What Can Be Learned From Current Large-Scale Assessment Programs to Inform Assessment of the Next Generation Science Standards?

Alicia C. Alonzo

September 2013

K-12 Center at ETS

Invitational Research Symposium on Science Assessment

What Can Be Learned From Current Large-Scale Assessment Programs to Inform Assessment of the Next Generation Science Standards?

Alicia C. Alonzo
Michigan State University

With the publication of *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas* (National Research Council [NRC], 2012), followed by the *Next Generation Science Standards* (NGSS; NGSS Lead States, 2013), we are poised to make a significant change in the way that students learn science in K-12 schools. "Seldom has such a consistent message been sent as to the need for change in what we expect students to know and be able to do in science" (Pellegrino, 2013, p. 320). While the vision provided by the framework and new standards has the potential to make students' science learning more coherent and more consistent with the discipline of science, "the vision cannot be realized unless the standards permeate the education system and guide curriculum, instruction, teacher preparation and professional development, and student assessment" (NRC, 2012, p. 241). As states (e.g., Michigan Department of Education, 2013; Washington's Regional Science Coordinators, 2013; Vermont State Board of Education, 2013) and organizations (e.g., National Science Teachers Association, 2013) gear up to prepare teachers to teach to the new standards, it may be easy to put off work on assessment as part of a second wave of reform, with the rationale that students must be given the opportunity to learn the new standards before they (and their teachers) are held accountable for achievement relative to those standards.

However, this delay is problematic for at least two reasons. First, when used wisely, assessment can provide valuable information to stakeholders about students' current achievement, thus, making efforts to reform science teaching and learning more targeted to students' current learning needs. Second, policy research has clearly demonstrated the great impact that assessment can have on curriculum (e.g., Au, 2007) and instruction (e.g., Hamilton & Berends, 2006). As recognized in the assessment framework for the National Assessment of Educational Progress (NAEP), an assessment that is low stakes for teachers and students, "assessment frameworks… signal to the public and to teachers the elements of a subject that are important" (2008, p. 9). This effect is likely to be even more dramatic for assessments that have high-stakes consequences for students, teachers, and/or schools. Even the best innovations in curriculum and instruction may be subverted if assessments do not fully reflect the type of science learning recommended in the NRC Framework and NGSS.

> Assessment is a key element in the process of educational change and improvement… Done poorly, it sends the wrong signals and skews teaching and learning. Our greatest danger may be a rush to turn the NGSS into sets of assessment tasks for use on high-stakes state accountability tests before we have adequately engaged in research, development, and validation of the range of tasks and tools needed to get the job done properly. (Pellegrino, 2013, p. 323)

To ensure that high-quality assessments will be available to support and measure progress towards the NGSS, we cannot delay work to develop these assessments. Therefore, as new curricula and models of teacher preparation and professional development are created to support implementation of the NGSS, it is vital that we also consider how to assess student learning relative to the new standards.

The framework and associated standards represent several departures from the status quo, which can be summarized by three features:

> First, it is built on the notion of learning as a developmental progression… Second, the framework focuses on a limited number of core ideas in science and engineering, both within and across the disciplines… Third, the framework emphasizes that learning about science and engineering involves integration of the knowledge of scientific explanations (i.e., content knowledge) and the practices needed to engage in scientific inquiry and engineering design. (NRC, 2012, pp. 10–11)

The first feature requires viewing student learning along a progression spanning multiple grade levels.

This is an unfamiliar idea in the realm of science assessments, which have more often been viewed as simply measuring whether students know particular grade-level content. It means that assessments must strive to be sensitive both to grade-level appropriate performances to intermediate performances that may be appropriate at somewhat lower or higher grade levels (Pellegrino, 2013, p. 320).

This challenge is among those discussed in a recent chapter that explores what would be required to incorporate a learning progression[1] approach into large-scale science assessments (Alonzo, Neidorf, & Anderson, 2012).

The second feature also has implications for assessment. Focus on a smaller number of core ideas presents the opportunity to assess students' understanding in more depth. The same assessment time could be used to elicit detailed information about students' understanding of a limited number of ideas, rather than cursory information about their understanding of many ideas. In order to fully represent the content in their frameworks, assessments have typically emphasized breadth over depth;

[1] There are some important differences between *learning progressions* as defined in this chapter and the progressions defined in the NRC Framework and NGSS, such that not all aspects of the chapter are relevant to assessing the NGSS; however, the basic challenges entailed in a progression approach are explored in this chapter.

thus, a challenge may arise in assessing particular ideas more deeply. “The assessment of the NGSS should be on understanding the full Disciplinary Core Ideas—not just the pieces” (NGSS Lead States, 2013, Introduction, p. 6). Thus, some attention may need to be paid to covering fully the depth of ideas represented in the NGSS; however, this work has a clear foundation in earlier attempts to assess student achievement in science.

The third feature encompasses two critical elements. First,

> [the] *Framework* specifies that each performance expectation must combine a relevant practice of science or engineering with a core disciplinary idea and crosscutting concept… That guideline is perhaps the most significant way in which the NGSS differs from prior standards documents. In the future, science assessments will not assess students’ understanding of core ideas separately from their abilities to use the practices of science and engineering. (NGSS Lead States, 2013, Appendix F, p. 1)

Although tension between emphases on science knowledge and practice has been part of debates about science education for decades (NRC, 2012, p. 41), the NGSS represent one of the first attempts to write standards that integrate scientific knowledge and practices.[2] Although other science assessment frameworks—such as those for NAEP and Trends in International Mathematics and Science Study (TIMSS)—call for students’ knowledge to be assessed through the performance of various practices, these assessments report student achievement relative to content standards, rather than standards that integrate content and practices. In other words, in many large-scale assessment programs, practices serve as the means through which students demonstrate knowledge of content, rather than being inextricably linked to content. What is different about the view of science achievement entailed in the NGSS is that the standards themselves—rather than being defined in terms of content ideas—are defined in terms of performances, which represent both content and practices. The formulation of standards in the NGSS presents significant assessment challenges. “A disjuncture exists between students’ knowledge of science facts and procedures, as assessed by typical achievement tests, and their understanding of how that knowledge can be applied through the practices of scientific reasoning, argumentation, and inquiry” (Pellegrino, 2013, p. 320).

Second, the NRC Framework and NGSS include not only scientific practices, which have a relatively long history in the assessment of student achievement, but also practices of engineering design, which “have not received the same level of attention in science curricula, assessments, or the education of new science teachers as the traditional science disciplines have” (NGSS Lead States, 2013, Appendix A, p. 4). A significant difference between the NGSS and other science standards is the stance that “engineering design is as much a part of learning science as engagement in the practices of science” (NRC, 2012, p. 12). The intent is to “[raise] engineering design to the same level as scientific inquiry in

---

[2] As explored in more detail below, the new frameworks for *AP*® examinations in biology, physics, and chemistry also take this approach.

science classroom instruction at all levels" (NGSS Lead States, 2013, Executive Summary, p. 1). However, the NAEP Technology and Engineering Literacy (TEL) Framework (National Assessment Governing Board [NAGB], 2010) notes that "assessing technology and engineering literacy" is "under-developed" and, thus, that "there are few existing sample tasks to serve as examples for assessment development" (NAGB, 2010, p. xv). Although NAEP TEL items should be available to inform development of NGSS assessments, this challenge is likely to exist for the near future (for both groups of assessment developers).

Thus, the crossing of content and practices and the inclusion of engineering design present significant assessment challenges that must be addressed in tandem with work to develop curricula and teacher professional development in order to ensure that the NGSS have the intended impact on K-12 science education.

> Most science assessments, whether intended for classroom or large-scale use, still employ paper-and-pencil presentation and responses formats that are amenable to only limited forms of problem types… Assessments of this type can measure some kinds of conceptual knowledge, and they also can provide a snapshot of some science practices. But they do not adequately measure other kinds of achievement… High-quality science assessments that are consistent with the framework… must target the full range of knowledge and practices described in this report. (NRC, 2012, pp. 262–263)

Although there is much to be learned before we can fully assess the NGSS, we are not starting from scratch. Significant advances have been made in the assessment of students' science achievement, particularly with respect to the relationship between content and practices. Therefore, the purpose of this paper is to explore what might be learned from innovations in current large-scale assessment programs that could inform efforts to assess the NGSS. The paper is divided into four main sections. In the first, I provide an overview of the 16 science and engineering practices described in the NRC Framework and used to generate performance expectations in the NGSS. In the second, I examine the assessment frameworks for five large-scale assessment programs in order to identify where fruitful examples might be found to inform assessment of the NGSS. In the third, I analyze illustrative examples from these assessments to explore how innovative item types are being used to operationalize selected practices that are related to those in the NGSS. In the final section, I provide a summary of what might be learned from existing large-scale assessments and propose two directions for further work towards developing rich, informative assessments of students' learning relative to the NGSS.

Practices in the NGSS

In the NRC Framework, “learning is defined as the combination of both knowledge and practice, not separate content and process learning goals” (NRC, 2012, p. 254), and the Executive Summary of the NGSS (NGSS Lead States, 2013) states that “practices alone are activities and content alone is memorization” (p. 2). As noted above, this is one of the most significant differences between the NGSS and earlier standards, such as the *National Science Education Standards* (NRC, 1996). In addition, while earlier standards and assessment frameworks refer to “process skills,” the NRC Framework reflects a deliberate decision to “use the term ‘practices’ instead of a term such as ‘skills’ to emphasize that engaging in scientific investigation requires not only skill but also knowledge that is specific to each practice” (NRC, 2012, p. 30). Thus, students’ enactment of the practices in the NGSS involves not only “doing” the practice, but also drawing upon content knowledge and understanding of the practice itself.

The NRC Framework identifies a set of eight science practices “that scientists employ as they investigate and build models and theories about the world” and eight engineering design practices “that engineers use as they design and build systems” (NRC, 2012, p. 30). Articulation of these practices is viewed as a key advance of the NRC Framework and NGSS for three reasons: (a) “[minimizing] the tendency to reduce scientific practice to a single set of procedures,” which “overemphasizes experimental investigation at the expense of other practices, such as modeling, critique, and communication” and avoiding procedures being taught as an end in and of themselves; (b) “[avoiding] the mistaken impression that there is one distinctive approach common to all science”; and (c) overcoming the obstacles to “[developing] the idea that science should be taught through a process of inquiry” posed by “the lack of a commonly accepted definition of [the] constituent elements of inquiry” (NRC, 2012, pp. 43–44).

Although the 16 practices are described in some detail in the NRC Framework and NGSS, descriptions of each practice—resulting from a synthesis of information about the practices from multiple places across the two documents—are presented in the text below and in Table 1. This synthesis permits comparisons with large-scale assessment frameworks in the next section.

The science and engineering practices are paired in the NRC Framework, sometimes with the same description (e.g., Practice 7: Engaging in Argument from Evidence) and sometimes with descriptions reflecting the different nature of work in science and engineering, for example, Science Practice 6 (Constructing Explanations) and Engineering Practice 6 (Designing Solutions). There is significant overlap between the science and engineering practices in a given pair; however, they are detailed separately below because work of scientists and of engineers entails different goals and, thus, assessment of these practices are likely to involve different tasks. Given that the goal of this paper is to consider where we might look to learn about how to assess the NGSS and that engineering practices are not as widely assessed, this approach ensures that both sets of practices are considered throughout the paper.

Table 1. Components of the NGSS Science and Engineering Practices and Their Inclusion in Large-Scale Assessment Frameworks

Science and engineering practices	NAEP Science	NAEP TEL	TIMSS Science	PISA Science	AP Science
Science Practice 1: Asking Questions					
Formulate empirically answerable questions			SR		SR
Evaluate questions for testability, relevance, and/or whether they are scientific or not				SR	SR
Ask probing questions about others' scientific work					
Engineering Practice 1: Defining Problems					
Ask probing questions that seek to refine a problem, including criteria and constraints		PR			
Science Practice 2: Developing and Using Models					
Construct and use models/simulations to help develop questions					
Construct and use models/simulations to help develop and/or test explanations	SR			SR	SR
Construct and use models to represent current understanding			SR		
Construct and use models to communicate ideas					SR
Move flexibly between model types					SR
Recognize limitations of models/simulations				SR	
Evaluate limitations of models/simulations				PR	
Refine models/simulations					SR
Engineering Practice 2: Developing and Using Models					
Use models/simulations to analyze existing and proposed systems to identify strengths and weaknesses of designs		PR			
Use models/simulations to test possible solutions		SR			
Recognize limitations of models		SR			
Refine models		PR			
Science Practice 3: Planning and Carrying Out Investigations					
Formulate a question that can be investigated (S1)			SR		SR
Frame a hypothesis based on a model or theory			PR	PR	
Identify relevant variables	PR		PR	SR	PR
Consider how variables might be observed and/or measured	PR		SR	PR	SR
Consider how variables might be controlled	PR		SR	SR	
Consider reliability and precision of data	PR			SR	
Observe and collect data to describe a phenomenon	PR				PR
Observe and collect data to test existing theories and explanations	PR	SR			SR

	Science and engineering practices	NAEP Science	NAEP TEL	TIMSS Science	PISA Science	AP Science
	Observe and collect data to revise and develop new theories and explanations	PR				
	Design plans for investigations individually	SR		SR	SR	SR
	Design plans for investigations collaboratively					
	Evaluate plans for investigations	SR		PR	SR	
	Revise plans for investigations					
Engineering Practice 3: Planning and Carrying Out Investigations						
	Identify relevant variables					
	Consider how variables might be measured					
	Gathering data to specify design criteria					
	Gathering data to test designs		SR			
	Gathering data to fix or improve the functioning of a technological system					
	Gathering data to compare different solutions					
	Design plans for investigations individually		PR			
	Design plans for investigations collaboratively					
	Evaluate plans for investigations					
	Revise plans for investigations					
Science Practice 4: Analyzing and Interpreting Data						
	Use tabulation to collate, summarize, and display data		PR	PR	SR	
	Use graphs to collate, summarize, and display data		PR	PR	SR	
	Use visualization to collate, summarize, and display data		PR	PR	SR	
	Use statistical analysis to collate, summarize, and display data		PR	PR	SR	
	Identify the significant features and patterns in data	SR	PR	SR	SR	SR
	Use data as evidence				PR	
	Distinguish between causal and correlational relationships				PR	
	Identify sources of error					PR
	Calculate degree of uncertainty				SR	PR
Engineering Practice 4: Analyzing and Interpreting Data						
	Identify patterns and interpret results to compare different solutions		PR			
Science Practice 5: Using Mathematics and Computational Thinking						
	Visually represent data (S4)		PR	PR	SR	
	Transform data between tabular and graphical forms				SR	
	Statistically analyze data (S4)		PR	PR	SR	
	Assess the significance of patterns in data					

Science and engineering practices		NAEP Science	NAEP TEL	TIMSS Science	PISA Science	AP Science
	Recognize quantitative relationships	PR		SR		PR
	Express quantitative relationships	PR		SR	PR	PR
	Apply quantitative relationships and mathematical concepts	PR		SR	SR	SR
	Recognize dimensional quantities and use appropriate units	PR		PR	PR	PR
	Use approximation to determine whether quantitative results make sense					SR
Engineering Practice 5: Using Mathematics and Computational Thinking						
	Use mathematical relationships, concepts, and/or processes as part of the design process		PR			
	Use mathematical relationships, concepts, and/or processes to describe designs					
	Use mathematical concepts to test and compare proposed solutions					
	Use mathematical relationships, concepts, and/or processes to support solutions					
Science Practice 6: Constructing Explanations						
	Apply standard explanations of phenomena	SR		SR	SR	SR
	Incorporate current understanding of science into explanations	SR		SR	SR	SR
	Construct explanations of phenomena consistent with evidence	SR	SR	SR	SR	SR
	Link evidence to claims	SR	SR	SR	SR	SR
	Use evidence to support or refute an explanatory account	SR		SR	SR	SR
	Identify gaps or weaknesses in explanatory accounts			PR		PR
Engineering Practice 6: Designing Solutions						
	Solve engineering problems	SR	SR	SR		
	Balance competing priorities	SR	SR	SR		
	Test a design (E3)		SR			
	Evaluate and critique competing design solutions	SR	SR	PR		
	Refine design ideas		SR			
	Optimize a design		SR			
	Select among alternative design features		SR			
	Consider possible unanticipated effects		SR	PR		
Science Practice 7: Engaging in Argument From Evidence						
	Engage in reasoning and argument to identify strengths and weaknesses in a line of reasoning about the best experimental design (S3)		SR			
	Engage in reasoning and argument to identify strengths and weaknesses in a line of reasoning about the most appropriate techniques of data analysis (S4)					
	Engage in reasoning and argument to identify strengths and weaknesses in a					

Science and engineering practices		NAEP Science	NAEP TEL	TIMSS Science	PISA Science	AP Science
	line of reasoning about the best interpretation of a given data set (S4)					
	Formulate evidence based on data (S4)				PR	
	Engage in reasoning and argument to identify strengths and weaknesses in a line of reasoning about how data support a claim (S6)	SR	SR	SR	SR	SR
	Engage in reasoning and argument to find the best explanation for natural phenomena individually (S6)	SR	PR	SR	PR	SR
	Engage in reasoning and argument to find the best explanation for natural phenomena collaboratively		PR			
	Analyze arguments to determine whether they emphasize similar or different evidence and/or interpretations					
	Provide critiques of others' scientific work		SR		SR	
	Identify flaws in one's own arguments					
	Modify one's own scientific work in light of evidence					
	Modify one's own scientific work in response to critiques		SR			
	Understand how claims are judged by the scientific community	PR	PR		SR	SR
	Identify strengths and weaknesses in media reports of science	SR			SR	PR
Engineering Practice 7: Engaging in Argument From Evidence						
	Engage in reasoning and argument to find the best possible solution individually	SR	SR	SR		
	Engage in reasoning and argument to find the best possible solution collaboratively					
	Consider a range of factors to find the best solution (E7)	PR	SR	SR		
	Make arguments from evidence to defend conclusions					
	Compare and evaluate competing ideas	SR	SR	SR		
	Compare and evaluate competing methods			SR		
	Formulate evidence based on test data					
	Revise designs		SR			
	Make an argument about performance of a technology based on empirical evidence		PR			
	Make an argument about the strengths and weaknesses of a technology as reported in the media		PR			
Science Practice 8: Obtaining, Evaluating, and Communicating Information						
	Communicate ideas orally					
	Communicate ideas in writing	SR	PR	PR	SR	PR
	Communicate ideas with the use of tables, diagrams, graphs, and equations	SR	SR		SR	PR

Science and engineering practices	NAEP Science	NAEP TEL	TIMSS Science	PISA Science	AP Science
Communicate ideas through extended discussions with peers					
Derive meaning from scientific papers		PR	PR	PR	
Derive meaning from scientific texts from the Internet	SR	PR	PR	PR	
Derive meaning from scientific information presented orally					
Identify flaws in reports about science in the press or on the Internet (S7)	SR	PR		SR	
Evaluate the validity of scientific information				SR	
Assess the credibility of sources of scientific information		PR		PR	
Assess the accuracy of sources of scientific information		PR		PR	
Assess possible bias in sources of scientific information		PR		PR	
Integrate information from multiple sources	PR	SR			
Engineering Practice 8: Obtaining, Evaluating, and Communicating Information					
Communicate advantages of designs orally		PR			
Communicate advantages of designs in writing		SR			
Communicate advantages of designs with the use of tables, diagrams, graphs, and equations		SR			
Communicate advantages of designs through extended discussions with peers					
Derive meaning from engineering texts		PR			
Evaluate information relevant to the design process		SR			
Apply information meaningfully to the design process		PR			

Note: In general, where overlap existed between the components of different practices, the component was listed with the practice for which it is more central. However, sometimes a component was central to two practices. These cases are indicated in the latter practice, using (SX) or (EY)—where X indicates the science practice number and Y indicates the engineering practice number. NAEP = National Assessment of Educational Progress; TEL = Technology and Engineering Literacy; TIMSS = Trends in International Mathematics and Science Study; PISA = Programme for International Student Assessment; AP = Advanced Placement; SR = represents substantial representation; PR = represents partial representation (either because only part of the component is mentioned in the framework or because the framework does not contain enough detail to ensure that substantial representation is intended).

Science Practice 1: Asking Questions

This practice entails "formulating empirically answerable questions about phenomena, establishing what is already known, and determining what questions have yet to be satisfactorily answered" (NRC, 2012, p. 50). Such questions may arise from "careful observation of phenomena," "examining models or a theory," or attempting "to determine relationships" (NGSS Lead States, 2013, Appendix F, p. 4). *Asking Questions* includes "evaluating question[s] to determine if [they are] testable and relevant" (NGSS Lead States, 2013, Appendix F, pp. 4–5) and distinguishing "between questions that can be answered empirically and those that are answerable only in other domains of knowledge or human experience" (NRC, 2012, p. 55). This practice also entails "[asking] probing questions that seek to identify the premises of an argument, request further elaboration… or challenge the interpretation of a data set" (NRC, 2012, p. 55).

Engineering Practice 1: Defining Problems

"While science begins with questions, engineering begins with defining a problem to solve" (NGSS Lead States, 2013, Appendix F, p. 4). This practice involves "[asking] probing questions that seek to… refine… [an] engineering problem" (NRC, 2012, p. 55). In particular, questions to clarify engineering problems are asked "to determine criteria for successful solutions and identify constraints to solve problems about the designed world" (NGSS Lead States, 2013, Appendix F, p. 17). Such criteria and constraints "may include social, technical, and/or environmental considerations" (NGSS Lead States, 2013, Appendix F, p. 5).

Science Practice 2: Developing and Using Models

This practice entails "the construction and use of… models and simulations to help develop explanations about natural phenomena" (NRC, 2012, p. 50). Such models include "diagrams, drawings, physical replicas, mathematical representations, analogies, and computer simulations" (NGSS Lead States, 2013, Appendix F, p. 19), and the practice includes the ability to "represent and explain phenomena with multiple types of models… and move flexibly between model types" (NRC, 2012, p. 58). In this context, models are "explicit representations that are in some ways analogous to the phenomena they represent" (NRC, 2012, p. 56) and allow scientists to make "predictions of the form 'if … then… therefore'… in order to test hypothetical explanations" (NRC, 2012, p. 50), "to better visualize and understand a phenomenon under investigation" (NRC, 2012, p. 56), "to represent their current understanding… of a system…, to aid in the development of questions and explanations, and to communicate ideas to others" (NRC, 2012, p. 57). As models "bring certain features into focus while minimizing or obscuring others" and "contain approximations and assumptions that limit the range of validity of their application and the precision of their predictive power" (NRC, 2012, p. 56), this practice entails recognizing and evaluating the limitation of models, as well as refining models "in light of empirical evidence or criticism to improve its quality and explanatory power" (NRC, 2012, p. 58).

Engineering Practice 2: Developing and Using Models

This practice entails "the use of… models and simulations to analyze existing systems" in order "to see where flaws might occur or to test possible solutions to a new problem..., to recognize the strengths and limitations of their designs" (NRC, 2012, p. 50), "to compare the effectiveness of different design solutions" (NRC, 2012, p. 58), and "to communicate a design's features to others" (NRC, 2012, p. 57). In this context, models are "explicit representations that are in some ways analogous to the phenomena they represent" and "allow… engineers to… develop a possible solution to a design problem" (NRC, 2012, p. 56). As models "bring certain features into focus while minimizing or obscuring others" and "contain approximations and assumptions that limit the range of validity of their application and the precision of their predictive power" (NRC, 2012, p. 56), this practice entails recognizing the limitation of models, as well as refining models to "better reflect a design's specification" (NRC, 2012, p. 58).

