Journal of Science Teacher Education

ISSN: 1046-560X (Print) 1573-1847 (Online) Journal homepage: http://www.tandfonline.com/loi/uste20

Designing for the Next Generation Science Standards: Educative Curriculum Materials and Measures of Teacher Knowledge

Jo Ellen Roseman, Cari F. Herrmann-Abell & Mary Koppal

JOURNAL OF SCIENCE TEACHER EDUCATION
2017, VOL. 28, NO. 1, 111–141
http://dx.doi.org/10.1080/1046560X.2016.1277598

Routledge
Taylor & Francis Group

Designing for the Next Generation Science Standards: Educative Curriculum Materials and Measures of Teacher Knowledge

Jo Ellen Roseman, Cari F. Herrmann-Abell, and Mary Koppal

American Association for the Advancement of Science, Washington, DC, USA

ABSTRACT

Focusing on the science knowledge and pedagogical content knowledge that teachers need to realize the vision of the Next Generation Science Standards (NGSS) in their classrooms, this article presents findings from efforts to (a) adapt existing criteria and design heuristics for educative features of curriculum materials and apply them in the development of materials to support NGSS and (b) develop an authentic measure of teachers' knowledge and practice based on an analysis of teachers' evaluations of their students' written explanations of phenomena. The study demonstrates that existing criteria and heuristics for designing educative features of curriculum materials can be used productively, with minor modifications, to design features that support teachers in their use of materials that support NGSS. It also provides quantitative and qualitative data to show that teachers' analyses of the explanation task produced useful information about their understanding of the science ideas targeted in the assessment; of the misconceptions their students held; and of their students' ability to reason from evidence, science ideas, and models in explaining phenomena. This study builds on and contributes to a body of work on the design and use of educative curriculum materials and the evaluation of teacher knowledge. It suggests a practical approach to the design of NGSS-aligned curriculum materials that support both student and teacher learning based on findings from analysis and empirical studies.

KEYWORDS
curriculum; educative; measures; standards

With the release of Next Generation Science Standards (NGSS; NGSS Lead States, 2013) and their adoption by 18 states and the District of Columbia, science educators are being asked to transform the way they think about the content they teach and the way that they teach it. Among other changes, NGSS—along with the National Research Council's (NRC; 2012) *A Framework for K–12 Science Education* that preceded it—emphasizes the interconnected nature of science as it is practiced and experienced in the real world, a coherent progression of science concepts from kindergarten through high school, and a deeper understanding of core science ideas and their usefulness in making sense of phenomena and developing solutions to problems.

To implement NGSS, teachers must reconsider the science content that is taught, how students build their understanding of that content, and how ideas fit together to tell a coherent story (Reiser, 2013). Although professional development (PD) can provide some help to teachers for these tasks, limited time and other resources make PD by itself an impractical solution for satisfying the needs of teachers across all grades and science disciplines. According to Achieve, Inc., a partner organization in the development of NGSS, appropriate instructional materials designed to support NGSS will have an important role to play. To help teachers understand the new standards and implement them effectively, says Achieve, Inc. (2015), materials will need to provide an extensive range of supports, from

CONTACT Mary Koppal mkoppal@aaas.org AAAS Project 2061, 1200 New York Avenue NW, Washington, DC 20005, USA.
Supplemental content for this article is available at www.tandfonline.com/uste.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/uste.

suggestions for how to engage students in developing explanations and constructing conceptual models of the natural world to learning progressions that map out students' development of science content knowledge over time. Indeed, because of their widespread use by teachers (Horizon Research, 2013), curriculum materials have enormous potential for leveraging teachers' role as enactors of curriculum to the transfer of research findings, best practices, and reforms such as those proposed by NGSS into the classroom (Ball & Cohen, 1996; Remillard, 2005).

More than a decade ago, Davis and Krajcik (2005) broke new ground by proposing a set of heuristics for designing educative science materials that would support teachers in developing their (a) subject matter knowledge (SMK), (b) pedagogical content knowledge (PCK) for science ideas, and (c) PCK for scientific inquiry (i.e., their ability to engage and guide students in using science practices). Their work focused on the kinds of knowledge that teachers would need to meet the practical challenges of the classroom. Today, in the era of NGSS, teachers face new challenges and need examples of what effective instruction and curriculum might look like (NRC, 2015).

