investigation conducted by the student or someone else, direct observation, science text, archived data, or some other sources of information. The evidence needs to be both *appropriate* and *sufficient* to support the claim. • Reasoning is a justification that shows why the data counts as evidence to support the claim and includes appropriate scientific knowledge, principles, or theories.
Prior Knowledge	• Ability to use evidence in constructing explanations that specify variables that describe and predict phenomena • Ability to construct an explanation of observed relationships • Ability to identify the evidence that supports particular points in an explanation
Student Challenges	• Students often have difficulty supporting their scientific claims and providing reasoning for why they chose the evidence (McNeill, Lizotte, Krajcik, & Marx, 2006). • Students have difficulty in differentiating between appropriate and inappropriate evidence (McNeill & Krajcik, 2011).

Table 2: Unpacking the Science Practice of Modeling.

Claims	• Ability to develop or modify a model (based on evidence) to account for what happens if a variable or component of a system is changed • Ability to develop and/or revise a model to show the relationships among variables, including those that are not observable but that predict observable phenomena • Ability to develop and/or use a model to predict and/or describe phenomena
Evidence Required to Demonstrate Practice	Students must demonstrate in their models: the components of the model, relationships among the components, and how their model helps to explain or make predictions about phenomena. • Component: Models include specific variables or factors within the system under study. • Relationship: Models need to represent the relationship among components in order to provide an account of why the phenomenon occurs. • Connection: Models need to be connected to causal phenomena or scientific theory that students are expected to explain or predict.
Prior Knowledge	• Ability to develop and revise simple models (e.g., analogies, examples) and use these models to represent events and design solutions
Student Challenges	• Students often have difficulty making connections between different types of representational displays including graphs, tables, charts, data sets, and real-world events (Friel and Bright, 1996). • The predictive, interpretive, and analytic aspects of models often are ignored (Ost, 1987). • Students tend to view models primarily as physical copies of phenomena rather than as tools in the service of theory construction and testing (Grosslight et al., 1991). • Students struggle to coordinate their understanding of scientific phenomena and representations of those phenomena (Rappoport & Ashkenazi, 2008). • Students have difficulty explaining how models, including diagrams and illustrations, can be used to explain observed macroscopic phenomena (Stieff, 2011).

Domain Analysis – Unpacking Disciplinary Core Ideas

Disciplinary core ideas help inform a progression of learning, and as such guide the content of curriculum, instruction and assessment. Unpacking disciplinary core ideas entails thoughtful consideration of ideas in relation to students' grade level, or expected level of expertise. For example, when focusing on an aspect of a disciplinary core idea, such as Structure and Properties of Matter, at the middle school level, we unpack it by:

- *elaborating the meaning of key terms* (e.g., *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 understandings for the targeted student level* (e.g., At the middle school level, students should learn that each pure substance has characteristic physical and chemical properties, for any bulk quantity under given conditions, that can be used to identify it.);
- discussing boundaries (e.g., 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.);
- *identifying background knowledge* that is expected of students to develop a grade-level appropriate understanding of a disciplinary core idea (e.g., Students should have learned how to make observations and measurements to identify materials based on their properties.); and
- *considering research-based problematic student ideas* (e.g., The total mass decreases during chemical reaction when a gas is produced).

Domain Modeling – Specifying a Knowledge-in-Use Assessment Argument

Leveraging the unpacking of science practices and disciplinary core ideas described above, we then move toward specifying a knowledge-in-use assessment argument. In this step, we consider relationships about 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 through engagement in a science practice. In Table 3 we present a knowledge-in-use assessment argument for integrating disciplinary content knowledge about structure and properties of matter with scientific explanation.

Table 3: Knowledge-in-Use Assessment Argument.

Claim	• Ability to construct an explanation in which substances are identified based upon characteristic properties
Additional Knowledge, Skills and Abilities	• Knowledge that some properties can be used to identify substances. These properties are called characteristic properties (e.g., density, melting point, boiling point). • Knowledge that temperature, volume, and mass cannot be used to identify substances and are not characteristic properties • Ability to identify which data can be used as valid and appropriate evidence • Knowledge that a scientific explanation includes a claim, evidence, and reasoning
Evidence Required to Demonstrate Claim	• Claim: Statement that substances (e.g., Liquid A and B) are the same/different. • Evidence: Identification of at least two properties to support claim. • Reasoning: Statement that the same substance must have the same set of characteristic properties or that different substances have different characteristic properties.
Characteristic Task Features	• Assessment is limited to analysis of the following characteristic properties: density, melting point, boiling point, solubility, flammability, and odor. • The term “substance” means a pure material (not a mixture of substances). • Some properties do change with changing conditions (e.g., changing atmospheric pressure affects boiling point). All assessment items will make comparisons between substances where it is clear that the conditions, such as temperature and pressure, are constant. • Tasks provide data about characteristic properties of substances. • Tasks provide a motivating context.
Variable Task Features	• Types of properties included as data/evidence • State of matter of substances (i.e., solid, liquid, or gas state) • Inclusion of irrelevant data (e.g., non-characteristic properties) • Level of scaffolding to develop claim, evidence, and reasoning