Science Practice 3: Planning and Carrying Out Investigations

This practice includes observing and collecting data "to describe a phenomenon" (NGSS Lead States, 2013, Appendix F, p. 7) or "to test existing theories and explanations or to revise and develop new ones" (NRC, 2012, p. 50). These goals require "designing… inquiries that are appropriate to answering the question being asked or testing a hypothesis that has been formed" (NRC, 2012, p. 59). Thus, this practice also includes "formulating a question that can be investigated" and/or "[framing] a hypothesis… based on a model or theory" (NRC, 2012, p. 60) and "identifying the relevant variables and considering how they might be observed, measured, and controlled" (NRC, 2012, p. 59). *Planning and Carrying out Investigations* also involves considerations related to "how much data are needed to produce reliable measurements," as well as "limitations on the precision of data" (NRC, 2012, p. 60). Because data "aren't evidence until used in the process of supporting a claim" (NGSS Lead States, 2013, Appendix F, p. 7), this practice also involves the reasoning that justifies the use of data as evidence and "what counts as data" (NGSS Lead States, 2013, Appendix F, p. 21). Consistent with scientists' work, this practice requires both individual and collaborative planning, as well as the evaluation of and revision to plans for investigations (NGSS Lead States, 2013, Appendix F, p. 7).

Engineering Practice 3: Planning and Carrying Out Investigations

This practice includes gathering data to "[specify] design criteria or parameters and to test their designs" (NRC, 2012, p. 50) in order to "fix or improve the functioning of a technological system or to compare different solutions to see which best solves a problem" (NGSS Lead States, 2013, Appendix F, p. 7). These goals require "[identifying] relevant variables and [deciding] how they will be measured" (NRC, 2012, p. 50). The goal of engineers' investigations is to "identify how effective, efficient, and durable their designs may be under a range of conditions" (NRC, 2012, p. 50). Consistent with scientists' work, this practice requires both individual and collaborative planning, as well as the evaluation of and revision to plans for investigations (NGSS Lead States, 2013, Appendix F, p. 7).

Science Practice 4: Analyzing and Interpreting Data

In order to "derive meaning" from data, this practice entails using "a range of tools—including tabulation, graphical interpretation, visualization, and statistical analysis" (NRC, 2012, p. 51)—"to collate, summarize, and display data" (NRC, 2012, p. 63) and "to identify the significant features and patterns in… data" (NRC, 2012, p. 51). "Because raw data as such have little meaning," this practice is needed for data to "be used as evidence" (NRC, 2012, p. 61). This practice also involves "distinguish[ing] between causal and correlational relationships" (NRC, 2012, p. 63) and considering the quality of the data: identifying sources of error and calculating degree of certainty (NRC, 2012, p. 51) in order to "evaluate the strength of conclusion[s] that can be inferred from [a] data set" (NRC, 2012, p. 63) and/or "to seek to improve precision and accuracy of data with better technological tools and methods" (NGSS Lead States, 2013, Appendix F, p. 9).

Engineering Practice 4: Analyzing and Interpreting Data

This practice entails using "a range of tools to identify the major patterns and interpret the results" in order to "compare different solutions and determine how well each one meets specific design criteria" (NRC, 2012, p. 51). Physical models may be used to "collect data… and analyze the performance under a range of conditions" (NRC, 2012, p. 63). Rather than "rely[ing] on trial and error" (NRC, 2012, p. 62), data serves as evidence to make engineering decisions, including drawing conclusions about different designs and "defin[ing] an optimal operational range" for a proposed design (NGSS Lead States, 2013, Appendix F, p. 9).

Science Practice 5: Using Mathematics and Computational Thinking

This practice entails using mathematics and computation to perform tasks such as "visually representing data" (NRC, 2012, p. 65) and "transforming … data between various tabular and graphical forms" (NRC, 2012, p. 66) "in ways that allow the exploration of patterns" (NRC, 2012, p. 65); "statistically analyzing data"; "assessing the significance of patterns" in data (NRC, 2012, p. 51); and "recognizing, expressing, and applying quantitative relationships" (NRC, 2012, p. 51). Thus, this practice has a "structural function, which allows for logical deduction" (NRC, 2012, p. 64). In addition, because mathematics and computation allow one to "[represent] physical variables and their relationships" (NRC, 2012, p. 51), this practice has a "communicative function, as one of the languages of science," allowing "ideas to be expressed in a precise form" (NRC, 2012, p. 64). Both functions require the "recogni[tion] of dimensional quantities," use of "appropriate units in scientific applications" (NRC, 2012, p. 65), and "apply[ing] mathematical concepts and/or processes (e.g., ratio, rate, percent, basic operations, simple algebra)" (NGSS Lead States, 2013, Appendix F, p. 10) and "express[ing] relationships and quantities in appropriate mathematical or algorithmic forms" (NRC, 2012, p. 65). The "mathematics of probability and of statistically derived inferences" (NRC, 2012, p. 64) and, thus, the development, use, and understanding of simulations are also part of this practice. *Using Mathematical and Computational Thinking* also entails comparison "with what is known about the real world" to see if "mathematical expressions… or simulations… 'make sense'" (NRC, 2012, p. 66). This practice can be enhanced by the use of "computers and digital tools," for example, to "automat[e] calculations, approximat[e] solutions… and analyz[e] large data sets" and to "observ[e], measur[e], [record]," "organize," "search," and "[process] data" (NGSS Lead States, 2013, Appendix F, p. 10).

Engineering Practice 5: Using Mathematics and Computational Thinking

This practice includes using both (a) "established relationships and principles" (NRC, 2013, p. 51) and "mathematical concepts and/or processes (e.g., ratio, rate, percent, basic operations, simple algebra)" (NGSS Lead States, 2013, Appendix F, p. 10) as part of the design process and (b) "simulations of designs" to develop and improve designs (NRC, 2013, p. 51). It also includes "us[ing] mathematical representations to describe and/or support… design solutions," "using digital tools and/or mathematical concepts to test and compare proposed solutions," and "creat[ing] algorithms (a series of ordered steps) to solve a problem" (NGSS Lead States, 2013, Appendix F, p. 10).

Science Practice 6: Constructing Explanations

This practice entails "construct[ing] logically coherent explanations of phenomena that… are consistent with the available evidence" (NRC, 2012, p. 52). Scientific explanations "include a claim that relates how a variable or variables relate to another variable or set of variables," "often… in response to a question" (NGSS Lead States, 2013, Appendix F, p. 11). Scientific explanations can be used to "predict… and/or describe… phenomena" (NGSS Lead States, 2013, Appendix F, p. 11); they "are explicit applications of theory to a specific situation or phenomenon, perhaps with the intermediary of a theory-based model" (NRC, 2012, p. 52). Students' scientific explanations should "incorporate their current understanding of science or a model that represents it" (NRC, 2012, p. 52); they are also "expected to… apply standard explanations" (NGSS Lead States, 2013, Appendix F, p. 11). "Deciding on the best explanation is a matter of argument," based on "how well any given explanation fits with all available data, how much it simplifies what would seem to be complex, and whether it produces a sense of understanding" (NRC, 2012, p. 68). Thus, this practice includes "appl[ying] scientific reasoning to link evidence to … claims to assess the extent to which the reasoning and data support the explanation" (NGSS Lead States, 2013, Appendix F, p. 11), using "scientific evidence and models to support or refute an explanatory account of a phenomenon," and "identif[ying] gaps or weaknesses in explanatory accounts" (NRC, 2012, p. 69).

Engineering Practice 6: Designing Solutions

This practice entails using an iterative and "systematic process for solving engineering problems…. based on scientific knowledge and models of the material world… and balancing competing criteria of desired functions, technological feasibility, cost, safety, aesthetics, and compliance with legal requirements" (NRC, 2012, p. 52). *Designing Solutions* also involves "testing a design"; "evaluat[ing] and critique[ing] competing design solutions"; "refining design ideas based on the performance of a prototype or simulation"; "optimizing [performance] of a design by prioritizing criteria, making trade-offs, testing, revising, and retesting" (NGSS Lead States, 2013, Appendix F, p. 12); and "selecting among alternative design features to optimize the achievement of design criteria" (NRC, 2012, p. 69). Throughout the process of *Designing Solutions*, "possible anticipated effects" must be considered (NGSS Lead States, 2013, Appendix F, p. 12).

Science Practice 7: Engaging in Argument From Evidence

This practice entails "reasoning and argument" to identify "strengths and weaknesses of a line of reasoning" and find "the best explanation for… natural phenomen[a]" (NRC, 2012, p. 52). Argumentation is used "to listen, compare, and evaluate competing ideas and methods based on their merits" (NGSS Lead States, 2013, Appendix F, p. 13). Arguments for and against particular explanations may involve "showing how data support a claim" (NRC, 2012, p. 72). In addition, arguments may be made about "the best experimental design, the most appropriate techniques of data analysis, or the best interpretation of a given data set" (NRC, 2012, p. 71). "Arguments [may] be based on deductions from premises, on inductive generalization of existing patterns, or on inferences about the best possible explanation" (NRC, 2012, p. 71). They may be communicated orally or in writing (NGSS Lead States, 2013, Appendix F, p. 13). As part of this practice, scientists "defend their explanations, formulate evidence based on a solid foundation of data" (NRC, 2012, p. 52), "identify flaws in their own arguments and modify and improve them" (NRC, 2012, p. 73) "in light of the evidence" (NRC, 2012, p. 52) "and in response to criticism" (NRC, 2012, p. 73), and "collaborate with peers in searching for the best explanation for… phenomen[a]" (NRC, 2012, p. 52).Thus, scientists must "respectfully provide and receive critiques about [their] explanations, procedures, models, and questions by citing relevant evidence and posing and responding to questions that elicit pertinent elaboration and detail" (NGSS Lead States, 2013, p. 13).

Another part of this practice, important for both scientists and citizens, is "the knowledge and ability to detect 'bad science'" (NRC, 2012, p. 71), an understanding of "how claims to knowledge are judged by the scientific community," and, thus, the ability to "read media reports of science in a critical manner so as to identify their strengths and weaknesses" (NRC, 2012, p. 73). Doing so requires the ability to "compare and critique arguments and analyze whether they emphasize similar or different evidence and/or interpretations of facts" (NGSS Lead States, 2013, Appendix F, p. 13).

Engineering Practice 7: Engaging in Argument From Evidence

This practice entails reasoning and argument to find "the best possible solution to a problem" (NRC, 2012, p. 52), by "listen[ing], compar[ing], and evaluat[ing] competing ideas and methods" "based on scientific ideas and principles, empirical evidence, and/or logical arguments" (NGSS Lead States, 2013, Appendix F, pp. 13–14). As part of this practice, "engineers collaborate with their peers throughout the design process," particularly in order to "[reach] agreements about" (NGSS Lead States, 2013, Appendix F, p. 13) "the most promising solution among a field of competing ideas" (NRC, 2012, p. 52). They

> use systematic methods to compare alternatives, formulate evidence based on test data, make arguments from evidence to defend their conclusions, evaluate critically the ideas of others, and revise their designs in order to achieve the best solution to [a] problem. (NRC, 2012, p. 52)

As part of this practice, engineers "use cost-benefit analysis, an analysis of risk, an appeal to aesthetics, or predictions about market reception to justify why one design is better than another" (NRC, 2012, p. 72) and consider relevant factors such as "economic, societal, environmental, [and] ethical considerations" (NGSS Lead States, 2013, Appendix F, p. 14). Their arguments may be communicated orally or in writing (NGSS Lead States, 2013, Appendix F, p. 13).

This practice includes the ability to "make an argument that supports or refutes the advertised performance of a device… based on empirical evidence concerning whether or not the technology meets relevant criteria and constraints" (NGSS Lead States, 2013, Appendix D, p. 13) and the ability to "read media reports of technology in a critical manner so as to identify their strengths and weaknesses" (NRC, 2012, p. 73).

Science Practice 8: Obtaining, Evaluating, and Communicating Information

This practice entails "reading, interpreting, and producing text" and has two main components. First, it includes "the communication of ideas and the results of inquiry" through multiple modes, including "orally"; "in writing"; "with the use of tables, diagrams, graphs, and equations"; and through "extended discussions with scientific peers" (NRC, 2012, p. 53). In particular, scientists "describe observations precisely, clarify their thinking, and justify their arguments" (NRC, 2012, p. 74). Second, it includes "deriv[ing] meaning" from a range of "scientific texts," "such as papers, the Internet, symposia, and lectures"; "evaluat[ing] the scientific validity of the information thus acquired," and integrat[ing]… information from multiple sources (NRC, 2012, p. 53). In other words, the practice includes "determining the central ideas" of scientific texts" and offering "simple but still accurate" "paraphrasing" (NGSS Lead States, 2013, p. 15) of these texts

Of particular importance is the use of this ability to be "a critical consumer of science," being able to "view reports about science in the press or on the Internet and to recognize the salient science, [and] identify sources of error and methodological flaws" (NRC, 2012, p. 75) and the ability to "synthesize multiple claims, methods, and/or designs" (NGSS Lead States, 2013, Appendix F, p. 15). This includes "gather[ing]… information from multiple appropriate sources"—including those that seem to offer "competing information or accounts"—"assess[ing] the credibility, accuracy, and possible bias of each publication.., and describ[ing] how they are supported or not supported by the evidence" (NGSS Lead States, 2013, Appendix F, p. 15).

Engineering Practice 8: Obtaining, Evaluating, and Communicating Information

This practice entails "reading, interpreting, and producing text" and has two main components. First, it includes communicating "the advantages of their designs...clearly and persuasively" through multiple modes, including "orally"; "in writing"; "with the use of tables, graphs, drawings, or models"; and through "extended discussions with scientific peers" (NRC, 2012, p. 53). Second, it includes "deriv[ing] meaning" from engineering texts, "evaluat[ing] information, and apply[ing] it usefully" (NRC, 2012, p. 52). These texts may include "handbooks, specific to particular engineering fields, that provide detailed information... on how best to formulate design solutions to commonly encountered engineering tasks" (NRC, 2012, p. 75).

Mapping Large-Scale Assessment Frameworks Onto the NGSS Science and Engineering Practices

In this section, I provide a brief overview of five large-scale assessment programs and show how each of the practices in the NRC Framework is reflected in the five assessment frameworks. Table 1 provides a summary of my judgments concerning the extent to which each large-scale assessment framework contains components of the science practices identified in the NRC Framework and NGSS. It is important to note that this is necessarily a somewhat flawed process, in that the assessment frameworks vary in the level of detail provided about the practices and because identifying the presence of a particular component cannot speak to the depth or quality of its representation in the framework or assessment. Below, I have tried to note places where a particular assessment framework did not explicitly include mention of a particular component that might be inferred from the framework. Although this helps somewhat, Table 1 should not be interpreted as a judgment of the quality of any large-scale assessment program or its overall alignment with the NGSS.

I start with large-scale assessments intended for making national and international comparisons: NAEP, TIMSS, and the Programme for International Student Assessment (PISA). Pellegrino (2013, p. 322) noted that large-scale assessment programs such as NAEP and PISA include "sets of simple and complex science assessment tasks that demand reasoning about science content as described in the NRC Framework and NGSS." All three assessments are designed to provide a picture of achievement across entire countries and/or states (and other local regions).

In contrast, *AP®* examinations, which are discussed next, are taken by a select subset of US high school students (those who seek college credit for advanced coursework). The frameworks for AP biology, chemistry, and physics courses and assessments have recently been redesigned. The structure of these frameworks "parallels that of the core ideas and science practices in the NRC Framework" (Pellegrino, 2013, p. 322), and "performance expectations… within each discipline… reflect the blending of core ideas with science practices" (p. 323).

Overview of Large-Scale Assessments

NAEP Science. "The National Assessment of Educational Progress (NAEP) is the largest nationally representative and continuing assessment of what America's students know and can do in various subject areas" (National Center for Education Statistics [NCES], 2012e). "NAEP results serve as a common metric for all states and selected urban districts" and "provide a clear picture of student academic progress over time" (NCES, 2012e)

> ...for populations of students (e.g., all fourth-graders) and groups within those populations (e.g., female students, Hispanic students). NAEP does not provide scores for individual students or schools… NAEP results are based on representative samples of students at grades 4, 8, and 12 for the main assessments… (NCES, 2012e)
>
> Historically, NAEP has administered a science assessment every 4 to 5 years.
>
> The content of the NAEP science assessment is guided by the NAEP science framework... In 2009, a new framework was introduced that replaced the one used for the 1996, 2000, and 2005 science assessments. The 2009 and 2011[3] assessments were developed using the same framework…. (NCES, 2012a)

The main NAEP science assessment consists of selected and constructed-response items; in addition, "a subset of students sampled… receive an additional 30 minutes to complete hands-on performance or interactive computer tasks" (NAGB, 2008, p. viii).

The NAEP Science framework includes not only a content dimension but also a dimension that is "defined by four science practices: identifying science principles, using science principles, using scientific inquiry, and using technological design" (NAGB, 2008, p. viii). *Identifying Science Principles* and *Using Science Principles* can be considered to be "knowing science" (NAGB, 2008, p. 11), as "both require students to correctly state or recognize science principles contained in the content statements" (NAGB, 2008, p. 69). *Using Scientific Inquiry* and *Using Technological Design* can be considered to be "the application of knowledge to 'doing science' and 'using science to solve real-world problems"

[3] "The 2011 science assessment was administered at grade 8 only so that results from both the NAEP mathematics and science assessments in 2011 could be linked to results from the 2011 Trends in International Mathematics and Science Study (TIMSS)" (NCES, 2012b).

(NAGB, 2008, p. 11). Thus, the latter two practices have the greatest overlap with the NRC Framework. The NAEP Science framework recognized a criticism frequently levied at tasks that purport to measure students' inquiry skills in large-scale assessments: "Rather than tap into students' ability to inquire into a problem, typical performance assessments instead measure students' ability to follow step-by-step instructions to arrive at the expected answer" (NAGB, 2008, p. 106). While acknowledging that "assessment developers are likely to create these 'recipe' types of exercises because they must take into account the vast differences" in students' prior knowledge and experiences (NAGB, 2008, p. 106), the assessment framework calls for NAEP hands-on performance tasks to be "content-rich…, requir[ing] knowledge of science principles to carry them out" (NAGB, 2008, p. 107). In contrast to other science assessments, NAEP Science includes some attention to engineering, through the practice of *Using Technological Design*. However, the framework emphasizes that the assessment focuses on science, such that "technology and technological design are included in the framework but are limited to that which has a direct bearing on the assessment of students' science achievement" (NAGB, 2008, p. 9). Thus, items are "limited to those that reveal students' ability to apply science principles in the context of technological design" (NAGB, 2008, p. 76).

As in the NRC (2012) framework, "performance expectations are derived from the intersection of content statements and science practices" (NAGB, 2008, p. 82). "The NAEP Science assessment… focus[es] on how students bring science content… to bear as they engage in the practices" (NAGB, 2008, p. 82); therefore, on the NAEP science assessment, "neither the content statements… nor the practices… [are] assessed in isolation" (NAGB, 2008, p. 84). However, despite this similarity to the NRC Framework, the NAEP Science assessment framework prioritizes students' knowledge, focusing on "students' conceptual understanding, that is, their knowledge and use of science facts, concepts, principles, laws, and theories" (NAGB, 2008, p. vii). "At all grades, the greatest emphasis [is] on identifying and using science principles" (NAGB, 2008, p. viii); "30 percent of the items address *using scientific inquiry*, and 10 percent of the items address *using technological design*" (NAGB, 2008, p. 95). Indeed, rather than assessing standards that are written to encompass both content and practices, although the NAEP achievement levels are defined in terms of what students know and can do, scores are reported only for overall achievement and for achievement in each content area (NCES, 2012d).

NAEP TEL. In 2014, NAEP will, for the first time, assess students in the area of Technology and Engineering Literacy (TEL), which "involves a range of knowledge and capabilities whose assessment requires having students… [use] diverse tools to solve problems and meet goals within rich, complex scenarios that reflect realistic situations" (NAGB, 2010, p. xiii). The assessment will be totally computer-based because the assessment

> will rely primarily on scenario-based assessment sets that test students through their interaction with multimedia tasks that include conventional item types… and also monitor student actions as they manipulate components of the systems and models that are presented as part of the task. (NAGB, 2010, p. 4-1)

The NAEP TEL framework defines three key areas of "knowledge and skills": "Technology and Society; Design and Systems; and Information and Communication Technology" (NAGB, 2010, p. xii), which are subdivided into subareas as shown in Table 2. Like the NAEP science framework, the NAEP TEL framework describes *practices*, in this case "particular ways of thinking and reasoning when approaching a problem" (NAGB, 2010, p. xii). These practices are: Understanding Technological Principles, Developing Solutions and Achieving Goals, and Communicating and Collaborating (NAGB, 2010, p. xiii). As for Science, the NAEP TEL framework calls for the reporting of "technological content areas" (NAGB, 2010, p. 1-12), rather than assessing standards (such as NGSS) that combine content and practices. In addition, the NAEP TEL framework specifies that tasks should "provide whatever prior knowledge is required to answer the question" (p. 1-16); thus, rather than expecting students to combine scientific knowledge and engineering practices (as in the NRC Framework), the NAEP TEL assessment would provide the scientific knowledge required to respond.

Table 2***. Major Areas and Subareas of 2015 NAEP Technology and Engineering Literacy Assessment***

Technology and society	Design and systems	Information and communication technology (ICT)
A. Interaction of technology and humans B. Effects of technology on the natural world C. Effects of technology on the worlds of information and knowledge D. Ethics, equity, and responsibility	A. Nature of technology B. Engineering design C. Systems thinking D. Maintenance and troubleshooting	A. Construction and exchange of ideas and solutions B. Information research C. Investigation of problems D. Acknowledgement of ideas and information E. Selection and use of digital tools

Note. Adapted from "Technology and Engineering Literacy Framework for the 2014 National Assessment of Educational Progress: Pre-Publication Edition," by National Assessment Governing Board, 2010, p. 2-2, Table 2.1).

TIMSS Science. TIMSS is an international assessment that "provide[s] a comprehensive picture of the mathematics and science achievement of fourth- and eighth-grade students in each participating country" (Mullis, Martin, Ruddock, O'Sullivan, & Preuschoff, 2009, p. 121). As for NAEP, each student participating in TIMSS "is presented with only a sample of the items" (Mullis et al., 2009, p. 121) by using a "matrix-sampling approach" (Mullis et al., 2009, p. 123). The TIMSS assessment includes two item formats: multiple-choice and constructed-response. "At least half of the total number of points represented by all of the questions" (Mullis et al., 2009, p. 127) come from the former, which means that over half of the questions on the assessment are in this format. The TIMSS Science framework acknowledges that these items "do not allow for students' explanations or supporting statements" (Mullis et al., 2009, p. 128).

> The science assessment framework for TIMSS consists of a content dimension specifying the subject matter domains to be assessed within science… and a cognitive dimension specifying the cognitive domains or skills and behaviors (that is, knowing, applying, and reasoning) expected of students as they engage with the science content. (Mullis et al., 2009, p. 49)

TIMSS reports "achievement in each of the content and cognitive domains, as well as overall mathematics and science achievement" (Mullis et al., 2009, p. 121). This approach is similar to that taken in the NAEP frameworks (NAGB, 2008, 2010). Objectives in the TIMSS Science framework are "written in terms of behaviors to be elicited by items that exemplify the understandings and abilities expected of students" (Mullis et al., 2009, p. 63), and, as with the NAEP frameworks and the NGSS, TIMSS Science "takes the position that the understandings and abilities required to engage in [scientific inquiry]" should not be assessed in isolation but rather "in the context of one or other of the TIMSS Science content domains" (Mullis et al., 2009, p. 51). However (as indicated by verbs like *state*, *identify, recognize, describe, relate, compare,* and *demonstrate basic knowledge of*), most of the objectives are ways to demonstrate knowledge of content, rather than engagement in scientific practices as described in the NRC Framework and NGSS. There are some objectives that use the verb *explain*; however, these may also require recall of scientific information, rather than the construction of evidence-based scientific explanations. The TIMSS Science assessment focuses solely on science (and applications of science). There is no explicit attention paid to engineering design in the TIMSS Science framework, although, as noted below, some expectations of the framework do align with engineering practices in the NRC Framework and NGSS.

PISA Science. PISA is an international assessment that measures 15-year-olds' scientific literacy, along with literacy in other areas. The content of the assessment is defined by "a response to the question: *What is important for young people to know, value, and be able to do in situations involving science and technology?*" (Organisation for Economic Co-operation and Development [OECD], 2013b, p. 4). Scientific literacy refers to both "*a knowledge of science"* and a knowledge of "*science-based technology*" (OECD, 2013b, p. 3) and "requires not just knowledge of the concepts and theories of science but also a knowledge of the common procedures and practices associated with scientific enquiry and how these enable science to advance" (OECD, 2013b, pp. 3–4). The PISA Science assessment framework defines the "construct of scientific literacy" "in terms of a set of competencies that a scientifically literate individual would be expected to display" (OECD, 2013b, p. 4): "Explain phenomena scientifically, evaluate and design scientific enquiry, and interpret data and evidence scientifically" (OECD, 2013b, p. 5). The "target distribution of score points" for the competencies is 40% to 50% *explaining*, 20% to 30% *evaluating and designing*, and 30% to 40% *interpreting* (OECD, 2013b, p. 47). The framework notes that "all of these competencies require knowledge" (OECD, 2013b, p. 5). In this sense, the PISA Science assessment framework shares a similarity with the NRC Framework and NGSS, in that competencies or practices are defined to include knowledge underlying their meaningful enactment. The PISA Science assessment framework identifies three types of knowledge: content, procedural, and epistemic , comprising 54% to 66%, 19% to 31%, and 10% to 22% of the science assessment, respectively (OECD, 2013b, p. 46). *Knowledge* and *competencies* are joined by *contexts* and *attitudes* as “four interrelated aspects” that characterize PISA’s definition of scientific literacy (OECD, 2013b, p. 11). All PISA Science items “will require the use and application of the scientific competencies and knowledge within a context” (OECD, 2013b, p. 43). PISA Science incorporates “a range of personal, local, national and global contexts,” a perspective that “differs from that of many school science programmes which are often dominated by content knowledge” because “the framework is based on a broader view of the kind of knowledge of science required by participating members of contemporary society” (OECD, 2013b, p. 6).

The PISA Science assessment is divided into test units, which are “defined by specific stimulus material, which may be a brief written passage”; “text accompanying a table, chart, graph, or diagram”; or “non-static… material, such as animations and interactive simulations” (OECD, 2013b, p. 44). Associated with each stimulus material is “a set of independently scored questions of various types” (OECD, 2013b, p. 44). This structure “facilitate[s] the employment of contexts that are as realistic as possible, reflecting the complexity of real situations, while making efficient use of testing time” (OECD, 2013b, p. 44). Three types of item response formats are present in the PISA Science assessment: simple multiple-choice, complex multiple-choice, and constructed response, “ranging from a phrase to a short paragraph (e.g., two to four sentences of explanation)” or a drawing (OECD, 2013b, p. 45). Each format makes up about one-third of the assessment. As with NAEP and TIMSS, each student takes only a subset of the available assessment items. In 2015, “each student will spend one hour on scientific literacy, with the remaining time assigned” to one of the other tested domains (OECD, 2013b, p. 46). In addition, for the 2015 assessment, “computer-based assessment will be the primary mode of delivery”; however, countries will have the option of administering a paper-based assessment, including only trend items—not those developed for the 2015 assessment (OECD, 2013b, p. 46).

AP Physics. The College Board’s AP program “enables students to pursue college-level studies while still in high school” (College Board, 2013, p. 1). “Each AP course is modeled upon a comparable college course” and “concludes with a college-level assessment developed and scored by college and university faculty, as well as experienced AP teachers” (College Board, 2013, p. 1). “Most four-year colleges and universities in the United States… grant students credit, placement, or both on the basis of successful AP Exam scores” (College Board, 2013, p. 1).

As mentioned above, the College Board’s AP program is in the midst of a curriculum redesign, with AP Science courses, such as biology, chemistry, and physics, being substantially revised to “[shift] away from a traditional ‘content coverage’ model of instruction to one that focuses on the big ideas in an introductory, college-level” course or sequence and “provides students with enduring, conceptual understandings of foundational…principles” (College Board, 2012, p. 1).

Like the NGSS, the new AP Science frameworks emphasize science practices and articulate learning objectives that “provide a clear and detailed articulation of what students should know and be able to do," "integrate science practices with specific content," and "fully define what will be assessed" on the corresponding exams (College Board, 2012, p. 2). The frameworks treat "content, inquiry, and reasoning" as "equally important" (College Board, 2012, p. 1). In these frameworks, a practice is defined as "a way to coordinate knowledge and skills in order to accomplish a goal or task," and each "capture[s] important aspects of the work that scientists engage in, at the level of competence expected of AP… students" (College Board, 2012, p. 2) Each AP Science framework includes the same set of seven science practices, with descriptions tailored to the particular science discipline. For simplicity, in this paper, I will examine only the AP Physics framework as a means of understanding the general alignment between the AP Science frameworks and the NGSS.

How the NGSS Are Reflected in Large-Scale Assessment Frameworks

In considering the alignment between the NGSS and existing large-scale assessment frameworks, it is important to note that, although the practices are delineated separately, "they intentionally overlap and interconnect" (NGSS Lead States, 2013, Appendix F, p. 2); therefore, although the practices have been described separately above, there is overlap in the consideration of their alignment with other assessment frameworks below. This is especially true for practices 6 and 7 (*Constructing explanations*/*Designing solutions* and *Engaging in argument from evidence*).

Science Practice 1: Asking Questions

TIMSS Science. *"*Specific behaviors to be elicited by items that are aligned" (Mullis et al., 2009, p. 81) with the cognitive domain of *Reasoning* include "hypothesize/predict" (Mullis et al., 2009, p. 86). This behavior includes "combining knowledge of science concepts with information from experience or observation to formulate questions that can be answered by investigation" (Mullis et al., 2009, p. 86). "Formulating questions" is also identified as one of five "major aspects of the scientific inquiry process" that are expected in the TIMSS Science assessment (Mullis et al., 2009, p. 88).

PISA Science. In the PISA 2015 framework, the competency *Evaluate and design scientific enquiry* includes "us[ing] a knowledge and understanding of scientific enquiry" to "identify questions that can be answered by scientific enquiry" (OECD, 2013a, p. 5). In particular, this competency requires students to "[demonstrate] the ability to: identify the question explored in a given scientific study" and to "distinguish questions that are possible to investigate scientifically," which relies upon the ability to "discriminate scientific questions from other forms of enquiry" and "to recognise questions that could be investigated scientifically in a given context" (OECD, 2013a, pp. 15-16). *Epistemic knowledge* includes recognizing "what constitutes a scientific or technological question" (OECD, 2013a, p. 21).

AP Science. Science Practice 3 in the AP Physics framework is "The student can engage in scientific questioning to guide investigations within the context of the AP course" and includes "3.1 The student can *pose scientific questions*"; "3.2 The student can *refine scientific questions*"; and "3.3 The student can *evaluate scientific questions*" (College Board, 2012, p. 125).

Summary. This science practice is partially represented in the reviewed large-scale assessment frameworks. The abilities to formulate and to evaluate questions that drive scientific investigations are each addressed in two of the large-scale assessment frameworks. None of the frameworks specifically includes the ability to ask probing questions of others' scientific work, although several do include the ability to critique others' scientific work.

Engineering Practice 2: Defining Problems

NAEP TEL. For this practice, perhaps the most relevant aspect of the NAEP TEL framework is *Engineering Design,* a subarea of *Design and Systems*. As part of the engineering design process, students are expected to "define the problem by identifying criteria and constraints" (D.8.8) and "predicting how these will affect the solution" (D.12.8; NAGB, 2010, p. 2-25). The NAEP TEL framework identifies the following as examples of criteria and constraints: "materials, cost, safety, reliability, performance, maintenance, ease of use, aesthetic considerations…, policies" (NAGB, 2010, p. 2-23), "natural laws," and "available technologies" (NAGB, 2010, p. 2-25). Students are expected to know that criteria "may be weighted in various ways" (D.12.7; NAGB, 2010, p. 2-25). Underlying students' work to define engineering problems, *Engineering Design* includes the key principles that "designing includes identifying and stating the problem, need, or desire" and that "requirements for a design challenge include the criteria for success, or goals to be achieved, and the constraints or limits that cannot be violated in a solution" (NAGB, 2010, p. 2-23). In addition, in the *Technology and Society* area of the TEL framework, the *Interaction of Technology and Humans* subarea includes knowledge that "technological solutions are developed on the basis of criteria and constraints" (NAGB, 2010, p. 2-6), and the *Effects of Technology on the Natural World* subarea includes the ability to "identify a complex global environmental issue" (T.12.7; NAGB, 2010, p. 2-11).

Summary. Although NAEP TEL does include problem definition as a central part of the design process, it does not specify that students should be able to ask probing questions as part of this process. Therefore, this practice does not seem to be well represented in the large-scale assessment frameworks.

Science Practice 2: Developing and Using Models

NAEP Science. One type of NAEP Science interactive computer task (administered to only a subset of students) will ask students to use a simulation that models a system, "manipulat[ing] variables and predict[ing] and explain[ing] resulting changes in the system" (NAGB, 2010, p. ix).

TIMSS Science. *"*Specific behaviors to be elicited by items that are aligned" (Mullis et al., 2009, p. 81) with the *Applying* cognitive domain include "use models" (Mullis et al., 2009, p. 86), which is elaborated as "use a diagram or model to demonstrate understanding of a science concept, structure, relationship, process, or biological or physical system or cycle (e.g., food web, electrical circuit, water cycle, solar system, atomic structure)" (Mullis et al., 2009, p. 83).

PISA Science. The ability to "identify, use and generate explanatory models and representations" is included in the 2015 PISA Science competency *Explain Phenomena Scientifically* (OECD, 2013b, p. 15). The framework notes that "a scientifically literate person should be expected to draw on standard scientific models to construct simple representations to explain everyday phenomena… and use these to make predictions" (OECD, 2013b, p. 15). *Epistemic Knowledge* includes the knowledge that "the construction of models, be they directly representational, abstract or mathematical, is a key feature of science and that such models are akin to maps rather than accurate pictures of the material world" (OECD, 2013b, p. 20). Those with epistemic knowledge understand "the use and role of physical, system and abstract models and their limits" (OECD, 2013b, p. 21).

AP Physics. Science Practice 1 in the AP Physics framework is: "The student can use representations and models to communicate scientific phenomena and solve scientific problems" (College Board, 2012, p. 123). The framework specifies that "models can be both conceptual and mathematical" and that "inherent in the construction of models that physicists invent is the use of representations," including "pictures, motion diagrams, force diagrams, graphs, energy bar charts," "ray diagrams," and "mathematical representations such as equations" (College Board, 2012, p. 123). As part of this practice, students are expected to "*create*" (1.1), "*describe*" (1.2), and "*refine*" (1.3) "representations of natural or man-made phenomena and systems in the domain"; to "*use representations and models* to analyze situations or solve problems qualitatively and quantitatively" (1.4); and to "*re-express key elements* of natural phenomena across multiple representations in the domain" (College Board, 2012, p. 123). In addition, Science Practice 6—"The student can work with scientific explanations and theories" (College Board, 2012, p. 128)—includes "6.4 The student can *make claims and predictions about natural phenomena* based on scientific… models," and Science Practice 7—"The student is able to connect and relate knowledge across various scales, concepts, and representations in and across domains" (College Board, 2012, p. 129)—includes "7.1 The student can *connect phenomena and models* across spatial and temporal scales" (College Board, 2012, p. 130).

Summary. The construction and use of models/simulations to help develop explanations is most consistently included in the reviewed large-scale assessment frameworks. Less attention is paid to other purposes for models/simulations (such as developing questions, representing current understanding, and communicating ideas). There is also relatively little explicit attention to the limitations of models/simulations, and only one assessment framework includes the expectation that students will refine models/simulations.

Engineering Practice 2: Developing and Using Models

NAEP TEL. In the NAEP TEL framework, the *Design and Systems* area contains two subareas with relevance to this practice: *Engineering Design* and *Systems Thinking*. The former includes "making drawings, models, and prototypes," while the latter includes "the ability to use simulations… to predict the behavior of systems" (NAGB, 2010, p. 2-18). For example, students are expected to "simulate tests of various materials to determine which would be best to use for a given application" (D.8.4; NAGB, 2010, p. 2-21) and to "construct and test… models to see if they meet the requirements of a problem" (D.12.9; NAGB, 2010, p. 2-25). The NAEP TEL framework notes that modeling—"try[ing] out [a] solution by constructing a model, prototype…, or simulation and then testing it to see how well it meets the criteria and falls within the constraints"—allows ideas "to be tested before too much time, money, or effort" has been invested (NAGB, 2010, p. 2-22).

In addition, the *Effects of Technology on the World of Information and Knowledge* subarea (within the *Technology and Society* area of the TEL framework) includes the key principle that "the emergence of intelligent information technologies and the development of sophisticated modeling and simulation capabilities are transforming the world of information and knowledge" (NAGB, 2010, p. 2-6). "Key principles" in *Investigation of Problems* (a subset of *Information and Communication Technology*) include "digital models can be used to create simulations and test solutions" and "digital tools can be used to… investigate practical problems" (NAGB, 2010, p. 2-38). Students are expected "to explore authentic issues by building models and conducting simulations in which they vary certain quantities to test 'what if' scenarios," to draw conclusions and "propose ways to reach a goal" (I.12.9; NAGB, 2010, p. 2-40) on the basis of simulations, to "explain how changes in [a] model result in different outcomes" (I.8.9; NAGB, 2010, p. 2-40), and to "critique the conclusions based on the adequacy of the model to represent the actual problem situation" (NAGB, 2010, pp. 2-38–2-39).

Summary. The NAEP TEL assessment includes—at least to some degree—all components of this engineering practice. Emphasis is placed on the use of models/simulations to test possible solutions, but identifying flaws is also implied by the description in the assessment framework. There appears to be less emphasis in the NAEP TEL assessment (as compared to the NGSS) on refining models and considering both strengths and weaknesses of designs.

Science Practice 3: Planning and Carrying Out Investigations

NAEP Science. The *Using Scientific Inquiry* practice in the NAEP Science framework includes "mak[ing] observations about the natural world" and "collect[ing]… relevant data" (NAGB, 2008, pp. 72–73). As part of this practice, students are expected to "design or critique aspects of scientific investigations" and to "conduct scientific investigations using appropriate tools and techniques… with appropriate levels of precisions" (NAGB, 2008, p. 73). In the NAEP Science hands-on performance tasks (administered to only a subset of students), students "manipulate selected physical objects and try to solve a scientific problem involving the objects, including determin[ing] scientifically justifiable procedures for arriving at the solution" (NAGB, 2008, p. ix). One type of interactive computer task is the "empirical investigation item," which "place[s] [a] hands-on performance [task] on the computer and invite[s] students to design and conduct a study to draw conclusions about a problem" (NAGB, 2008, p. ix). However, the NAEP Science framework also notes that "it is incorrect to assume that assessment of using scientific inquiry is best or only achieved through hands-on performance tasks and interactive computer tasks (NAGB, 2008, p. 74).

NAEP TEL. In the NAEP TEL framework, a key principle in the subarea *Investigation of Problems* includes the idea that "digital tools can be used to conduct experiments," and students are expected to be able "to use digital tools in testing hypotheses" (NAGB, 2010, p. 2-38).

TIMSS Science. The *Reasoning* cognitive domain includes the goal of students being able to "extend their knowledge" (Mullis et al., 2009, p. 84). "Specific behaviors to be elicited by items that are aligned" (Mullis et al., 2009, p. 81) with this cognitive domain include "hypothesize/predict" and "design" (Mullis et al., 2009, p. 86). *Hypothesis/predict* includes "formulat[ing] hypotheses as testable assumptions" (Mullis et al., 2009, p. 86). *Design* is elaborated as

> design or plan investigations appropriate for answering scientific questions or testing hypotheses; describe or recognize the characteristics of well-designed investigations in terms of variables to be measured and controlled and cause-and-effect relationships; make decisions about measurements or procedures to use in conducting investigations. (Mullis et al., 2009, p. 86)

Formulating hypotheses and designing investigations are also identified as two of the "major aspects of the scientific inquiry process" that are expected in the TIMSS Science assessment (Mullis et al., 2009, p. 88).

PISA Science. In the 2015 PISA Science framework, the competency *Evaluate and Design Scientific Enquiry* includes the ability to

> propose a way of exploring a given question scientifically; evaluate ways of exploring a given question scientifically; and describe and evaluate a range of ways that scientists use to ensure the reliability of data and the objectivity and generalizability of explanations. (OECD, 2013b, p. 15)

This competency requires "knowledge of the key features of a scientific investigation, for example, what things should be measured, what variables should be changed or controlled, or what action should be taken so that accurate and precise data can be collected" as well as "an ability to evaluate the quality of data, which in turn depends on recognizing that data are not always completely accurate" (OECD, 2013b, p. 16).

While *Evaluate and Design Scientific Enquiry* is clearly the competency in the PISA framework most clearly aligned with this practice, the other two competencies are also implicated in this practice (although to a lesser degree). As part of the competency *Interpret Data and Evidence Scientifically*, students should be able to judge whether "the procedures that have been applied to obtain any data set" are appropriate (OECD, 2013b, p. 9). In addition, the competency *Explain Phenomena Scientifically* includes the ability to "make and justify appropriate predictions" and "offer explanatory hypotheses"

(OECD, 2013b, p. 15). As noted in the NRC (2012, p. 60) Framework, *Planning and Carrying out Investigations* may entail developing a hypothesis that may be tested in a scientific investigation.

Both *Procedural Knowledge* and *Epistemic Knowledge* are required for *Planning and Carrying out Investigations*. *Procedural Knowledge* in the 2015 PISA Science framework is considered to be "knowledge of the standard procedures scientists use to obtain reliable and valid data" (OECD, 2013b p. 19). The PISA science assessment may test the following "general features of procedural knowledge" (OECD, 2013b, p. 19):

- The concept of variables including dependent, independent, and control variables
- Concepts of measurement, e.g., quantitative [measurements], qualitative [observations]
- Ways of assessing and minimising uncertainty such as repeating and averaging measurements
- Mechanisms to ensure the replicability (closeness of agreement between repeated measures of the same quantity) and accuracy of data (the closeness of agreement between a measured quantity and a true value of the measure)
- The control of variables strategy and its role in experimental design or the use of randomised controlled trials to avoid confounded findings and identify possible causal mechanisms
- The nature of an appropriate design for a given scientific question, e.g., experimental, field based or pattern seeking

The science practice *Planning and Carrying Out Investigations* is also reflected in the 2015 PISA Science framework as *Epistemic Knowledge*:

> Whereas procedural knowledge is required to explain what is meant by the control of variable strategy, being able to explain *why* the use of the control of variables strategy or replication of measurements is central to establishing knowledge in science is epistemic knowledge. (OECD, 2013b, p. 20)

As reflected in the competency *Evaluate and Design Scientific Enquiry,* epistemic knowledge includes an understanding of "what constitutes… appropriate data" (OECD, 2013b, p. 21).

AP Physics. Science Practice 4 of the AP Physics framework is: "The student can plan and implement data collection strategies in relation to a particular scientific question" (College Board, 2012, p. 126). This practice is subdivided as follows: "The student can *justify the selection of the kind of data* needed" (4.1), "*design a plan* for collecting data" (4.2), "*collect data*" (4.3), and *evaluate sources of data*" (4.4) "to answer a particular scientific question" (College Board, 2012, pp. 126–127). In addition, Science Practices 6 ("The student can work with scientific explanations and theories") is described as including the expectation that students can “design experiments to test alternative explanations of phenomena by comparing predicted outcomes” (College Board, 2012, p. 128).

Summary. Although only the NAEP assessments and the PISA science assessment involve students in actually carrying out investigations, the science practice *Planning and Carrying Out Investigations* is well represented across the five reviewed large-scale assessment frameworks. None of the assessment frameworks calls for students to engage in live collaborations, but work with virtual collaborators is included in frameworks for both the NAEP TEL (NAGB, 2010) and PISA Collaborative Problem Solving (OECD, 2013a) assessments.

Engineering Practice 3: Planning and Carrying Out Investigations

NAEP TEL. In the NAEP TEL framework, the engineering practice *Developing Solutions and Achieving Goals* includes “collecting… data to develop a solution and complete a project” (NAGB, 2010, p. 1-9). The *Engineering Design* subarea defines the process of design as including testing ideas in an iterative manner through the use of models and prototypes (NAGB, 2010, p. 2-25), and students are expected to test “models to see if they meet the requirements of a problem” (D.12.9; NAGB, 2010, p. 2-25). In the *Systems Thinking* subarea, students are expected to “test a manufacturing system composed of several machines” (D.12.15; NAGB, 2010, p. 2-28). In addition, a key principle in the subarea *Investigation of Problems* is “digital tools can be used to… investigate practical problems” (NAGB, 2010, p. 2-38).

Summary. Perhaps not surprisingly, given that only one reviewed large-scale assessment framework has an explicit focus on engineering, fewer components of this practice are represented than the corresponding science practice. This may be largely due to less specificity in the NAEP TEL framework, as compared to the NRC Framework and NGSS. The former includes the broad practice of using investigations as part of the design process but does not include as much detail as the NGSS regarding what this practice entails.

Science Practice 4: Analyzing and Interpreting Data

NAEP Science. The *Using Scientific Inquiry* practice in the NAEP Science framework includes “identify[ing] patterns in data” and “the use of logical reasoning” (NAGB, 2008, pp. 72–73).

NAEP TEL. In the NAEP TEL framework, the *Investigation of Problems* subarea (part of *Information and Communication Technology*) includes the expectation that students “analyze and display data” “in order to test hypotheses” (I.8.8, I.12.8; NAGB, 2010, p. 2-28).

TIMSS Science. The TIMSS Science framework notes that “considerable scientific reasoning is… involved in… analyzing and interpreting data” (Mullis et al., 2009, p. 85). A specific behavior to be elicited by items that are aligned with the *Reasoning* cognitive domain is “draw conclusions” (Mullis et al., 2009, p. 86). This behavior includes “detect[ing] patterns in data, describ[ing] or summariz[ing] data trends, and interpolat[ing] or extrapolat[ing] from data or given information” (Mullis et al., 2009, p. 86). Representing data and analyzing and interpreting data are also identified as two of the "major aspects of the scientific inquiry process" that are expected in the TIMSS Science assessment (Mullis et al., 2009, p. 88).