Task Example

The following example illustrates a task aligned with the knowledge-in-use assessment argument presented above in Table 3. Immediately below the task is a description of the elements required in a complete response.

Steven found four different bottles filled with unknown pure liquids. He measured the properties of each liquid. The measurements are displayed in the data table below.

Liquid	Density	Color	Volume	Boiling Point
1	1.0 g/cm^3	Clear	6.1 cm^3	100 C$^\circ$
2	0.89 g/cm^3	Clear	6.1 cm^3	211 C$^\circ$
3	0.92 g/cm^3	Clear	10.2 cm^3	298 C$^\circ$
4	0.89 g/cm^3	Clear	10.2 cm^3	211 C$^\circ$

Steven wonders if any of the liquids are the same substance. Use the data in the table to:
1) Write a claim stating whether any of the liquids are the same substance.
2) Provide at least two pieces of evidence to support your claim.
3) Provide reason(s) that justify why the evidence supports your claim.

For full credit:

- Claim explicitly states that Liquid 2 and 4 are the same substance.
- Evidence includes at least two of the following pieces of evidence: density, boiling point, or color of Liquid 2 and 4 are the same.
- Reasoning indicates that density, boiling point, and color are characteristic properties; same substances have the same set of characteristic properties; and Liquid 2 and 4 have the same set of characteristic properties, so they are the same substance.

Implications and Conclusions

Developing a coherent and consistent approach to science education depends upon having high-quality, aligned assessments of student learning. At this time, a critical need exists for research and development of high-quality assessments that align with knowledge-in-use learning goals and that are instructionally suitable for teachers’ use in classrooms. Having exemplary knowledge-in-use assessments will be important to multiple stakeholders. Assessment researchers want to better understand the design principles and psychometric properties of assessments that integrate core ideas and science practices. Learning sciences researchers want to use the assessments to better understand larger issues that wide-spread adoption of a knowledge-in-use perspective would entail, including developing and evaluating new curricula. Science educators and policy makers want assessments that help them to better understand students' knowledge and abilities and also to inform changes in classroom instruction. As a comprehensive assessment design methodology, ECD is uniquely suited to assuring that necessary new assessments accurately measure the *integration* of core ideas with science practices in a coherent and consistent manner.

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

Bell, P. & Linn, M. (2000). Scientific arguments as learning artifacts: Designing for learning from the Web with KIE. *International Journal of Science Education*, 22 (8), 797-817.

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.

Friel, S. N., & Bright, G. W. (1996). *Building a theory of graphicacy: How do students read graphs?* Paper presented at the annual meeting of the American Educational Research Association, New York, NY.

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.

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

McNeill, K. L. & Krajcik, J. (2011). *Supporting grade 5-8 students in constructing explanations in science: The claim, evidence and reasoning framework for talk and writing.* New York, NY: Pearson Allyn & Bacon.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, *23*(2), 13–23.

Mislevy, R. J., & Riconscente, M. M. (2006). Evidence-centered assessment design: Layers, structures, and terminology. In S. Downing & T. Haladyna (Eds.), *Handbook of test development* (pp. 61–90). Mahwah, NJ: Erlbaum.

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). *Taking 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.

National Research Council (2013). *Developing assessments for the Next Generation Science Standards*. Washington, DC: National Academies Press.

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

Ost, D. H. (1987). Models, modeling and the teaching of science and mathematics. *School Science and Mathematics*, *87*(5), 363-370.

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

Rappoport, L. T., & Ashkenazi, G. (2008). Connecting levels of representation: Emergent versus submergent perspective. *International Journal of Science Education*, *30*(12), 1585-1603.

Stieff, M. (2011). Improving representational competence using molecular simulations embedded in inquiry activities. *Journal of research in science teaching*, *48*(10), 1137-1158.

Acknowledgments

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.