PISA Science. *Analyzing and Interpreting Data* has the closest alignment with the 2015 PISA Science competency *Interpret Data and Evidence Scientifically*, which is defined as including "analys[ing] and evaluat[ing] data… in a variety of representations" (OECD, 2013b, p. 7). Students are expected to "interpret and make sense of basic forms of scientific data and evidence" and "to interpret the meaning of scientific evidence… in their own words, using diagrams or other representations" (OECD, 2013b, p. 16). This competency includes "looking for patterns, constructing simple tables and graphical visualisations such as pie charts, scatterplots or Venn diagrams" (OECD, 2013b, p. 9), as well as "transforming data from one representation to another" (OECD, 2013b, p. 16), and using "the analytical tools offered by spreadsheets and statistical packages" (OECD, 2013b, p. 9). This competency requires "a substantial body of knowledge" to "recognize what constitutes reliable and valid evidence and how to present data appropriately" (OECD, 2013b, p. 9).

Procedural knowledge to be included in the 2015 PISA Science assessment includes knowledge of "common ways of abstracting and representing… data using tables, graphs and charts and their appropriate use," as well as "ways of assessing and minimizing uncertainty" and "the control of variable strategy and its role in experimental design or the use of randomised controlled trials to avoid confounded findings and identify possible causal mechanisms" (OECD, 2013b, p. 19).

AP Physics. Science Practice 5 ("The student can perform data analysis and evaluation of evidence") includes "5.1 The student can *analyze data* to identify patterns and relationships" and "5.2 The student can *refine observations and measurements* based on data analysis" (College Board, 2012, p. 127).

Summary. All five of the reviewed assessment frameworks require students to analyze and interpret data. Although the frameworks are not always as explicit about the individual components of this practice, students are expected to be able to use a variety of formats to display data and to identify patterns in data. However, there is less emphasis on reasoning with respect to the quality of the data being analyzed. These components are partially represented in the PISA and AP Science assessment frameworks.

Engineering Practice 4: Analyzing and Interpreting Data

NAEP TEL. In the NAEP TEL framework, the engineering practice *Developing Solutions and Achieving Goals* includes "represent[ing] data" (NAGB, 2010, p. 3-17), "analyzing… data, [and] interpreting results" (NAGB, 2010, p. 3-5) "to develop a solution and complete a project" (NAGB, 2010, p. 1-9). In the *Investigation of Problems* subarea (part of *Information and Communication Technology*), students are expected to "analyze and display data" (D.12.15; NAGB, 2010, p. 2-28).

The *Engineering Design* subarea defines the process of design as including testing ideas in an iterative manner through the use of models and prototypes (NAGB, 2010, p. 2-25), and students are expected to test “models to see if they meet the requirements of a problem” (D.12.9; NAGB, 2010, p. 2-25).

Summary. This NAEP TEL assessment framework clearly includes the analysis and interpretation of data in order to investigate a particular design solution. It is less clear whether students are also expected to engage in this practice in order to compare solutions, although this can perhaps be inferred from other parts of the framework.

Science Practice 5: Using Mathematics and Computational Thinking

NAEP Science. Although the use of mathematics and computational thinking is not identified as one of the four NAEP Science practices, the framework does note that items should “probe students’ ability to use… quantitative reasoning skills in science... at all three grade levels” (NAGB, 2008, p. 97). In addition, the *Using Science Principles* practice in the NAEP Science framework includes “predict[ing] observations of phenomena… including quantitative predictions based on science principles that quantify quantitative relationships among variables” (NAGB, 2008, p. 69).

TIMSS Science. Both the *Applying* and *Reasoning* cognitive domains include mathematics and computational thinking. “Items aligned with” *Applying* may “involve the direct application or demonstration of relationships, equations, and formulas in contexts likely to be familiar in the teaching and learning of science concepts” (Mullis et al., 2009, p. 83). This cognitive domain includes the specific behavior *find solutions*, which is defined as “identify[ing] or us[ing] a science relationship, equation, or formula to find a qualitative or quantitative solution involving the direct application/demonstration of a concept” (Mullis et al., 2009, p. 84). *Reasoning* involves the specific behaviors *integrate/synthesize*, which include “integrat[ing] mathematical concepts or procedures in the solutions to science problems” (Mullis et al., 2009, p. 85), and *generalize*, which includes “determin[ing] general formulas for expressing physical relationships” (Mullis et al., 2009, p. 87). Eighth-grade students are expected to “demonstrate… analysis skills in selecting and applying appropriate mathematical techniques” (Mullis et al., 2009, p. 89).

PISA Science. The competency *Interpret Data and Evidence Scientifically* “requires the use of mathematical tools to analyse or summarise data” (OECD, 2013b, p. 16), and *Epistemic Knowledge* includes knowing that the construction of models (including mathematical models) “is a key feature of science” (OECD, 2013b, p. 20). However, the PISA Science assessment framework specifies that “questions within the domain of science that assess… mathematical literacy will be avoided” (OECD, 2013b, p. 45).

AP Physics. This scientific practice from the NRC Framework and NGSS is reflected in Science Practice 2 of the AP Physics framework: “The student can use mathematics appropriately” (College Board, 2012, p. 124). This practice includes “*justify[ing] the selection of a mathematical routine* to solve a problem” (2.1) and “*applying mathematical routines* to quantities that describe natural phenomena” (2.2), and “*estimat[ing] numerically* quantities that describe natural phenomena” (2.3; College Board, 2012, p. 123).

In addition, Science Practice 1 (“The student can use representations and models to communicate scientific phenomena and solve scientific problems”) includes mathematical models and “mathematical representations such as equations” (College Board, 2012, p. 123), as expressed in “1.4 The student can *use representations and models* to analyze situations or solve problems qualitatively and quantitatively” (College Board, 2012, p. 123).

Summary. The TIMSS Science and AP Physics assessment frameworks include the most explicit discussion of students’ *Us[e] of Mathematics and Computational Thinking*. However, NAEP and PISA also appear to have this expectation, and, across these four assessment programs, most of the components of this practice are represented. However, approximation to check whether quantitative results make sense is only explicitly included in the AP Physics framework. While all four assessment frameworks include identifying patterns in data, none mentions students’ assessing the significance of these patterns.

Engineering Practice 5: Using Mathematics and Computational Thinking

Although none of the reviewed large-scale assessment frameworks included explicit attention to the use of mathematics and computational thinking in engineering design, it is reasonable to assume that some of the tasks in the NAEP TEL assessment and some of the technological design tasks in the NAEP Science assessment will require this practice.

Science Practice 6: Constructing Explanations

NAEP Science. In the NAEP Science framework, the practice *Using Science Principles* includes “explain[ing] observations of phenomena (using science principles from the content statements)” (NAGB, 2008, p. 68). The *Using Scientific Inquiry* practice includes “propos[ing] explanations to account for patterns,” (NAGB, 2008, p. 72), “relat[ing] patterns in data to theoretical models,” and “using empirical evidence to validate or criticize conclusions about explanations and predictions” (NAGB, 2008, p.73).

NAEP TEL. The NAEP TEL framework (in the *Investigating Solutions* subarea of *Information and Communication Technology*) specifies that students “use digital tools in testing hypotheses” or “conduct[ing] investigations,” which includes “explain[ing] the implications of the results” and “drawing conclusions” (NAGB, 2010, pp. 2-38–2-39). They are expected to “draw and report conclusions consistent with observations” (I.8.8) and to “justify conclusions based on observed patterns in the data” (I.12.8; NAGB, 2010, p. 2-40).

TIMSS Science. The *Applying* cognitive domain involves “the direct application of knowledge and understanding of science in straightforward situations” (Mullis et al., 2009, p. 82). The set of items used to measure this cognitive domain includes those “that require students… to use and apply their understanding of science concepts and principles to find a solution[4] or develop an explanation" (Mullis et al., 2009, pp. 82–83). This cognitive domain includes the specific behavior *explain*, defined as "provid[ing] or identify[ing] an explanation for an observation or natural phenomenon, demonstrating understanding of the underlying concept, principle, law, or theory" (Mullis et al., 2009, p. 84).

The TIMSS Science framework identifies "develop[ing] explanations" and "draw[ing] conclusions" as goals of science education included in the cognitive domain *Reasoning* (Mullis et al., 2009, p. 84). The TIMSS Science assessment may require students "to draw conclusions from scientific data and facts, providing evidence of both inductive and deductive reasoning and of an understanding of cause and effect" (Mullis et al., 2009, pp. 84–85). Eighth-grade students "should consider and evaluate alternative explanations" (Mullis et al., 2009, p. 85). The *Reasoning* cognitive domain includes two specific behaviors with relevance to the scientific practice *Constructing Explanations*. *Draw conclusions* includes "mak[ing] valid inferences on the basis of evidence and/or understanding of science concepts" and "draw[ing] appropriate conclusions that address questions or hypotheses" (Mullis et al., 2009, p. 87). *Evaluate* includes the evaluation of "alternative explanations" and "results of investigations with respect to sufficiency of data to support conclusions" (Mullis et al., 2009, p. 87). "Drawing conclusions and developing explanations" is one of the "five major aspects of the scientific inquiry process" addressed in the TIMSS Science assessment (Mullis et al., 2009, p. 88).

PISA Science. This practice aligns with two competencies in the 2015 PISA Science framework. *Explain Phenomena Scientifically* includes "recognising" and "offer[ing]… explanations for a range of natural and technological phenomena" (OECD, 2013b, p. 7), with a particular focus on "everyday phenomena" (OECD, 2013b, p. 15). This requires students to "recall appropriate content knowledge in a given situation and use it to interpret and provide an explanation for the phenomena of interest" (OECD, 2013b, p. 15). *Interpret Data and Evidence Scientifically* includes "draw[ing] appropriate scientific conclusions" (OECD, 2013b, p. 7). Expanding upon this definition, students should "be able to judge whether… claims are justified" based upon a given data set and "to construct claims that are justified by data" (OECD, 2013b, p. 9). *Epistemic knowledge* includes recognizing that "the purpose and goals of science" are "to produce explanations of the natural world" (OECD, 2013b, p. 21) and "understand[ing] that scientists draw on data to advance claims to knowledge" (OECD, 2013b, p. 20).

AP Physics. *Constructing Explanations* is most closely aligned with the Science Practice 6 ("The student can work with scientific explanations and theories") from the AP Physics framework (College Board, 2012, p. 128). Associated with this practice, the framework clarifies:

> Scientific explanations may specify a cause-and-effect relationship between variables or describe a mechanism through which a particular phenomenon occurs… A scientific explanation,

[4] Although the term *solution* is also used in the NGSS engineering practices, my read of the TIMSS framework is that its use here is not intended to imply a process of engineering design but rather the application of science knowledge to practical situations.

> accounting for an observed phenomenon, needs to be experimentally testable. One should be able to use it to make predictions about a new phenomenon… Students should be prepared to offer evidence, to construct reasoned arguments for their claim from the evidence, and to use the claim or explanation to make predictions. (College Board, 2012, p. 128)

Included as part of this AP Physics practice are "6.1 The student can *justify claims with evidence*," "6.2 The student can *construct explanations of phenomena based on evidence* produced through scientific practices," "6.4 The student can make *claims and predictions about natural phenomena* based on scientific theories and models," and "6.5 The student can *evaluate alternative scientific explanations*" (College Board, 2012, p. 129).

In addition, *Constructing Explanations* is reflected in part of Science Practice 4, "The student can plan and implement data collection strategies in relation to a particular scientific question" (College Board, 2012, p. 126): "4.4 The student can *evaluate sources of data* to answer a particular scientific question" (College Board, 2012, p. 127). The expectations in Science Practice 5 ("The student can perform data analysis and evaluation of evidence") also include components of *Constructing Explanations*: that students will "be able to revise their reasoning based on… new data… that… may appear anomalous" and (5.6) that students "can *evaluate the evidence provided by data sets* in relation to a particular scientific question" (College Board, 2012, p. 127).

Summary. This practice is well represented in the reviewed large-scale assessment frameworks. Almost all of the components of this practice were included in all four large-scale assessment programs. However, identifying gaps or weaknesses in explanatory accounts—which I take to mean evaluating not only the available evidence but also the reasoning used to justify claims on the basis of this evidence—was not explicitly included in any of the frameworks.

Engineering Practice 6: Designing Solutions

NAEP Science. The NAEP Science practice *Using Technological Design* is roughly equivalent to the NGSS practice *Designing Solutions*. In particular, the NAEP Science framework defines *Using Technological Design* as "propos[ing] or critique[ing] solutions to problems, given criteria and constraints"; "identify[ing] scientific trade-offs in design decisions and choose among alternative solutions"; and "apply[ing] science principles or data to anticipate effects of technological design" (NAGB, 2008, pp. 75–76).

NAEP TEL. In the NAEP TEL framework, examples of the practice *Developing Solutions and Achieving Goals* include "apply[ing]… concepts of… engineering design and information technology… to solve problems and meet goals" (NAGB, 2010, p. 1-9), including "generating ideas, selecting between alternatives, optimizing, evaluating the design, and redesigning if needed" (NAGB, 2010, p. 2-18).

Engineering Design (a subarea of *Design and Systems*) includes "using a process of informed decision-making" to "[compare] different solutions to the requirements of the problem and either [choose] the most promising solution or [synthesize] into an even more promising potential solution" (NAGB, 2010, p. 2-22). Key principles of *Engineering Design* include the idea that "there are several possible ways of addressing a design challenge" and that "optimization, which is sometimes part of designing, means finding the best possible solution when some criterion or constraint is identified as the most important and others are given less weight" (NAGB, 2010, p. 2-23). Students "should be able to carry out a full engineering design process… They should be able to…generate [and weigh] alternative solutions," "use the concept of trade-off to balance competing values," and redesign so as to arrive at an optimal solution" (NAGB, 2010, pp. 2-23–2-24). As part of the *Engineering Design* subarea, students should be able to "identify the benefits of a design as well as the possible unintended consequences (D.8.10; NAGB, 2010, p. 2-25).

Other subareas of *Design and Systems* are also relevant to this practice. *Nature of Technology* includes the expectation that students should be able to "redesign an existing tool to make it easier [or more efficient] to accomplish a task" (D.8.5, D.12.5) and to "take into account trade-offs among several factors when selecting a material for a given application" (D.12.4; NAGB, 2010, p. 2-12). *Systems Thinking* includes "examin[ing] a system to predict how it will perform with a given set of inputs in a given situation and how performance will change if the components of interactions of the system are changed" (D.12.13) and "redesign[ing] the system to optimize its efficiency (D.12.15; NAGB, 2010, p. 2-28).

When applying the practice *Developing Solutions and Achieving Goals* in the area of *Design and Systems*, students may be asked "to develop designs, to propose or critique solutions to problems after being given criteria and constraints, to select appropriate resources by considering trade-offs,… or to determine the consequences of making a change in a system" (NAGB, 2010, p. 3-12). Because students may use digital tools to aid in the design process, the *Investigating Problems* subarea of *Information and Communication Technology* includes using digital tools to "investigate alternative solutions" (NAGB, 2010, p. 2-40), to "identify and compare different possible solutions," and to "fully investigate the pros and cons of different approaches" (NAGB, 2010, pp. 2-38–2-39).

In addition, within *Technology and Society*, three subareas have relevance to this practice: Interaction of Technology and Humans; Effects of Technology and the Natural World; and Ethics, Equity, and Responsibility. In the first, key principles include the ideas that "technological decisions should take into account both costs and benefits" and that "when considering technological decisions that involve competing priorities, it is helpful to consider the trade-offs among alternative solutions" (NAGB, 2010, p. 2-6). In the second, "students are expected to recognize that technological decisions involve competing priorities and also to consider the consequences of alternative decisions in developing sustainable solutions to environmental problems" (NAGB, 2010, p. 2-4). In addition, this subarea involves the ability "to investigate the environmental effects of alternative decisions by… considering the trade-offs in different technologies" and to "generate innovative sustainable solutions" to "complex global environmental issues" (NAGB, 2010, p. 2-11). The last subarea "addresses the fact that technological decisions made by some people have impacts on others" and includes "the knowledge and skills that students should have for analyzing the issues, gathering evidence that could support multiple perspectives, and presenting alternative solutions to technological issues that have ethical implications" (NAGB, 2010, p. 2-5). Students "should be able to take into account both intended and unintended consequences in making technological decisions" (NAGB, 2010, p. 2-16) and recognize that "these consequences may be difference for different groups of people" and that "different points of view" should be considered in decisions about technology use (T.12.2; NAGB, 2010, p. 2-17).

When applying the practice *Understanding Science Principles* in the area of *Technology and Society*, students are expected to "[describe] local and global effects of technologies," "[analyze] beneficial and negative impacts," "compar[e] costs and benefits of technologies," and "[predict] potential impacts on society and the environment" (NAGB, 2010, p. 3-5). When applying the practice *Developing Solutions and Achieving Goals* in the same area, students are expected to "develop alternative proposals for a new technology based on an analysis of potential positive and negative impacts" (NAGB, 2010, p. 3-5).

Finally, when applying the practice *Developing Solutions and Achieving Goals* in the area of *Information and Communication Technology*, students may be asked "to use ICT tools to plan an approach to solving a problem" or "to access and use information and data to solve a problem or achieve a goal" (NAGB, 2010, p. 3-17).

TIMSS Science. The *Reasoning* cognitive domain includes *analyze*, defined as "analyz[ing] problems to determine the relevant relationships, concepts, and problem-solving steps" (Mullis et al., 2009, p. 85); *integrate/synthesize*, which includes "provid[ing] solutions to problems that require consideration of a number of different factors or related concepts" (Mullis et al., 2009,p. 85); and *evaluate*, which includes "weigh[ing] advantages and disadvantages to make decisions about alternative processes, materials, and sources; consider[ing] scientific and social factors to evaluate the impact of science and technology on biological and physical systems"; and "evaluat[ing]… problem-solving strategies or solutions" (Mullis et al., 2009,p. 87).

Summary. All of the components of this practice are included in the NAEP TEL assessment framework. In addition, both NAEP Science and TIMSS Science include important components of this practice; however, these two assessment frameworks do not seem to include the iterative refinement and optimization of designs.

Science Practice 7: Engaging in Argument from Evidence

In the NRC Framework, there is considerable overlap between the practices *Constructing Explanations* and *Engaging in Argument from Evidence*. Therefore, there is also overlap in terms of the parts of the various large-scale assessment frameworks that can be thought to align with *Engaging in Argument from Evidence*.

NAEP Science. The NAEP Science framework includes "explain[ing] observations of phenomena (using science principles from the content statements)" (NAGB, 2008, p. 68) as part of the *Using Science principles* practice. It also includes "propos[ing] explanations to account for patterns" (NAGB, 2008, p. 72), "relat[ing] patterns in data to theoretical models," and "using empirical evidence to validate or criticize conclusions about explanations and predictions" (NAGB, 2008, p. 73) as part of the *Using Scientific Inquiry* practice. In addition, *Using Scientific Inquiry* includes "reading or listening critically to assertions in the media, deciding what evidence to pay attention to and what to dismiss, and distinguishing careful arguments from shoddy ones" (NAGB, 2008, p. 73). In addition, in hands-on performance tasks, students are expected to "provide solid data to be used in arguing for and justifying a problem solution" (NAGB, 2008, p. 106).

NAEP TEL. The NAEP TEL framework (in the *Investigating Solutions* subarea of *Information and Communication Technology*) specifies that students "use digital tools in testing hypotheses" or "conduct[ing] investigations," which include "explain[ing] the rationale for the approaches they used in designing the investigation as well as the implications of the results" (NAGB, 2010, pp. 2-38–2-39). They are expected to "draw and report conclusions consistent with observations" (I.8.8) and to "justify conclusions based on observed patterns in the data" (I.12.8; NAGB, 2010, p. 2-40).

When applying the *Communicating and Collaborating* practice in the area of *Information and Communication Technology*, students are expected to "integrate input… from multiple [virtual] collaborators who are peers or experts" and to "integrate feedback from others" and "provide constructive criticism" (NAGB, 2010, p. 3-17).

TIMSS Science. As part of the cognitive domain *Reasoning*, TIMSS Science assessment requires students to demonstrate three specific behaviors: generalize, evaluate, and justify (Mullis et al., 2009, p. 87). *Generalize* includes "mak[ing] general conclusions that go beyond the experimental or given conditions"; *evaluate* includes "evaluat[ing] alternative explanations"; and *justify* is defined as "us[ing] evidence and scientific understanding to justify explanations" and "construct[ing] arguments to support the reasonableness of… conclusions from investigations or scientific explanations" (Mullis et al., 2009, p. 87).

PISA Science. The PISA 2015 Science framework notes that "the scientifically literate individual would understand the function and purpose of argument and critique and why it is essential to the construction of knowledge in science" and would be able "to identify any flaws in the arguments of others" (OECD, 2013b, p. 9). This practice overlaps with all three 2015 PISA Science competencies. *Explaining Phenomena Scientifically* is implicated in that part of its definition that includes "evaluat[ing] explanations for a range of natural and technological phenomena" (OECD, 2013b, p. 7). *Interpret Data and Evidence Scientifically* requires students to "evaluate… claims and arguments in a variety of representations" (OECD, 2013b, p. 7) and "from different sources (e.g., newspaper, internet, journals)" and to "identify the assumptions, evidence, and reasoning in science-related text" (OECD, 2013b, p. 16). This competency "may also involve evaluating alternative conclusions using evidence" and "giving reasons for or against a given conclusion using procedural or epistemic knowledge" (OECD, 2013b, p. 16). In other words, the 2015 PISA Science framework expects students "to be able to identify logical or flawed connections between evidence and conclusions" (OECD, 2013b, p. 17).

Epistemic knowledge includes the understanding of "how scientific claims are supported by data and reasoning in science" (OECD, 2013b, p. 21), understanding that "argument is a commonplace feature of science" (OECD, 2013b, p. 20), and an understanding of both "how measurement error affects the degree of confidence in scientific knowledge" and "the role and significance of peer review as the mechanism that the scientific community has established for testing claims to new knowledge" (OECD, 2013b, p. 20).

AP Physics. In the description of Science Practice 6 ("The student can work with scientific explanations and theories"), the AP Physics framework specifies that "students should be prepared to offer evidence" and "to construct reasoned arguments for their claim from the evidence" (College Board, 2012, p. 128). Students are expected to "*justify claims with evidence*" and "*evaluate alternative explanations*" (College Board, 2012, p. 129).

Summary. All of the assessment frameworks expect students to engage in some form of argumentation. However, the focus of students' arguments in the assessment frameworks seems somewhat narrower than in the NRC Framework and NGSS. The assessment frameworks have a heavy emphasis on argumentation in order to support/refute both explanations and connections between claims and data. The NRC Framework and NGSS also include argumentation about experimental design, techniques of data analysis, interpretation of a given data set, and use of data as evidence. In addition, *Engaging in Argument From Evidence* in the NRC Framework includes reasoning collaboratively about explanations, while the assessment frameworks focus on individual reasoning. Although the NAEP TEL assessment framework does call for students to work with virtual collaborators, this takes the form of providing and receiving feedback, in which the focus is on an individual—rather than a collaborative—explanation. The NRC Framework and NGSS also include students' abilities to reflect on their own arguments and to modify their own work in light of evidence and in response to critiques. While the NAEP TEL assessment framework does include responding to critiques, overall—perhaps due to time limitations in on-demand large-scale assessments—there is little attention to reflection and revision.

Engineering Practice 7: Engaging in Argument from Evidence

In the NRC Framework, there is considerable overlap between this engineering practice and *Designing Solutions*. Therefore, there is also overlap in terms of the parts of the various large-scale assessment frameworks that can be thought to align with *Engaging in Argument From Evidence*.

NAEP Science. The NAEP Science practice *Using Technological Design* includes "propos[ing] or critique[ing] solutions to problems, given criteria and constraints" and "identify[ing] scientific trade-offs in design decisions and choose among alternative solutions" (NAGB, 2008, pp. 75–76).

NAEP TEL. In the NAEP TEL framework, examples of the practice *Developing Solutions and Achieving* Goals include “us[ing] multiple processes and diverse perspectives to explore alternative solutions” and “evaluat[ing] claims and mak[ing] intelligent decisions” (NAGB, 2010, p. 1-9).

Engineering Design (a subarea of *Design and Systems*) includes “using a process of informed decision-making” to “[compare] different solutions to the requirements of the problem and either [choose] the most promising solution or [synthesize] into an even more promising potential solution” (NAGB, 2010, p. 2-22). This process includes “selecting between alternatives,” “optimizing,” and “evaluating the design” (NAGB, 2010, p. 2-18). Key principles of *Engineering Design* include the idea that “there are several possible ways of addressing a design challenge” and that “optimization, which is sometimes part of designing, means finding the best possible solution when some criterion or constraint is identified as the most important and others are given less weight” (NAGB, 2010, p. 2-23). Students “should be able to carry out a full engineering design process… They should be able to…weigh alternative solutions” and “use the concept of trade-off to balance competing values” and “redesign so as to arrive at an optimal solution” (NAGB, 2010, pp. 2-23–2-44).

When applying this engineering practice *Understanding Technological Systems* in the area *Design and Systems,* students are expected to be able to “evaluate multiple representations of a system” (NAGB, 2010, p. 3-11). When applying the practice *Developing Solutions and Achieving Goals* in this area, students may be asked to “critique solutions to problems after being given criteria and constraints, to select appropriate resources by considering trade-offs,… or to determine the consequences of making a change in a system” (NAGB, 2010, p. 3-12).

As with the engineering practice *Designing Solutions*, the area *Technology and Society* also has expectations related to trade-offs involved in engineering design. The subarea *Effects of Technology on the Natural World* addresses “the positive and negative ways that technologies affect the natural world” (NAGB, 2010, p. 2-2), and “students are expected to recognize that technological decisions involve competing priorities and also to consider the consequences of alternative decisions in developing sustainable solutions to environmental problems” (NAGB, 2010, p. 2-4). The subarea *Ethics, Equity, and Responsibility* “concerns the profound effects that technologies have on people, how those effects can widen or narrow disparities, and the responsibility that people have for the societal consequence of their technological decisions” (NAGB, 2010, p. 2-2); “the framework identifies the knowledge and skills that students should have for analyzing the issues, gathering evidence that could support multiple perspectives, and presenting alternative solutions to technological issues that have ethical implications” (NAGB, 2010, p. 2-5).

When applying the practice *Understanding Science Principles* in the area of *Technology and Society*, students are expected to “[describe] local and global effects of technologies,” “[analyze] beneficial and negative impacts,” “compar[e] costs and benefits of technologies,” and “[predict] potential impacts on society and the environment” (NAGB, 2010, p. 3-5). When applying the practice *Developing Solutions and Achieving Goals* in the same area, students are expected to “[analyze] the uses of [a] new technology and evaluate alternatives" (NAGB, 2010, p. 3-5). When applying the practice *Communicating and Collaborating* in this area, students are expected to "use a variety of modalities to represent and exchange data, ideas, and arguments about the advantages and disadvantages of technology" (NAGB, 2010, p. 3-5).

Finally, the *Information Research* subarea of *Information and Communication Technology* includes the ability to "evaluate the credibility of information and data sources" (NAGB, 2010, p. 2-36). Because students may use digital tools to aid in the process of evaluating design solutions, the *Investigating Problems* subarea of *Information and Communication Technology* includes using digital tools to "investigate alternative solutions" (NAGB, 2010, p. 2-40), to "identify and compare different possible solutions," and "fully investigate the pros and cons of different approaches" (NAGB, 2010, pp. 2-38–2-39). Similarly, the *Investigation of Academic and Practical Problems* subarea includes "present[ing] findings in terms of pros and cons of two or more innovative sustainable solutions" (I. 12.7; NAGB, 2010, p. 2-40).

TIMSS Science. The TIMSS Science assessment requires students to demonstrate the specific *Reasoning* behaviors *evaluate* and *justify*. *Evaluate* includes "weigh[ing] advantages and disadvantages to make decisions about alternative processes, materials, and sources; consider[ing] scientific and social factors to evaluate the impact of science and technology on biological and physical systems" and "evaluat[ing] alternative… problem-solving strategies and solutions" (Mullis et al., 2009, p. 87).

Summary. The NAEP TEL framework contains most of the components of *Engaging in Argument From Evidence*. In addition, both NAEP Science and TIMSS Science require students to engage in some aspects of argumentation about design solutions. However, as with the science practice, none of the assessment frameworks includes collaborating to find design solutions to the extent represented in the NRC Framework.

Science Practice 8: Obtaining, Evaluating, and Communicating Information

NAEP Science. Although communicating is not identified as one of the four NAEP Science practices, the framework does note that the expectation of "the ability to communicate accurately and effectively" is "a strand that runs across the practices" (NAGB, 2008, p, 66) and that items should "probe students' ability to use communication skills... at all three grade levels" (p. 97). According to the NAEP Science framework, "accurate and effective communication" includes "writing clear instructions that others can follow to carry out an investigation"; "organizing data in tables and graphs"; "using audio, video, multimedia, and other technologies to access, process, and integrate scientific findings; using language and scientific terms appropriately; drawing pictures or schematics to aid in descriptions of observations; summarizing the results of scientific investigations; and reporting to various audiences about facts, explanations, investigations, and data-based alternative explanations and designs" (p. 66).

The NAEP science framework also includes the expectation that students will obtain and evaluate information. The "ability to communicate accurately and effectively" includes "reading… data in tables and graphs" and "locating information in computer databases" (NAGB, 2008, p. 66). In addition, one type of interactive computer task (taken by a subset of students), is the "information search and analysis" item, which "pose[s] a scientific problem and ask[s] students to query an information database and analyze relevant data to address the problem" (NAGB, 2008, p. ix).

NAEP TEL. In the *Investigation of Problems* subarea (part of *Information and Communication Technology*), students are expected to "analyze and display data" (D.12.15; NAGB, 2010, p. 2-28). In addition, the area *Information and Communication Technology* includes skills such as "research and information fluency" (NAGB, 2010, p. 1-8). Within this area, both *Construction and Exchange of Ideas and Solutions* and *Information Research* have the expectation that students are able to "synthesize information from different sources" (NAGB, 2010, p. 2-34). In addition, the latter "includes the capability to employ technologies and media to find, evaluate, analyze, organize, and synthesize information from different sources" (NAGB, 2010, p. 2-34). The subarea *Investigation of Problems* includes the knowledge that "information can be distorted, exaggerated, or otherwise misrepresented" and that, to "[ensure] quality of information," it is important to "[assess] the source of information" and to "[use] multiple sources to verify" information (NAGB, 2010, p. 2-36).

TIMSS Science. The *Applying* cognitive domain includes the specific behavior *interpret information*, defined as "interpret[ing] relevant textual, tabular, or graphical information in light of a science concept or principle" (Mullis et al., 2009, p. 83).

PISA Science. The competency *Interpret Data and Evidence Scientifically* includes being "able to interpret the meaning of scientific evidence and its implications to a specified audience in their own words, using diagrams or other representations" (OECD, 2013b, p. 16). The definition of scientific literacy in the 2015 PISA Science framework rests upon the premise that "in their lifetimes, individuals will need to acquire knowledge, not through scientific investigations, but through the use of resources such as the libraries and the internet" (OECD, 2013b, p. 6). "The competency of [*E*]*valuating and Designing scientific enquiry* is required to evaluate reports of scientific findings and investigations critically" (OECD, 2013b, p. 15). This competency includes "understand[ing] the importance of developing a sceptical disposition to all media reports in science," "recognising that all research builds on previous work, that the findings of any one study are always subject to uncertainty, and that the study may be biased by the sources of funding" (OECD, 2013b, p. 16). This competency draws upon procedural and epistemic knowledge (OECD, 2013b, p. 16), which "are essential to deciding whether the many claims to knowledge that pervade contemporary media have been derived using appropriate procedures and are warranted" (OECD, 2013b, p. 6).

In addition, the competency *Interpret Data and Evidence Scientifically* includes being able to "evaluate scientific arguments and evidence from different sources (e.g., newspaper, internet, journals)" and "accessing scientific information and… evaluating alternative arguments based on scientific evidence" (OECD, 2013b, p. 16).

AP Physics. Although not explicitly mentioned in the AP Physics framework, the free-response section included on the AP Physics assessment does require students to produce substantial written work.

Summary. All of the assessment frameworks require students to provide some written responses to open-ended questions; thus, all include the expectation that students can communicate ideas in writing and/or other tools. However, the assessment frameworks vary in the amount of writing students are expected to produce. In addition, none mentions the ability to communicate ideas orally or through extended discussions with peers.

With the exception of AP Science, the reviewed assessment frameworks include the expectation that students will derive meaning from, critically evaluate, and/or synthesize sources of scientific information. The NAEP TEL assessment framework includes almost all of the components of *Obtaining and Evaluating Information*, as defined in the NRC Framework and NGSS. However, similar to *Communicating Information*, none of the assessment frameworks includes the expectation that students will work with oral forms of scientific information.

Engineering Practice 8: Obtaining, Evaluating, and Communicating Information

NAEP TEL. In the NAEP TEL framework, the area *Information and Communication Technology* and practice *Communicating and Collaborating* most clearly align with this practice. The former "includes the capability to communicate ideas and solutions" (NAGB, 2010, p. 3-17). Within the subarea *Construction and Exchange of Ideas and Solutions*, students are expected to "communicate information and ideas effectively using a variety of media, genres, and formats for multiple purposes and a variety of audiences" (I.8.3; NAGB, 2010, p. 2-35). This requires the knowledge that "communicating always involves understanding the audience—the people for whom the message is intended" (NAGB, 2010, p. 2-34) and the ability to "take into account the perspectives of different audiences" (NAGB, 2010, p. 2-35). The latter (the practice of *Communicating and Collaborating*) includes a very similar expectation: "communicat[ing] information and ideas effectively to multiple audiences using a variety of media and formats" (NAGB, 2010, p. 1-9). In addition, this practice includes "develop[ing] representations and "[sharing] ideas, designs, data, explanations, models, arguments, and presentations" (NAGB, 2010, p. 3-3).

The subarea *Engineering Design* includes communicating results. The NAEP TEL framework notes that "designing usually concludes with a presentation to clients or other interested parties… on the preferred solution," and one of the key principles of this subarea is that "engineering design usually requires one to develop and manipulate representations," including "drawings, charts, and graphs" (NAGB, 2010, p. 2-23). Students are expected to be able to "communicate the entire design process from problem definition to evaluation of the final design" (D.12.10) and to "communicate the results of a design process and articulate the reasoning behind design decisions by using verbal and visual means" (D.8.10; NAGB, 2010, p. 2-23).

In the NAEP TEL framework, the area *Information and Communication Technology* includes skills such as “research and information fluency” (NAGB, 2010, p. 1-8). Within this area, both *Construction and Exchange of Ideas and Solutions* and *Information Research* have the expectation that students are able to “synthesize information from different sources” (NAGB, 2010, p. 2-34). In addition, the latter “includes the capability to employ technologies and media to find, evaluate, analyze, organize, and synthesize information from different sources” (NAGB, 2010, p. 2-34). The subarea *Investigation of Problems* “concerns the use of information and communication technology to define and solve problems” (NAGB, 2010, p. 2-3) and requires that students are able to “formulate a set of questions that will guide them in their search,” “to formulate efficient search strategies,” and to judge the relevance of information and data “to the question at hand” (NAGB, 2010, p. 2-36). As part of this subarea, students are expected to know that “information can be distorted, exaggerated, or otherwise misrepresented” and that, to “[ensure] quality of information,” it is important to “[assess] the source of information” and to “[use] multiple sources to verify” information (NAGB, 2010, p. 2-36).

In addition, the subarea *Engineering Design* includes researching ideas. In defining *researching ideas*, the NAEP TEL framework specifies that “next steps” after a challenge has been defined “are often to investigate relevant scientific and technical information and the way that similar challenges have been solved in the past” (NAGB, 2010, p. 2-22).

Summary. Most components of communicating information are at least implied by the NAEP TEL assessment framework; however, the NAEP TEL framework is less explicit (as compared to the NRC Framework and NGSS) about students’ ability to communicate orally. Consistent with other collaborative components of the NRC Framework—the NAEP TEL assessment framework does not appear to require students to communicate about design solutions through extended discussions with peers. The NAEP TEL assessment framework has a heavy emphasis on evaluating information to be used in the design process.

Overall Summary

Across the five reviewed large-scale assessment frameworks, most of the components of the science practices in the NRC Framework and NGSS are included in at least one assessment framework, and many are addressed in multiple assessment frameworks. Although not surprising, it is worth noting that components of the engineering practices are not as well represented in the large-scale assessment frameworks. However, the NAEP TEL assessment framework does seem to have quite a bit of overlap with the NRC Framework and NGSS, particularly with respect to the three practices highlighted in the NGSS (NGSS Lead States, 2013, Introduction, p. 4): *Developing and Using Models, Designing Solutions, and Engaging in Argument From Evidence*. Thus, while assessment items assessing the engineering practices are not widespread, there is reason for optimism, in that efforts are already underway to assess components of these practices.

While the alignment between the NRC Framework/NGSS and the assessment frameworks is overall quite positive, two caveats are in order. First, it is important to examine the components of each practice that are *not* well represented in the reviewed large-scale assessment frameworks. As reflected in the analysis above, the NRC Framework and NGSS appear to require students to engage in a wider range of practices than is indicated by the reviewed large-scale assessment frameworks. Particularly for the practices of *Developing and Using Models* and *Engaging in Argument from Evidence*, the assessment frameworks expect students to be able to use these practices for relatively narrow purposes, while the NRC Framework and NGSS present a broader view of how these practices are used in the whole process of scientific inquiry or engineering design. For example, while students are expected to engage in argumentation to reason about how data support a claim and to engage in argument to find the best solution, the NRC Framework and NGSS also call for reasoning and argumentation about experimental design, techniques for data analysis, interpretations of a given data set, and the selection of data as evidence.

The NRC Framework and NGSS also include components of practices that may be difficult to assess in on-demand large-scale assessments. These include oral components of the practices, such as communicating ideas orally and engaging in extended and/or collaborative discussions with peers. These also include components that are part of longer term inquiry and design work, such as reflecting on one's own work and revising explanations and solutions, based on relevant evidence and interactions with others. As will be discussed in more detail below, further work will be needed to identify means of eliciting students' performances of these hard-to-assess, but crucially important, components of the science and engineering practices.

Second, what is most important in terms of assessing the NGSS is not the available assessment frameworks, but rather the available items. The assessment frameworks provide some insights into where we might look for examples of assessment items, but—ultimately—any consideration of lessons learned from large-scale assessment systems must also occur at the item level. Therefore, in the next section, I turn to a consideration of innovative large-scale assessment items.

Large-Scale Assessment Examples

As noted in the NRC Framework, "sophisticated models of learning do not by themselves produce high-quality assessment information. Also needed are methods and tools both for eliciting appropriate and relevant data from students and for interpreting the data collected about their performance" (NRC, 2012, p. 318). Therefore, in considering what can be learned to inform development of assessments relative to the NGSS, it is crucial that we consider not only the frameworks, but the items that are actually developed as a result.

Drawing from the framework analysis above, in this section, I examine examples from each of the assessments in order to explore the extent to which sample tasks and items from large-scale assessments seem consistent with the vision of science achievement articulated in the NRC Framework and NGSS. It is important to note that this analysis is intended to be illustrative, rather than comprehensive. With guidance from the respective assessment developers, I have selected a few independent items or one task (which could consist of a set of related items) per assessment to analyze below. These items were chosen to highlight innovative approaches to the assessment of students' science achievement.[5] Given the breadth of the assessment frameworks, a small sample from each assessment cannot be thought to fully represent the assessment framework. In addition, issues of test security limited the items available for this analysis. Thus, other items (whether released or unreleased) might align differently and/or better with the NGSS; this analysis is not intended to make broad claims about the alignment between these assessments and the NGSS. Rather, it is intended to give a sense for the extent to which existing assessment tasks might help to inform the design of assessments of the NGSS.

Assessment capabilities are changing rapidly, and computer-based assessments are a promising means for "measur[ing] not only deep conceptual understanding but also the science practices that are difficult to assess using paper-and-pencil tests or hands-on laboratory tasks" (NRC, 2012, p. 262). As noted in the NRC Framework, NAEP Science, NAEP TEL, and PISA all include at least some computer-based tasks. "There is hope that some of these early developments in large-scale testing contexts can be used as a springboard for the design and deployment of assessments… that support aspects of the framework" (NRC, 2012, p. 263). However, because these assessments are quite new, few examples of computer-based tasks are publicly available. As additional examples are generated and shared, it is likely that items developed for future iterations of large-scale assessments will provide additional guidance for the assessment of the NGSS.

In selecting the items for inclusion below, I tried to pay attention to aspects of the NRC Framework and NGSS that might pose particular assessment challenges, particularly those aspects that are not readily assessed using "paper-and-pencil presentation and response formats" (NRC, 2012, p. 262). Because "practices such as models, arguments, and explanations are more prominent throughout the standards in order to ensure rigorous content receives its due focus" (NGSS Lead States, 2013, Introduction, p. 4), the six corresponding science and engineering practices (*Modeling*, *Constructing Explanations/Designing Solutions*, and *Argumentation*) are emphasized below.

[5] In order to focus on innovations in the assessment of students' science achievement and because the NAEP, PISA, and AP assessments have recently undergone (or are in the process of undergoing) substantial changes, these assessments are the focus of the discussion below. However, it should be noted that the TIMSS Science framework includes some components of the science and engineering practices that are not well represented in the other assessments. Thus, even though the corresponding framework is not new, examination of specific items from the TIMSS Science assessment may be warranted.

NAEP Science

Six hands-on tasks from the 2009 NAEP Science assessment have been released, two from each grade level. Because NAEP Science is the only one of the reviewed large-scale assessments that includes hands-on tasks, I have selected a hands-on task for analysis. The Magnetic Fields task (NCES, n.d.b) was administered to grade 8 students and required them to

> design and conduct investigations based on observations of magnetic properties to determine what materials make up four metal bars. First, they use only the metal bars themselves. Students then repeat the investigation using a test magnet and compare the results of the investigations to confirm their conclusions. Finally, students design and conduct two different tests to compare the magnetic strength of a strong and a weak magnet. (NCES, 2012b, p. 4)

To complete this task, grade 8 students received a test booklet and the materials shown in Figure 1 (a test magnet, four metal bars, a bag of steel washers, a centimeter ruler, and a piece of white paper with a grid). Before beginning the task, students were informed that they would be asked to "design and conduct an investigation to identify four mystery metal bars using two different methods" and to "make measurements of the magnetic properties of these metal bars" (NCES, n.d.b, p. 5). The directions also specify that students will be scored on how well they (NCES, n.d.b, p. 5):

- design [their] procedures to identify metal bars,
- design [their] procedures to compare the strengths of different magnets,
- record [their] observations, and
- provide explanations based on [their] investigation.

In Part 1 (Identifying Metal Bars Using Only the Bars, shown in Figure 2), students must use only the four metal bars to identify which is a strong magnet, which is a weak magnet, which is a copper bar, and which is a steel (iron) bar (NCES, n.d.b, p. 6). In addition to identifying each bar, students are asked to "describe [their] procedures and observations (what [they] did and what [they] saw that helped [them] identify each bar)" (NCES, n.d.b, p. 6).

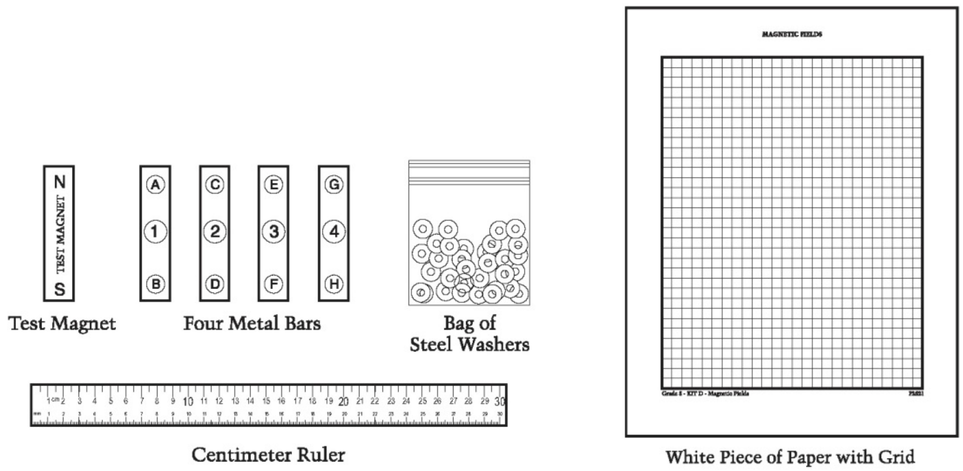


Figure 1. Hands-on materials students use to complete the task. From* The Nation's Report Card: ICT/HOT Grade 8 Magnetic Fields. Test Booklet*, by National Center for Education Statistics, n.d.a, Diagram 1, p. 2.

PART 1

In Part 1, you will identify the four metals bars (1, 2, 3, and 4).

- One is a strong magnet.
- One is a weak magnet.
- One is a copper bar.
- One is a steel (iron) bar.

1. You will conduct experiments to identify each numbered metal bar.

   You will use only the four numbered bars. You will record your results in Table 1 on page 7.

   Place the numbered metal bars (1, 2, 3, and 4) in front of you on the sheet of grid paper you were given.

   Identify all four numbered metal bars using only the four numbered metal bars and nothing else from the kit.

   In the middle column of Table 1, describe your procedures and observations (what you did and what you saw) that helped you identify each bar. Make sure you describe how you used each of the four numbered metal bars to make your identifications. In the last column, record the identity of each numbered metal bar.

   Remember: For Part 1 you will only need the four numbered metal bars. You will use the test magnet later in Part 2.

TABLE 1: IDENTIFYING METAL BARS

Bar Number	Procedures and Observations (What You Did and What You Saw)	Identity of Metal Bar (Strong Magnet, Weak Magnet, Copper Bar, or Steel Bar)
1	Procedure: Observations:	
2	Procedure: Observations:	
3	Procedure: Observations:	
4	Procedure: Observations:	

Figure 2. Directions and data table for Part 1: Identifying Metal Bars Using Only the Bars. From* The Nation's Report Card: ICT/HOT Grade 8 Magnetic Fields. Test Booklet*, by National Center for Education Statistics, n.d.b, pp. 6–7.

Part 1 (item 1) is scored in four parts, one for each bar. In each part, students' responses are scored in four levels (from complete to unsatisfactory/incorrect). Students are expected to correctly identify each bar and to provide "a complete procedure" and "correct observations" specific to the magnetic properties of the bar. To identify the steel bar, students should test "for attraction and repulsion," observing that "both ends of the bar are attracted to the bar identified as a magnet." To identify the strong/weak magnet, students should test "for attraction, repulsion, and relative strength," observing that the bar attracts one end of the magnet and "repels the other end" and that it "attracts [the steel bar] more strongly/weakly than" the weak/strong magnet does. To identify the copper bar, students should indicate that they have held the bar "next to any bar identified as a magnet" and that "the bar identified as a magnet has no effect" on the bar. Students receive a composite score for the item, based on their scores for each bar (NCES, n.d.a).

In Part 2 (Identifying Metal Bars Using the Test Magnet), students are asked to use only the test magnet to identify the four metal bars (item 2). As in Part 1, they record both their identifications and their procedures and observations in a table. They are cautioned:

> Your results might not be the same using the two different methods in Part 1 and Part 2. If your identifications of the metal bars changed from what you recorded in Table 1 [Part 1], keep your original answers in both Table 1 and Table 2. (NCES, n.d.b, p. 8)

In addition, in Part 2, students are asked to justify their work: "Explain how you identified each of the four metal bars using the test magnet. Refer to your observations in Table 2… to support your explanation" (NCES, n.d.b, p. 10). Then (item 3), they are asked to compare the results obtained in Parts 1 and 2, to make a final decision about the identity of each bar and to explain why their results were the same/different in Parts 1 and 2 (NCES, n.d.b, pp. 11–12). See Figure 3.

Item 2 is scored in five parts, one for each bar and one for the justification for their work. For each bar, students' responses are scored as in item 1, based on their identification of the bar and provision of *a complete procedure* and *correct observations* specific to the magnetic properties of the bar, similar to item 1. Students' justifications are scored at five levels (complete to unsatisfactory/incorrect). A complete justification

> provides a complete explanation for identifying all four numbered bars using the test magnet and refers to their observations to support each identity. The response mentions distinguishing the strong from the weak magnet (stronger force, pull, etc.), that magnets attract and repel, that steel only attracts the test magnet, and that copper does nothing (neither attracts nor repels the test magnet) as evidence to support the conclusions. (NCES, n.d.a)

As for item 1, students receive a composite score for the item, based on their scores for each bar and for their justification.

VC320080

3. Look at your results from the two different methods in Table 1 on page 7 and Table 2 on page 9.

In the second column of Table 3, for each bar write either "different" if you recorded a different identity or "same" if you recorded the same identity from this comparison of your results in Tables 1 and 2.

Then, in the last column of Table 3, fill in your final decision for the identity of each metal bar using the results from both methods. Base your decision on your observations in Tables 1 and 2.

TABLE 3: COMPARING THE IDENTITIES OF THE METAL BARS

Bar Number	Comparison of Identities in Tables 1 and 2 (Different or Same)	Identity of Metal Bar Based on Observations in Tables 1 and 2 (Strong Magnet, Weak Magnet, Copper Bar, or Steel Bar)
1		
2		
3		
4		

Remember: Keep your original identifications in Table 1 and Table 2. They will help you make comparisons in Table 3 and answer the question on the next page.

Are your identifications of the metal bars the same in Table 1 as in Table 2?

- (A) Yes, my identification of each bar is the same in Table 1 as in Table 2.
- (B) No, my identifications of some of the bars are different in Table 1 than in Table 2.
- (C) No, my identifications of all of the bars are different in Table 1 than in Table 2.

Explain why based on your experiments in Part 1 and Part 2. Include the bar numbers in your answer.

Figure 3. Item from Part 2: Identifying Metal Bars Using the Test Magnet. From* The Nation's Report Card: ICT/HOT Grade 8 Magnetic Fields. Test Booklet, *n.d.b, by National Center for Education Statistics, pp. 11–12.

Item 3 (Figure 4) is scored in two parts, one for the table and one for the explanation. Five levels (complete to unsatisfactory/incorrect) are available for the table, based on correct identification of the four metal bars and correct comparison between Tables 1 and 2. Three levels are available for the explanation, with a complete response providing "conclusive evidence for identifying at least two of the metal bars based on either or both of the investigations in Parts 1 and 2 of the task" (NCES, n.d.a). As for the previous two items, students receive a composite score for this item, based on scores for the two parts.

In Part 3 (Comparing the Strengths of the Magnets), students "design and conduct two different tests to compare the magnetic strength of the *strong magnet* and the *weak magnet* identified in Part 2. In item 2 (Figure 4), they are asked to "describe [their] materials and procedures" and to "record all measurements [they] made, including numbers and units" (NCES, n.d.b, p. 13). Finally, in Item 5, they identify which bar is the strong magnet and which is the weak magnet and "explain how [their] data and observations showed that the strong magnet is stronger than the weak magnet" (NCES, n.d.b, p. 15).

PART 3

In Part 3, you will design and conduct two different tests to compare the magnetic strength of the strong magnet and the weak magnet that you identified in Part 2. You will record your results in Table 4 on page 14.

- In the top row of Table 4 (columns 2 and 3), write down the numbers on the bars you identified as the strong magnet and the weak magnet in the spaces provided. Use the final identifications in Table 3 on page 11.

4. Think about how you can use the materials in your kit to test these magnets and demonstrate that the strong magnet is stronger than the weak magnet. You may use any of the materials in your kit.

   - Now conduct two different tests to compare the magnetic strength of the strong magnet and the weak magnet.
   - In Table 4, describe the materials and procedures you used. Record all measurements you made, including numbers and units in columns 2 and 3 under Results/Data.
   - Do not forget to write in the number of the metal bar that you think is the strong magnet and number of the magnetic bar you think is the weak magnet.

TABLE 4: COMPARING THE STRONG AND WEAK MAGNETS

	Strong Magnet _____ (number)	Weak Magnet _____ (number)
	Results/Data	Results/Data
Test 1: Materials: Procedure:		
Test 2: Materials: Procedure:		

Figure 4. Directions and data table for Part 3: Comparing the Strength of Magnets. From* The Nation's Report Card: ICT/HOT Grade 8 Magnetic Fields. Test Booklet*, n.d.b, by National Center for Education Statistics, pp. 13–14.

Item 4 is scored in three parts: "identification of strong and weak magnets" and one part for each test's "procedure and results." Three levels are available for each part of the item. For the two tests, students are expected to provide "complete procedure[s]" for two different tests, including "the materials to use and the steps to follow." Examples of such tests include "testing how many steel disks the magnets can hold" or "determining the smallest distance that a magnet can be placed near the unlike pole of the test magnet before one of the bars begins to be moved and pulled toward the other bar." Students are expected to provide "numerical data." The scores for the two tests are combined into a single composite score (NCES, n.d.a).

Item 5 is scored in two parts, one for their "identification of the strong and weak magnets" and one for their "evidence for identification." For both parts, three score levels are available. Complete evidence "provides a valid explanation of how the data and observations in Table 4 [Figure4] showed that the strong magnet is stronger than the weak magnet for at least one of the tests," using "numerical results." Students receive a composite score based on both parts of the item (NCES, n.d.a).

Considering Table 1, this task reflects the following science practices from the NRC Framework and NGSS:

- Science Practice 3: Planning and Carrying Out Investigations
  - Identify relevant variables
  - Consider how variables might be observed and/or measured
  - Observe and collect data to describe a phenomena
  - Design plans for investigations individually
- Science Practice 4: Analyzing and Interpreting Data
  - Use tabulation to collate, summarize, and display data
  - Use data as evidence
- Science Practice 5: Using Mathematics and Computational Thinking
  - Recognize dimensional quantities and use appropriate units
- Science Practice 6: Constructing Explanations
  - Link evidence to claims
- Science Practice 7: Engaging in Argument from Evidence
  - Formulate evidence based on data (S4)
- Science Practice 8: Obtaining, Evaluating, and Communicating Information
  - Communicate ideas in writing
  - Communicate ideas with the use of tables, diagrams, graphs, and equations

Thus, overall, this task includes components of almost all of the science practices from the NRC Framework and NGSS. In addition, it is important to note that, in order to identify the bars and design tests of the magnets' relative strengths, Parts 1 and 2 require students to use scientific knowledge about magnets (in particular the attractive/repulsive behavior of the two ends of a magnet and that iron is attracted to magnets but copper is not). Thus, students must draw upon content knowledge—and not simply content-free science practices—in order to complete this task. As such, it provides an example of the integration of content and practices that is required by the NRC Framework and NGSS.

NAEP TEL

As the NAEP TEL assessment has yet to be administered, only one sample task is available (NCES, 2013). In this task, grade 8 "students are asked to play the role of an engineer who is brought in to a remote village to find out why the local water well has stopped working" (NCES, 2013). Figures 5–15 display screenshots for this task. An associated scoring guide is not available for this task.

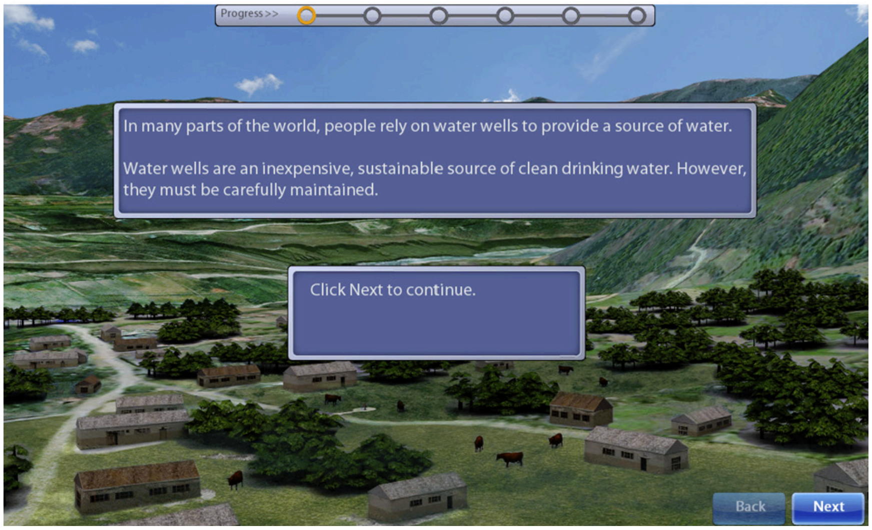


Figure 5. Sample NAEP TEL Task: Beginning of the task. From "Sample TEL Task," by National Center for Education Statistics, 2013.

The task begins with a series of screens, describing the problem to be solved. (Figure 5 shows the first screen in this sequence.) By clicking through the opening screens, students read the following scenario (NCES, 2013):

> In many parts of the world, people rely on water wells to provide a source of water.
>
> Water wells are an inexpensive, sustainable source of clean drinking water. However, they must be carefully maintained.
>
> In this activity, you'll be traveling to Ramnagar, a remote village in eastern Nepal.
>
> Unfortunately, the well in Ramnagar is not working. You're part of a team that will help repair the well.

Next, students are introduced to Kumar, who serves as a guide throughout the task. Figure 6 shows the first screen in this sequence.) By clicking through a series of screens, displaying what looks like a chat window with Kumar, students read the following introduction and additional information about the problem to be solved (NCES, 2013):

> Hi! I'm Kumar. I study engineering at Kathmandu University. Today we're going to fix the well in Ramnagar.
>
> The new well we dug for Ramnagar about three years ago isn't working anymore.

> The well is the main source of clean water for the village, so it's important to fix it.
>
> First, you need to know a little bit about how a well works.
>
> A well draws water from an aquifer—an underground water supply.

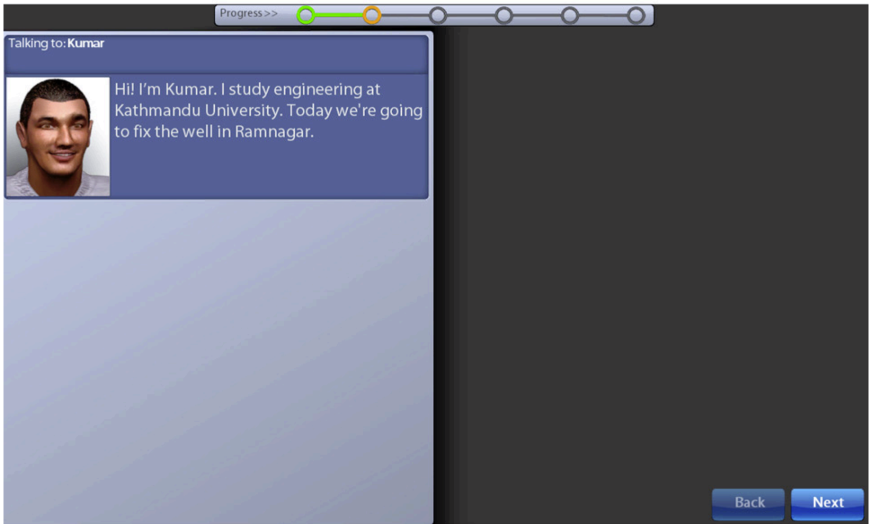


Figure 6. Sample NAEP TEL Task: Screenshot of Kumar (guide to the task). From "Sample TEL Task," by National Center for Education Statistics, 2013.

The next screen shows a chat window, in which Kumar has written "Ramnagar's well draws from the aquifer shown in the map on the right. Other villages draw water from the same aquifer," and a map (see Figure 7). Kumar then guides students through two animations, showing how water collects in an aquifer and how water is obtained from the aquifer using a well (Figures 8 and 9). As they watch the former animation, students read "Aquifers contain water from rainfall that has seeped underground." As they watch the latter animation, students read the following information (NCES, 2013):

> To get water out of the aquifer, a borehole is drilled into the ground.
>
> A pump is placed on top of the borehole.
>
> When you push on the pump's handle, it's supposed to bring water up to the surface.

Figure 7. Sample NAEP TEL Task: Screenshot of information about the location of the aquifer and surrounding villages. From "Sample TEL Task," by National Center for Education Statistics, 2013.

Figure 8. Sample NAEP TEL Task: Screenshot from animation of rainwater collecting in an aquifer. From "Sample TEL Task," by National Center for Education Statistics, 2013.

Figure 9. Sample NAEP TEL Task: Screenshot from animation of water being obtained from an aquifer. From "Sample TEL Task," by National Center for Education Statistics, 2013.

The screen following the one in Figure 9 shows the s same scenario, but with no water coming from the pump, and students read "But right now, no water is coming out of the well."

On the next screen, Kumar introduces a villager:

Let's talk to someone in the village and learn more about the problem.

Laxmi might be able to help us. She lives in Ramnagar, and she used the well every day until it quit working.

The next screen (Figure 10) shows students how to ask Laxmi a question. The first question (predetermined) is "What happens when you try to get water out of the well?" Laxmi responds, "When I try pushing the handle, no water comes out of the well."

Kumar then provides additional information and instructions:

Laxmi says no water is coming out of the well. That could be caused by two things.

First, there might not be enough water underground. This could happen if there has not been enough rain to refill the aquifer as water is pumped out.

Second, there could be a problem with the pump. It could be clogged or broken.

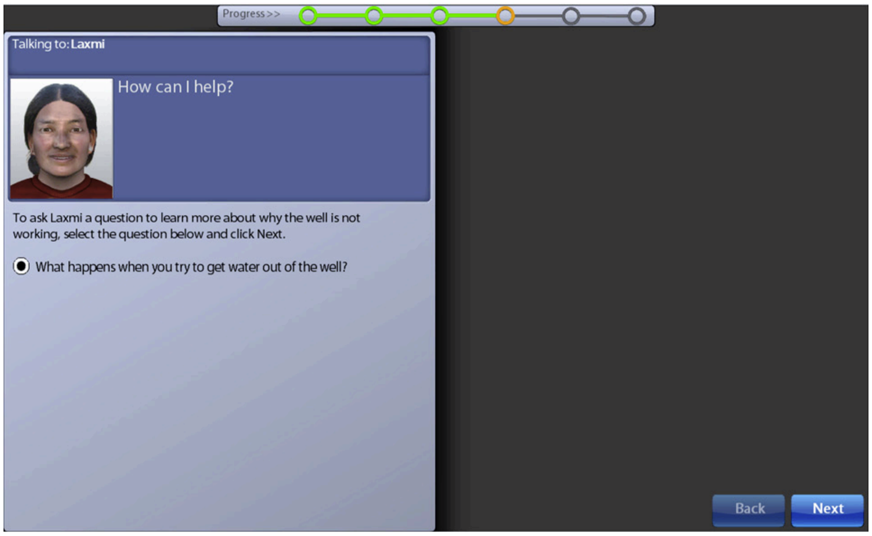


Figure 10. Sample NAEP TEL Task: Screenshot of the interface that allows students to ask Laxmi (a villager) questions about the well. From "Sample TEL Task," by National Center for Education Statistics, 2013.

As students read the two reasons, a summary of the "possible reasons the well is not working" appears to the right of the chat window. This text box remains on the screen throughout the time that students are working to determine why the well is not working. Kumar instructs students to "ask Laxmi more questions to help you understand what is causing the problem with the well." On the next screen (Figure 11), students read that they can select three (of four possible) questions to ask Laxmi. After Laxmi responds to each question, students are asked, "Does Laxmi's answer help you understand why the well is not working?" If they answer "yes," they are asked, "What does Laxmi's answer suggest about why the well is not working?" and must select either "There is probably not enough water underground" or "The pump is probably clogged or broken." After reading Laxmi's responses to their three questions and responding to the corresponding item(s), Kumar reappears with the following information: "Based on what Laxmi told us, there is probably plenty of water underground. The problem is probably with the pump."

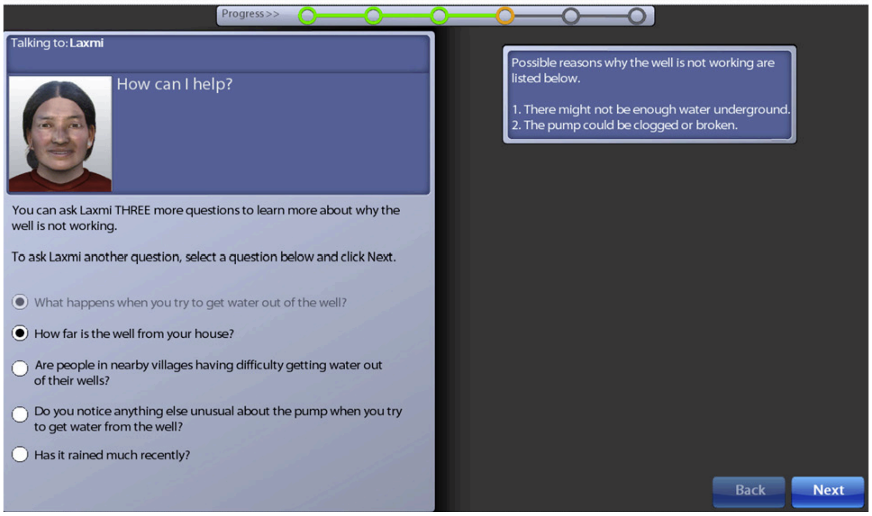


Figure 11. Sample NAEP TEL Task: Screenshot of questions students can ask Laxmi to identify which of the two reasons accounts for the well not working properly. From "Sample TEL Task," by National Center for Education Statistics, 2013.

Kumar then introduces the next part of the task: "We can use the pump repair manual to help us find out what's wrong with the pump." On the next screen, the text "The tutorial will show you how to use the pump repair manual" appears on the left, and a diagram of the pump appears on the right. As students click through the following screens, they read directions for using the manual—

> To use the manual, you will click on a problem to investigate it further.
>
> When you select a problem, the parts of the pump that are related to the problem are labeled for you.
>
> Instructions for investigating and repairing the problem are shown below.

—ending with the screen in Figure 12. Continuing the tutorial, students are told to "Click the Test Pump button to try using the pump. This might give you more information about what's going wrong." When students click the button, they see the screen in Figure 13, with the following description of the pump's operation: "You notice that the handle is difficult to push down, and there's a loud squeaking noise. No water comes out."

> The following screens give instructions for students' interaction with the repair manual:
>
> Now you are ready to use the pump repair manual to find out what's wrong with the pump.
>
> Use the manual to help you repair the pump. You should ONLY perform the checks and repairs that are necessary.

After you have repaired the pump, click Test Pump to make sure it is working.

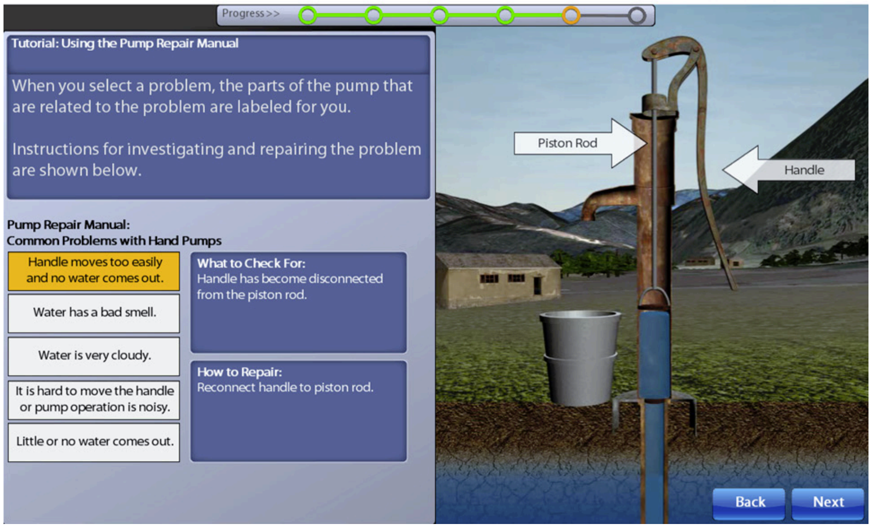


Figure 12. Sample NAEP TEL Task: Screenshot of tutorial on the use of the pump repair manual. From "Sample TEL Task," by National Center for Education Statistics, 2013.

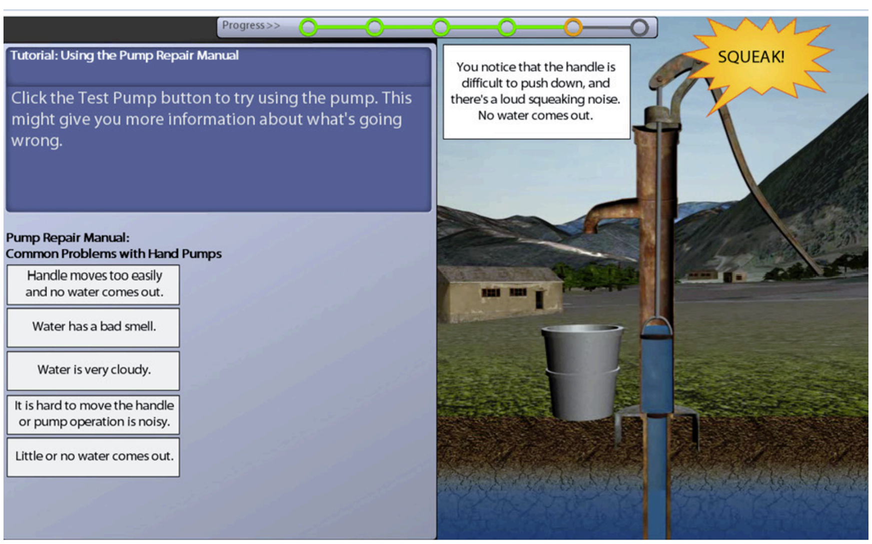


Figure 13. Sample NAEP TEL Task: Screenshot of the result when students click on the Test Pump button. From "Sample TEL Task," by National Center for Education Statistics, 2013.

When students click on one of the problems, they obtain information about "what to check for" and "how to repair." Clicking on the associated buttons produces an animated representation of checking and repairing the pump, respectively. Screenshots, showing the use of the repair manual, are displayed in Figure 14. As seen in Figure 14, fixing the initial problem does not completely repair the pump, so students must repeat the steps to solve the remaining problem ("no water is coming out"). When they have fixed the problem, the animation is labeled "The pump handle is now working more easily, the squeaking noise has stopped, and water is flowing freely from the pump" and the dialog box says "Good work! The pump has been completely repaired."

Kumar appears on the next screens, providing information about the next part of the task:

Thanks for helping us repair the well!

Next we need you to help Ramnagar plan for the future.

Currently, Ramnagar's well is the only safe source of drinking water nearby.

How can we make sure that Ramnagar has a reliable source of safe drinking water?

Here are some facts about Ramnagar's water supply.

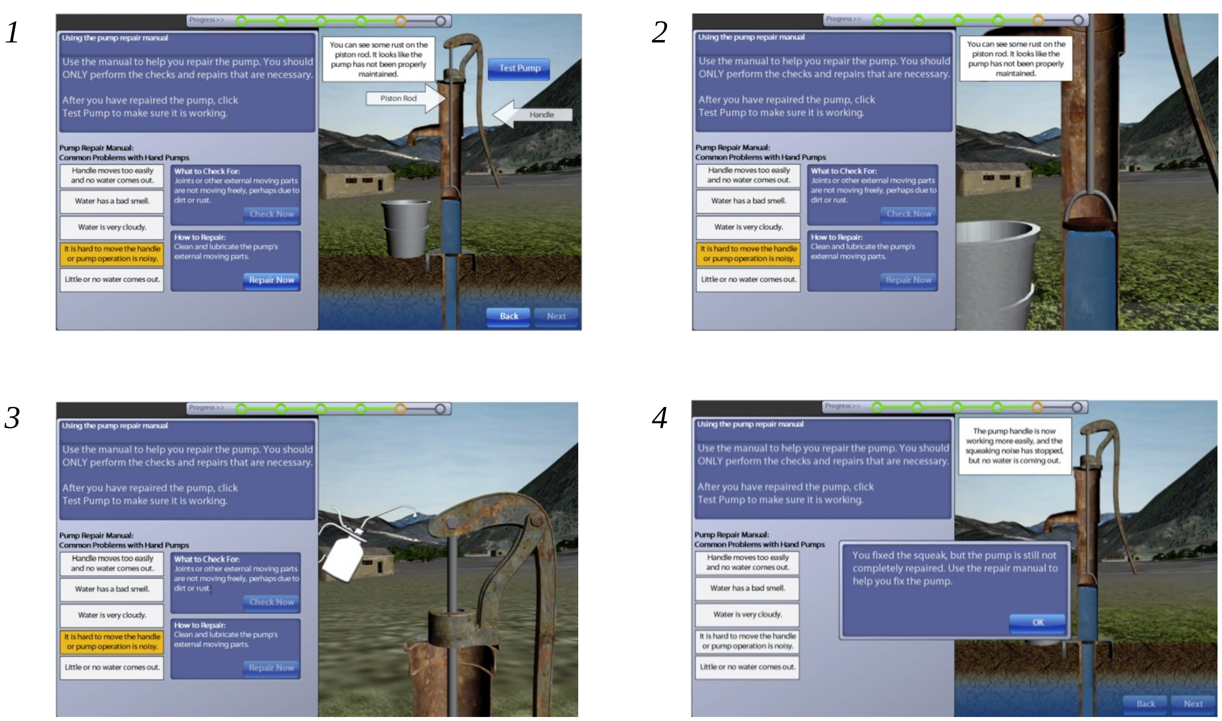


Figure 14. Sample NAEP TEL Task: Series of screen shots of using the repair manual. From "Sample TEL Task," by National Center for Education Statistics, 2013.

When the last instruction is given, students also see the facts shown on the right side of Figure 15. On this screen (Figure 15), Kumar asks, “What is the best way to ensure that Ramnagar has a reliable source of drinking water?” They are asked to select from four possible plans, using the criteria that “the plan should be as inexpensive as possible and make sure that Ramnagar will never be without a working well for more than a day” and constraints given as facts. Finally, students are asked to justify their plan: “Give Kumar two reasons why your suggestion is the best plan for maintaining Ramnagar’s well.” The task ends with a thank you from Kumar: “Thanks for your help! Ramnagar’s well will be a reliable source of drinking water for years to come.”

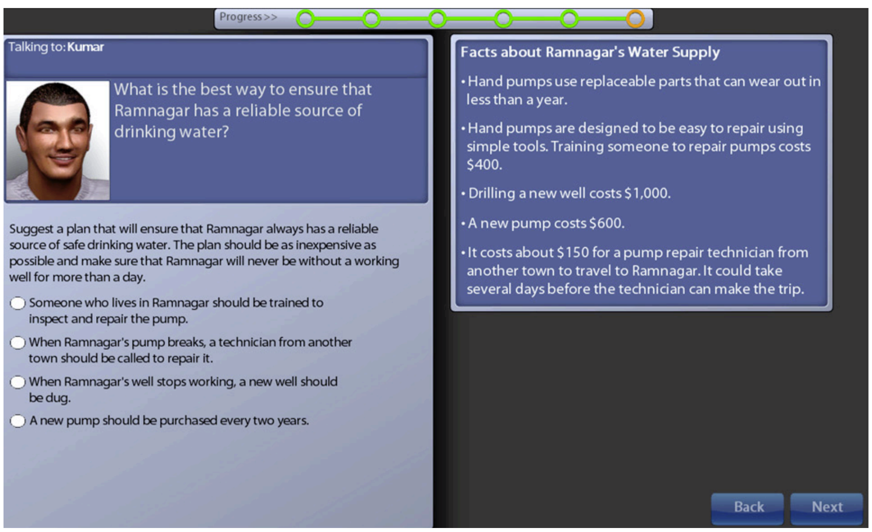


Figure 15. Sample NAEP TEL Task: Screenshot of the part of the task requiring students to select a plan to ensure that Ramnagar will have safe drinking water. From “Sample TEL Task,” by National Center for Education Statistics, 2013.

Considering Table 1, this task reflects the following science practices from the NRC Framework and NGSS:

- Engineering Practice 1: Defining Problems
    - Ask probing questions that seek to refine a problem, including criteria and constraints[6]
- Engineering Practice 2: Developing and Using Models

[6] This practice appears to be partially represented in the NAEP TEL task. Students are selecting questions from a set of choices, rather than asking their own questions. While the questions help to refine the problem (in terms of determining whether the problem is with the aquifer or with the pump), they do not involve much interaction with criteria and constraints.

    - Use models/simulations to test possible solutions
- Engineering Practice 3: Planning and Carrying Out Investigations
    - Gathering data to test designs
- Engineering Practice 6: Designing Solutions
    - Solve engineering problems
    - Balance competing priorities
- Engineering Practice 7: Engaging in Argument From Evidence
    - Consider a range of factors to find the best solution
    - Make arguments from evidence to defend conclusions
- Engineering Practice 8: Obtaining, Evaluating, and Communicating Information
    - Communicate ideas in writing.
    - Derive meaning from scientific texts from the Internet

Thus, overall, this task includes components of a number of engineering practices from the NRC Framework and NGSS. However, consistent with the NAEP TEL framework (NAGB, 2010), this task does not require students to draw upon scientific knowledge and, thus, does not integrate content and practices as required by the NRC Framework and NGSS. It is also important to note that the scoring guide will ultimately determine what students are required to do in order to receive credit for the various parts of the task.

PISA Science

The 2015 PISA Science framework “refines and extends the previous construct” of scientific literacy as defined in the “PISA 2006 framework that was used as the basis for assessment in 2006, 2009, and 2012” (OECD, 2013b, p. 3). In addition, as noted above, all items developed for PISA 2015 will be computer based. For both reasons, tasks included in the 2015 PISA Science assessment are likely to differ substantially from tasks included in previous PISA assessments. However, since the PISA 2015 assessment is still being developed, test security prevents the release of operational items. Therefore, this analysis focuses on abstracted items that could appear in the 2015 Science assessment. Table 3 shows how the abstracted items reflect practices from the NRC Framework and NGSS.

Although the 2015 PISA Science Framework (OECD, 2013b) did not explicitly include engineering practices, some of the abstracted items (see Table 3) that were set in a technological context appeared to align with several of the NRC/NGSS engineering practices, as well as many of the science practices. However, for most of the abstracted items, one could imagine students drawing upon the same procedural knowledge in a number of different, relatively interchangeable contexts. Therefore, it is not clear that these items integrate content and practices to the extent required by the NRC Framework and NGSS.

AP Physics

As the new AP Physics examination is still in the pilot stage, a limited number of items were available for review, and further work is still being conducted (e.g., to flesh out scoring guides). Thus, only a preliminary consideration of the affordances of the new examination is possible here.

“One of the requirements on the new exams,” is for students to “generate a coherent, paragraph length argument” (K. Lionberger, personal communication, August 22, 2013). An example of this requirement is the item in Figure 16 Based on an analysis of the information in Figures 16–18 (the item text, scoring guide, and example response), another item not available for public release, and Table 1, it appears that the items meeting the new requirement may align with the following science practices from the NRC Framework and NGSS:

- Science Practice 2: Developing and Using Models
  - Construct and use models to help develop and/or test explanations
- Science Practice 6: Constructing Explanations
  - Incorporate current understanding of science into explanations
- Science Practice 7: Engaging in Argument From Evidence
  - Engage in reasoning and argument to find the best explanation for natural phenomena individually (S6)
- Science Practice 8: Obtaining, Evaluating, and Communicating Information
  - Communicate ideas in writing

Table 3. Alignment Between Abstracted PISA Science Items and the NRC Framework and NGSS (Table 1)

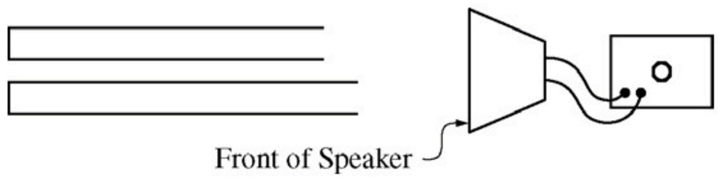
Item	Components of practices reflected in item
[With a simulation that allows the user to try out different machines in a testing device. A standard is provided that calls for balancing, in certain conditions, machine performance with energy use. Students can vary the machine and the condition.] Based on the results of the simulation for the different machines shown in the simulation, which machine meets the standard? • Machine A • Machine B • Machine C • Machine D Select a row of data in the simulation table to support your answer.	Engineering Practice 2: Developing and Using Models • Use models/simulations to test possible solutions Engineering Practice 3: Planning and Carrying Out Investigations • Consider how variables might be controlled[a] • Gathering data to test designs Engineering Practice 4: Analyzing and Interpreting Data • Identify patterns and interpret results to compare different solutions Engineering Practice 6: Designing Solutions • Balance competing priorities • Test a design (E3) Engineering Practice 7: Engaging in Argument From Evidence • Formulate evidence based on test data
[With a simulation] What would be the effect of increasing X on Y and Z? • Result A • Result B • Result C • Result D Select two rows of data to support your answer. Explain how the selected data supports your answer.	Science Practice 2: Developing and Using Models Science Practice 3: Planning and Carrying Out Investigations • Observe and collect data to describe a phenomenon Science Practice 4: Analyzing and Interpreting Data • Use data as evidence Science Practice 6: Constructing Explanations • Link evidence to claims Science Practice 7: Engaging in Argument From Evidence • Formulate evidence based on data Science Practice 8: Obtaining, Evaluating, and Communicating Information • Communicate ideas in writing
[With a simulation] When [conditions are specified], what is the effect of an increase in X on Y? • Y increases • Y decreases Select two rows of data to support your answer. What is the [biological] reason for this effect?	Science Practice 2: Developing and Using Models Science Practice 3: Planning and Carrying Out Investigations • Observe and collect data to describe a phenomenon Science Practice 4: Analyzing and Interpreting Data • Use data as evidence Science Practice 6: Constructing Explanations • Incorporate current understanding of science into explanations • Link evidence to claims Science Practice 7: Engaging in Argument From Evidence

Item	Components of practices reflected in item
	• Formulate evidence based on data Science Practice 8: Obtaining, Evaluating, and Communicating Information • Communicate ideas in writing
[With a simulation] Describe one advantage and one disadvantage [of experimenting with the simulation rather than the actual system].	Science Practice 2: Developing and Using Models • Recognize limitations of models/simulations[a]
[With a simulation] Scientists experiment with an actual system that is modeled by the computer simulation. They find that the results are different in way W from what is predicted by the simulation. What are two possible reasons for this difference?	Science Practice 4: Planning and Carrying Out Investigations • Consider reliability and precision of data Science Practice 7: Engaging in Argument From Evidence • Provide critiques of other's scientific work Science Practice 8: Obtaining, Evaluating, and Communicating Information • Communicate ideas in writing
[With static data] What are some possible sources of uncertainty in the students' data?	Science Practice 8: Obtaining, Evaluating, and Communicating Information • Communicate ideas in writing • Assess the credibility of sources of scientific information
[Following a text (and graphical) stimulus] The student decides to do an Internet search on [topic] and finds the sources listed below. Which source is likely to be reliable? [Options are different sources, such as a blog, an article, and a website, described with key details relevant to evaluating their credibility.] Explain why the source you chose is most likely to be reliable.	Science Practice 8: Obtaining, Evaluating, and Communicating Information • Communicate ideas in writing • Assess the credibility of sources of scientific information

Note. All abstracted items were created by Eric Steinhauer (personal communication, August 2, 2013). If a practice is listed without a component, the item aligns generally with the practice but not specifically with any of the components in Table 1. Consideration of alignment between each item and the science and engineering practices included scoring criteria, where available.

[a] Indicates practice and components that may or may not be elicited by an item, depending on the specific context of an abstracted item and/or on the way that the student chooses to respond to the item.

It is important to note, however, that the extent to which this item reflects the vision in the NRC Framework and NGSS greatly depends on how raters operationalize the criteria in the scoring guides such as “explaining the condition for resonance in a tube closed at one end” (see Figure 17) and “a coherent argument that leads to a correct conclusion” (K. Lionberger, personal communication, August 22, 2013).



(7 points, suggested time about 13 minutes)

The figure above shows two tubes that are identical except for their slightly different lengths. Both tubes have one open end and one closed end. A speaker connected to a variable frequency generator is placed in front of the tubes, as shown. The speaker is set to produce a note of very low frequency and then turned on. The frequency is then slowly increased to produce resonances in the tubes. Students observe that at first only one of the tubes resonates at a time. Later, as the frequency gets very high, there are times when both tubes resonate.

In a clear, coherent, paragraph-length answer, explain why there are some high frequencies, but no low frequencies, at which both tubes resonate. You may include diagrams and/or equations as part of your explanation.

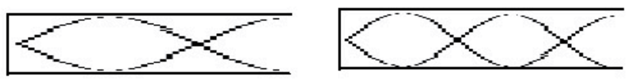
Figure 16. Sample AP Physics Task: Item prompt. Copyright © 2013 The College Board. Reproduced with permission. http://apcentral.collegeboard.com.

2 pts For explaining the condition for resonance in a tube closed at one end
1 pt For comparing wavelengths at low frequency to the tube lengths
1 pt For linking the above two concepts to explain why only one resonance occurs at a time
1 pt For indicating that as frequency goes up, wavelength goes down
2 pts For indicating how smaller wavelengths relate to differences in tube length, explaining how both tubes can now meet boundary conditions

Figure 17. Sample AP Physics Task: Scoring guide. Copyright © 2013 The College Board. Reproduced with permission. http://apcentral.collegeboard.com.

Example:

In order to resonate, the length of a tube must be an odd multiple of a quarter wavelength of the sound, as shown below.



For resonance at low frequencies, the wavelength of the sound is of the order of the length of the tubes. So the match can occur for only one tube at a time – the difference in tube lengths is much smaller than a half wavelength. As the frequency increases, the wavelength decreases and many more wavelengths fit inside a tube. When half the wavelength becomes of the order of the difference in tube lengths, the tubes can contain an odd multiple of quarter wavelengths for the same wavelength at the same time – it will just be a different multiple.

Figure 18. Sample AP Physics Task: Sample response. Copyright © 2013 The College Board. Reproduced with permission. http://apcentral.collegeboard.com.

Implications and Conclusions

In this section, I consider both what might be learned from current large-scale assessments and what further considerations might be required to fully assess the NGSS.

> Comprehensive sets of assessment examples that align completely with the NGSS performance expectations do not exist. Many of the tasks that have been used for classroom assessment, and those found in large-scale state, national, and international tests, focus primarily on science content or on aspects of scientific inquiry separate from content. With few exceptions, such assessments do not integrate core concepts and science practices in the ways intended by the NRC Framework or NGSS. (Pellegrino, 2013, pp. 321–322)

However, as illustrated in the previous section, some innovative approaches being used in large-scale assessment programs may be useful in beginning the work of designing assessments aligned with the NGSS. In this final section, I discuss both what we might learn from existing large-scale assessment programs, as well as where further work is needed.

Affordances of Current Assessments

Innovative tasks and item types are currently being designed and incorporated into large-scale assessment programs. As described above, many of these tasks allow us to elicit important components of the science and engineering practices in the NRC Framework and NGSS. In particular, consistent with their emphasis in the NGSS and the alignment between the NGSS and the assessment frameworks, the sample assessment items appear to elicit many important components of *Modeling*, *Developing*

Explanations/Designing Solutions, and *Argumentation*. The use of computer-based assessments allows students (a) to engage with a wider range of investigations and design tasks and (b) to engage more fully with models/simulations, both as compared to purely paper-and-pencil assessments. Thus, assessment developers for the NGSS have a rich (and growing) body of development work to draw upon in beginning to develop new assessments. Released items from other large-scale assessments may also be useful in communicating to teachers expectations of the NGSS.

Constraints of Current Assessments and Directions for Future Work

While current large-scale assessments elicit some important aspects of the science and engineering practices in the NRC Framework and NGSS, "high-quality science assessments that are consistent with the framework… must target the full range of knowledge and practices described in this report" (NRC, 2012, p. 263). This mandate faces at least two important challenges. First, some aspects of the NRC Framework and NGSS may be difficult to assess in on-demand assessments. Second, for some practices—particularly those associated with engineering design, which have been underrepresented in K-12 settings—we simply may not know enough about how to elicit students' performances. Below, I discuss both challenges and proposals for addressing each.

Limitations of On-Demand Assessments: Considering the Integration of Classroom and Large-Scale Assessment

As acknowledged in the NAEP Science framework, on-demand assessments "[ascertain] what students know and can do in a limited amount of time… and with limited access to resources," such that "important outcomes of science education that are difficult and time-consuming to measure… but valued by scientists, science educators, and the business community" are often "only partially represented" (NAGB, 2008, p. 8) in large-scale assessment frameworks. However, large-scale assessments "ought to be statements about what scientists, educators, policy-makers, and parents want students to become" because "what we choose to assess will end up being the focus of instruction" (Pellegrino, 2013, p. 320). While existing large-scale assessments permit the elicitation of some important aspects of the NGSS, full engagement with the practices of science and engineering cannot be squeezed into the limited time available for on-demand assessments.

On-demand assessments may be constrained in two ways. First, while large-scale assessments can—as discussed above—elicit important components of the NRC Framework and NGSS, time limitations and other concerns result in the assessment of individual components of the practices, rather than the full practice (sum of its components) or the full process of scientific inquiry or engineering design (the sum of the practices). A focus on the components, rather than on full engagement in scientific inquiry or engineering design may send a problematic message to teachers about what is valued, resulting in students' engagement in bits and pieces of scientific and engineering work, without a full sense for or ability to engage fully in inquiry or design. In addition, measuring the components separately may be misleading. Further research is needed to determine the extent to which students' performance of individual components of practices can be used as a proxy for their ability to engage in the full process of scientific inquiry or engineering design. Factors such as persistence may be important for the latter but may not be assessed in tasks that are designed to be completed in a relatively short period of time.

Second, some components of the NRC Framework and NGSS may simply be difficult to assess in a standardized manner in large-scale assessment contexts. These are likely to be the components of the NRC Framework and NGSS that are not well represented in the assessment frameworks (as shown in Table 1). As discussed above, these are often reflective components that are crucial to the processes of inquiry and design, as well as those that require students to interact with others (by discussing ideas and/or collaborating to construct an explanation or design a solution).

Therefore, in order to fully assess students' achievement with respect to the NGSS, it may be particularly important to consider how students' classroom work might be leveraged to inform large-scale assessment of their achievement relative to the NGSS. Although this approach is mentioned in the NRC Framework (NRC, 2012, p. 319), it represents a significant departure from current assessment practices in the United States. However, such an approach is being used elsewhere in the world. A useful example is the National Certificate of Educational Achievement (NCEA), "the main national qualification for secondary school students in New Zealand" (New Zealand Qualifications Authority [NZQA], n.d.a). The NZQA describes this system as follows (NZQA, n.d.a, "How It Works" section):

- Each year, students study a number of courses or subjects.
- In each subject, skills and knowledge are assessed against a number of standards.
- Schools use a range of internal and external assessments to measure how well students meet these standards.
- When a student achieves a standard, they gain a number of credits. Students must achieve a certain number of credits to gain an NCEA certificate.
- There are three levels of NCEA certificate, depending on the difficulty of the standards achieved. In general, students work through levels 1 to 3 in years 11 to 13 at school.
- Students are recognized for high achievement at each level by gaining NCEA with Merit or NCEA with Excellence.

"The mix of assessment varies for each student. It depends on the courses the school offers and the subjects the student chooses to study" (NZQA, n.d.c, p. 1). "In each subject, there will be a maximum of three externally assessed standards," and "across the country in recent years, there has been an almost equal mix of internal and external assessments" (NZQA, n.d.c, p. 2). "Subject and assessment experts decide the most appropriate way to assess the knowledge and skills in the various achievement standards" (NZQA, n.d.d, p. 1). The NZQA recognizes that "often internal assessment is the only way to assess particular skills and knowledge."

The following are examples of achievement standards that are assessed internally (Ministry of Education, 2012):

- Science 1.4 Investigate implications of heat for everyday life
- Science 1.12 Investigate the biological impact of an event on a New Zealand ecosystem
- Chemistry 1.8 Investigate selected chemical reactions
- Physics 1.1 Carry out a practical physics investigation that leads to a linear mathematical relationship, with direction

Figures 19 and 20 contain an internal assessment of achievement standard Science 1.4. This assessment was created based on guidelines provided by the NZQA (2011), which include the following clarification of the standard:

*Investigat*e involves showing awareness of how science is involved in an issue that students encounter in their everyday lives. This requires at least one of the following:

- The collection of primary evidence from an investigation and relating it to the scientific theory relevant to the issue.
- The collection of secondary data and the identification of the scientific theory relevant to the issue under investigation. The issue must involve two different views, positions, perspectives, arguments, explanations, or opinions.

This assessment permits students to engage in a more extended investigation than would be possible with limited time in an on-demand assessment. Although the task itself is somewhat constrained, one might imagine a similar task in which students are required to make more decisions regarding the design of their investigations. In addition, students are allowed to work collaboratively on some parts of the assessment, which is significant since—as noted above—collaboration with actual, as opposed to virtual, peers is not currently included in any of the reviewed large-scale assessment programs. By combining internal and external assessments, the NCEA permits the assessment of standards that would pose significant challenges to large-scale assessment. A similar combination of internal and external assessments may be useful for assessing hard-to-assess performance expectations in the NGSS. However, two important features of the New Zealand context must be considered.

Wellington Girls' College
AS 90943Science 1.4 | Version 1 | August 2012
Credits: Four | Investigate implications of heat for everyday life. | Internal

How does clothing worn in cold weather retain heat?

Achievement Criteria Achievement	Achievement with Merit	Achievement with Excellence
Investigate implications of heat for everyday life.	Investigate, in depth, implications of heat for everyday life.	Investigate, comprehensively, implications of heat for everyday life.

Student Instructions:

Part A: Dressing in Layers

Introduction: To keep warm in cold weather it is important to dress in layers.
In this activity you will investigate the effect of adding layers of fleece fabric to cans of hot water.

What you need:

- Four identical polystyrene cups
- Lids for cups (can be made from ice cream tub lids)
- Hot water from a kettle
- A thermometer
- Fleece fabric
- scissors
- sellotape

What you do:

This stage may be done on your own or in groups of up to three.

1. Leave one cup with no fleece. Cover the other cups with one layer, two layers or three layers of fleece fabric as shown in the diagram.

When the cups are covered they should look like the one in the photo. This cup is covered with two layers of fleece. It has a plastic lid, resting on top.

Put a hole in the lid just big enough for the thermometer.
All the cups should have similar lids to make the test fair.

2. Lift the lid up and pour the same volume of hot water into each of the cups. Measure the water temperature immediately. Write the temperature in the table.

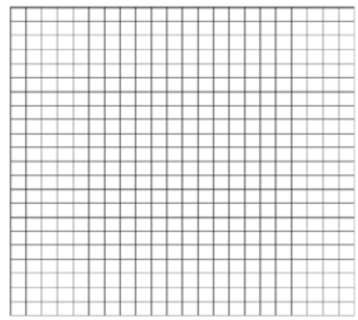
3. Measure the water temperature every five minutes for at least 40 minutes

1

4. Record your temperature readings and the times they were taken in the table below.

	Temperature of water in cup (°C)			
Time (minutes)	no fleece	one layer of fleece	two layers of fleece	three layers of fleece
0				

5. Graph your own or your group's results. Time will go along the horizontal axis.
Plot a new line for each set of results.



2

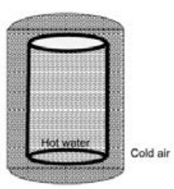
Wellington Girls' College
AS 90943Science 1.4 | Version 1 | August 2012
Investigate implications of heat for everyday life.

Analysis of your results:

You must do this part on your own.

The science behind the experiment:

1. What happened to the temperature of the water in all of the cups?
Make a general statement.



2. Draw an arrow on the diagram to show the direction of heat.

3. Label the materials that the heat travels through.

4. Name the heat transfer process involved.

Describe how the heat is transferred between the materials.

3

Wellington Girls' College
AS 90943Science 1.4 | Version 1 | August 2012
Credits: Four | Investigate implications of heat for everyday life. | Internal

5. The water in which cup cooled the most quickly? Explain why.
Use the data you gathered in your explanation.

4

Figure 19. New Zealand NCEA Sample Internal Task. From "How Does Clothing Worn in Cold Weather Retain Heat?" by Wellington Girls' College, 2012. Reproduced with permission.

Wellington Girls' College
AS 90943Science 1.4 | Version 1 | August 2012
Investigate implications of heat for everyday life.
Credits: Four | Internal

Part B: Can air keep you warm?

Introduction: Some of the best fabrics for keeping us warm are full of air.
In this activity you will investigate the effect of air pockets on heat insulation.

What you need:

- Two identical polystyrene cups, with lids
- Hot water from a kettle
- A thermometer
- Bubble wrap (enough to cover two cups with a double layer of bubble wrap)
- Scissors and sellotape

What you do:

This stage may be done on your own or in groups of up to three.

1.Cut the bubble-wrap in half. Pop all the bubbles in one of the pieces.
(Squash them with your fingers or use a pin)

2. Cover each cup with the same amount of bubble wrap
– one with air-filled bubbles,
- one with popped bubbles.
Remember to make the same lids and bases for both of the cups.
Put a hole in the lid just big enough for the thermometer.

3. Pour the same amount of hot water into both cups.
Measure the temperature in each cup immediately.

Write it in the table on the next page.

4. Measure the water temperature every five minutes for 30 – 40 minutes.
5. Record your temperature readings and the times in the table below.

5

	Temperature of water in cup (°C)	
Time (minutes)	air-filled bubbles	popped bubbles
0		

6. Graph your own or your group's results. Time will go along the horizontal axis.
Plot a new line for each set of results.

6

Wellington Girls' College
AS 90943Science 1.4 | Version 1 | August 2012
Investigate implications of heat for everyday life.
Credits: Four | Internal

Analysis of your results:

You must do this part on your own.

The science ideas:

1. The water in which cup cooled the most quickly? Use the data you gathered in your answer.

2. Write down an explanation for your answer in 1.
Explain how the air pockets affect loss of heat from the cups.

7

Wellington Girls' College
AS 90943Science 1.4 | Version 1 | August 2012
Investigate implications of heat for everyday life.
Credits: Four | Internal

Part C: "How does clothing worn in cold weather retain heat?"

You must do this part on your own.

Fluffy fabrics like fleece or woollen fabric like merino trap air among their fibres.

Discuss how layers of fleece or merino keep you warm.
Use the data you gathered plus the science ideas from both Parts A and B in your answer.

HAND IN ALL YOUR PAGES OF WORK TO YOUR TEACHER WHEN YOU HAVE FINISHED.

(continue on another sheet of paper if you need it) 8

Figure 20. N New Zealand NCEA Sample Internal Task. From "How Does Clothing Worn in Cold Weather Retain Heat?" by Wellington Girls' College, 2012. Reproduced with permission.

First, the NZQA assessments are designed to make relatively coarse determinations of student achievement than is typical in large-scale assessments (whether the focus is on students, schools, states, or countries). There are only four assessment results: not achieving the standard, achieving the standard, achieving the standard with Merit, and achieving the standard with Excellence (NZQA, n.d.a). In contrast, large-scale assessments typically produce scaled scores. Even when these scores are used to determine the percentage of students meeting particular performance expectations, such as Basic, Proficient, and Advanced for NAEP assessments (NCES, 2012c), the scores may also be used to provide numerical scores for individual students, schools, states, and/or countries. Alignment of internal and external assessments may be much more challenging when more precise measures of student achievement are required.

Second, the NZQA notes that "compared with teachers in many other countries, New Zealand teachers are assessment experts" (NZQA, n.d.e, p. 1). This does not occur by accident. Teachers in New Zealand receive significant professional development associated with the country's assessment program. In particular, the quality assurance built into the system includes "external moderation—making sure that teachers are making consistent internal assessment decisions across the country... by providing feedback and professional development" (NZQA, n. d.e, p. 1). "Moderators check each school's assessment tasks and activities, and the judgements schools are making when they assess student work," and "NZQA works with [schools] to improve… internal assessment processes" as needed (NZQA, n. d.e., p. 2). Moderators also "run assessment workshops for teachers," and "in many regions, schools… enable teachers to compare notes with others teaching their subject" (NZQA, n. d.e, p. 2), which serves as an important form of professional development. It is important to note that "all teachers are part of the moderation system. It's a process of being explicit about what [is] expected from students" (NZQA, n. d.e, p. 1).

Being explicit about what is expected could have significant benefits as teachers of science (both elementary teachers who teach science as one of many subjects and secondary teachers who teach only or primarily science courses) adjust to new expectations in the NGSS. Thus, a system that involves all teachers in assessing the NGSS may be useful not only for assessing scientific and engineering practices by also in terms of supporting all teachers in gaining a full understanding of the NGSS. We might think of professional development as being embedded as part of a moderation process, including both external and internal processes. As described above, New Zealand uses an external process for quality assurance, and teachers engage in an internal process to develop shared understandings of the assessment criteria. There is evidence that the latter may serve as sites for teacher learning. In contexts in which teachers are expected to participate in scoring student work, "moderation meetings are a rich and valuable teacher development strategy," perhaps "useful in ways that go well beyond assessment" (Wilson & Sloane, 2000, p. 205). In these meetings,

> Teachers discuss student work and the scores they have given that work, making sure that the scores are being interpreted in the same way by all teachers… [They] discuss the scoring,

> interpretation, and use of student work and make decisions regarding standards of performance and methods for reliably judging student work related to those standards. Moderation sessions also provide the opportunity for teachers to discuss implications of the assessment for their instruction, for example, by discussing ways to address common student mistakes or difficult concepts in their subsequent instruction. (Wilson & Sloane, 2000, p. 201)

A system that involves all teachers in assessing the NGSS may have benefits not only in our ability to assess valued scientific and engineering practices but also in terms of teachers' understanding of the standards.

Implementing a system of assessment such as this would require significantly more trust in teachers who, under current assessment conditions, are treated somewhat skeptically, rather than as partners in determining (a) what students know and can do and (b) how best to support student learning. However, implementation of the NGSS requires significant professional knowledge. As noted in the introduction to this paper, states and other organizations are beginning to consider the teacher professional development that would be required for the NGSS to be fully realized in the nation's K-12 classrooms. "Ultimately, the interactions between teachers and students in individual classrooms are the determining factor in whether students learn science successfully. Thus teachers are the linchpin in any effort to change K-12 science education" (NRC, 2012, p. 255). "Teaching science as envisioned by the framework requires that teachers have a strong understanding of the scientific ideas and practices they are expected to teach" (NRC, 2012, p. 256). In order for NGSS to be implemented as intended, we need to invest in the development of teachers' professional knowledge, such that they are capable both of teaching and assessing student understanding with respect to the new standards.

Limitations of the Current Knowledge Base: Continuing to Research and Develop Innovative Forms of Assessment

As explored above, large-scale assessments such as NAEP and PISA may provide an important basis on which to build additional assessment capabilities aligned with the NGSS. Pellegrino (2013, p. 322) identified an additional affordance of such assessments:

> Neither NAEP nor PISA represent[s] static assessment programs. Both undergo major revisions to the framework used to guide assessment design and task development, and both are increasingly moving to incorporate technology…Changes in both [frameworks] will ostensibly move in directions that even more closely align with the NRC Framework. Thus, both might constitute reasonable ways to monitor overall progress of teaching and learning in U.S. classrooms in ways consistent with implementation of the NRC Framework and NGSS.

Indeed, as NAEP unveils its computer-based TEL assessment in 2014 and PISA moves to a computer-based science assessment in 2015, new item types should become available to inform the

NGSS. Moving forward, collaboration between assessment developers working with assessment frameworks for large-scale assessments such as NAEP, PISA, and the AP program may wish to collaborate, both with each other and with those working on assessments of the NGSS, in order to share insights about designing hard-to-assess areas of the NGSS.

Drawing upon current and future research efforts will also be important for assessing the full range of practices in the NGSS. In particular, because research

> to create and evaluate assessment tasks and situations that can provide adequate evidence of the proficiencies implied in the NGSS… must be carried out in instructional settings where students have had an adequate opportunity to construct the integrated knowledge envisioned by the National Research Council Framework and the NGSS... (Pellegrino, 2013, p. 323)

conducting assessment research in status quo instructional contexts is unlikely to produce the type of innovative assessments that are needed. Indeed, results from the hands-on and interactive computer tasks on the 2009 NAEP science assessment reveal that students struggled to provide scientific explanations, a key practice in the NGSS (NCES, 2012b). As Pellegrino (2013) noted, "several projects have developed assessments for use in classroom instruction with a particular emphasis on the integration of core science concepts with one or more science practices" (p. 322). Thus, research projects could provide not only cutting-edge perspectives on assessment of the NGSS performance expectations but also a population of students who have had the opportunity to engage with the required knowledge and practices.

Summary

In conclusion, although there is much that can be learned from existing large-scale assessment programs to inform assessment of the NGSS, there is also much work to be done. Fully representing the treatment of scientific and engineering practices in the NGSS may require (a) rethinking mechanisms for obtaining data on student achievement, by considering the role of school-based assessments, and (b) partnerships among assessment developers and between assessment developers and researchers.

This task is too challenging for any one state to undertake on its own. Although the NGSS is state led, with each state making an independent choice about adoption, NGSS Lead States (2013) recommends that states adopt the standards without alteration. Therefore, as with the Common Core Standards for English language arts and mathematics, assessment consortia such as Partnership for Assessment of Readiness for College and Careers (www.parcconline.org) and the Smarter Balanced Assessment Consortium (www.smarterbalanced.org) would seem to be the only practical means of assessing the NGSS in ways that do not distort the vision offered by the NRC Framework.

References

Alonzo, A. C., Neidorf, T., & Anderson, C. W. (2012). Using learning progressions to inform large-scale assessment. In A. C. Alonzo & A. W. Gotwals (Eds.), *Learning progressions in science: Current challenges and future directions* (pp. 211–240). Rotterdam, The Netherlands: Sense Publishers.

Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher, 36*, 258–267.

College Board. (2012). *AP Physics 1: Algebra-based and AP Physics 2: Algebra-based Curriculum Framework 2014-2015*. New York, NY: Author.

College Board. (2013). *AP Chemistry: Course and exam description effective fall 2013.* New York, NY: Author.

Hamilton, L. S., & Berends, M. (2006, April). *Instructional practices related to standards and assessments*. Rand Education (Working Paper WR-374-EDU). Retrieved from http://www.rand.org/content/dam/rand/pubs/working_papers/2006/RAND_WR374.pdf

Michigan Department of Education. (2013, April). *Revised timeline for transition to NGSS*. Retrieved from http://www.michigan.gov/documents/mde/Rev_Timeline_for_posting_2-8-13_410943_7.pdf

Ministry of Education. (2012, December). *Science matrix*. Retrieved from http://ncea.tki.org.nz/Resources-for-aligned-standards/Science/Science

Mullis, I. V. S., Martin, M. O., Ruddock, G. J., O'Sullivan, C. Y, & Preuschoff, C. (2009, September). *TIMSS 2011 assessment frameworks*. Boston, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

National Assessment Governing Board (2008, September). *Science framework for the 2009 National Assessment of Educational Progress*. Washington, DC: US Government Printing Office.

National Assessment Governing Board. (2010, May). *Technology and engineering literacy framework for the 2014 National Assessment of Educational Progress (pre-publication edition)*. Retrieved from http://www.edgateway.net/cs/naepsci/view/lib/249

National Center for Education Statistics. (n.d.a). *NAEP science 2009: Grade 8 magnetic fields hands-on task (HOT): Administration and scoring materials.* Retrieved from http://nationsreportcard.gov/science_2009/hot_g8_scoring.aspNational Center for Education Statistics. (2012a, March 31). *What does the NAEP science assessment measure?* Retrieved from http://nces.ed.gov/nationsreportcard/science/whatmeasure.aspx

National Center for Education Statistics. (n.d.b). *The nation's report card: ICT/HOT Grade 8 magnetic fields. Test booklet.* Retrieved from http://nationsreportcard.gov/subject/science_2009/media/Grade8_Magnetic_Fields.pdf

National Center for Education Statistics. (2012b, June). *The Nation's Report Card: Science in action: Hands-on and interactive computer tasks from the 2009 science assessment* (NCES 2012-468). Washington, DC: Author.

National Center for Education Statistics. (2012c, July 12). *How results are reported.* Retrieved from http://nces.ed.gov/nationsreportcard/about/nathowreport.aspx

National Center for Education Statistics. (2012 d, August 3). *Interpreting NAEP science results*. Retrieved from http://nces.ed.gov/nationsreportcard/science/interpret-results.aspx

National Center for Education Statistics. (2012e, September 21). *NAEP overview.* Retrieved from http://nces.ed.gov/nationsreportcard/about/

National Center for Education Statistics. (2013). *Sample TEL task.* Retrieved from http://nces.ed.gov/nationsreportcard/tel/wells_item.aspx

National Research Council. (1996). *National science education standards*. Washington, DC: National Academy Press.

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

National Science Teachers Association. (2013). *NSTA Web Seminars: Next Generation Science Standards Web Seminar Series*. Retrieved from http://learningcenter.nsta.org/products/symposia_seminars/Ngss/webseminar.aspx

New Zealand Qualification Authority. (n.d.a). *How NCEA works*. Retrieved from http://www.nzqa.govt.nz/qualifications-standards/qualifications/ncea/understanding-ncea/how-ncea-works/

New Zealand Qualification Authority. (n.d.b). *Understanding NCEA*. Retrieved from http://www.nzqa.govt.nz/qualifications-standards/qualifications/ncea/understanding-ncea/

New Zealand Qualification Authority. (n.d.c). *Understanding NCEA #7: Internal and external assessment in NCEA*. Retrieved from http://www.nzqa.govt.nz/assets/qualifications-and-standards/qualifications/ncea/Understanding-NCEA/factsheet-7.pdf

New Zealand Qualification Authority. (n.d.d). *Understanding NCEA #8: Why are some achievement standards externally assessed, and some internally assessed?* Retrieved from http://www.nzqa.govt.nz/assets/qualifications-and-standards/qualifications/ncea/Understanding-NCEA/factsheet-8.pdf

New Zealand Qualification Authority. (n.d.e). *Understanding NCEA #10: How does NZQA make sure internal assessment is fair and consistent across the country* Retrieved from http://www.nzqa.govt.nz/assets/qualifications-and-standards/qualifications/ncea/Understanding-NCEA/factsheet-10.pdf

New Zealand Qualification Authority. (2011). *Achievement standards AS90943*. Retrieved from http://www.nzqa.govt.nz/nqfdocs/ncea-resource/achievements/2011/as90943.pdfOrganisation for Economic Co-operation and Development. (2013a, March). *PISA 2015 draft collaborative problem solving framework.* Retrieved from http://www.oecd.org/pisa/pisaproducts/Draft%20PISA%202015%20Collaborative%20Problem%20Solving%20Framework%20.pdf

NGSS Lead States. (2013). *Next Generation Science Standards: For states, by states*. Washington, DC: The National Academies Press.

Organisation for Economic Co-operation and Development. (2013b, March). *PISA 2015 draft science framework*. Retrieved from http://www.oecd.org/pisa/pisaproducts/Draft%20PISA%202015%20Science%20Framework%20.pdf

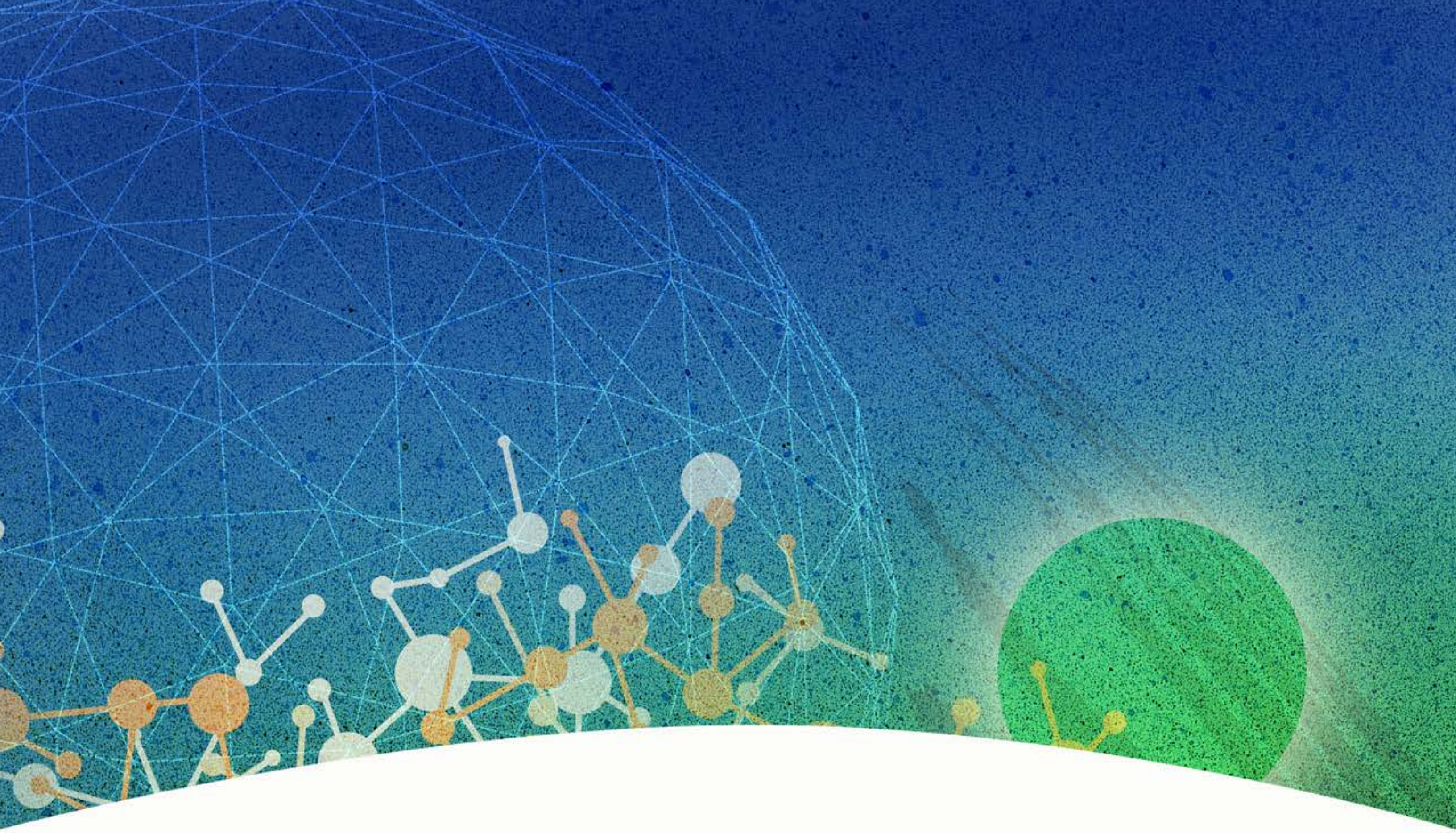
Pellegrino, J. W. (2013, April 19). Proficiency in science: Assessment challenges and opportunity. *Science, 340*, 320–323.

Vermont State Board of Education. (2013, June 25). Item H3: Will the State Board of Education vote to approve the Next Generation Science Standards as Vermont standards? Retrieved from http://education.vermont.gov/documents/EDU-SBE_2013_06_25_Item_H3.pdf#sthash.dhL0NNTb.dpuf

Washington's Regional Science Coordinators. (2013). *Preparing for the Next Generation Science Standards in 2013-2014: Facts and advice for Washington science educators.* Retrieved from http://www.washingtonesds.org/cms/lib4/WA07001775/Centricity/Domain/80/Preparing%20for%20NGSS.pdf

Wellington Girls' College. (2012, August). *How does clothing worn in cold weather retain heat?* Wellington, New Zealand: Author.

Wilson, M., & Sloane, K. (2000). From principles to practice: An embedded assessment system. *Applied Measurement in Education, 13*, 181–208.



The Center for K–12 Assessment & Performance Management at ETS creates timely events where conversations regarding new assessment challenges can take place and publishes and disseminates the best thinking and research on the range of measurement issues facing national, state, and local decision makers.

K-12 Center at ETS

Invitational Research Symposium on
Science Assessment