One such example, we believe, is a middle school curriculum unit that is designed to promote student achievement of the NGSS vision while also providing teachers with the educative features they need to implement that vision in their classrooms. Developed by a team of researchers and curriculum developers at the American Association for the Advancement of Science (AAAS) and Biological Science Curriculum Study (BSCS), the *Toward High School Biology* (THSB; AAAS, in press) unit is one of the first materials intentionally designed to realize the NGSS vision of three-dimensional learning.

To date, there has been little exploration of the role that educative curriculum materials can play in promoting the NGSS vision of science learning to a broad audience of teachers. In addition, the research literature on educative materials consistently points to difficulties that many investigators have had in developing appropriate measures for evaluating teacher knowledge and, hence, the effects of educative materials on teachers' knowledge and practice. This article aims to help fill that gap by providing evidence from the development and testing of the THSB unit to answer the following questions: (a) Can the design heuristics proposed by Davis and Krajcik (2005) be adapted and used to inform the design of educative materials for NGSS? (b) Can the same heuristics also guide the development of authentic measures to evaluate teacher knowledge and practice?

Theoretical foundation for educative curriculum materials

Design of the THSB unit as a whole was informed by the general view that curriculum materials can have a powerful impact on what and how science is taught for the benefit of students and teachers alike (Kali, Koppal, Linn, & Roseman, 2008). More specifically, the developers applied curriculum design principles that emerged from research on the coherence, quality, and effectiveness of middle and high school science textbooks (Herrmann-Abell, Koppal, & Roseman, 2016; Kesidou & Roseman, 2002; Roseman, Stern, & Koppal, 2010; Stern & Roseman, 2004) and from the learning research that underpins these principles (Bransford, Brown, & Cocking, 2000; Clements, 2007; NRC, 2007; Pashler et al., 2007).

With the release of the *Framework* and NGSS in 2012 and 2013, the THSB unit was revised to meet the criteria for measuring the alignment and quality of individual lessons and units with respect to NGSS found in the Educators Evaluating the Quality of Instructional Products (EQuIP) NGSS rubric (Achieve, Inc., 2014; Roseman, Fortus, Krajcik, & Reiser, 2015). Design of the educative features of the THSB unit drew on work exploring the potential of curriculum materials to serve as a source of teacher learning (Ball & Cohen, 1996; Davis & Krajcik, 2005; Schneider, Krajcik, & Marx, 2000).

Preservice teachers' use of educative curriculum materials

Given the novice level of preservice teachers, their use of curriculum materials is of particular interest to researchers exploring the role that educative materials might play in preparing teachers for the classroom. For example, Beyer and Davis (2009) examined the use of educative curriculum materials in improving the ability of preservice elementary teachers to critique and adapt science curriculum materials. They focused on educative features designed to support teachers in applying the principles of identifying, interpreting, and working with students' science ideas. Pre- and posttests found that teachers' use of the principles in their analyses was higher when the supports were present compared to when the supports were absent.

But simply providing educative features in materials may not be sufficient for teacher learning. A 2015 study by Land, Tyminski, and Drake found that even when educative features were present in materials, preservice mathematics teachers tended not to read them in educative ways (e.g., their interpretations of the materials were more descriptive than analytic, highlighting aspects of lessons that were fun). Confirming findings from their earlier study (Drake, Land, & Tyminski, 2014) and consistent with the work of Beyer and Davis (2009), Nicol and Crespo (2006), and Schwarz et al. (2009), the authors concluded that preservice teachers need help in making sense of educative curriculum materials and require help in using frameworks for analyzing materials productively.

In-service teachers' use of educative curriculum materials

Of particular relevance to the implementation of NGSS is the role that educative curriculum materials can play in helping teachers who are already in the classroom make changes in their instruction that will support their students' achievement of NGSS performance expectations. Findings have shown that in-service teachers do benefit from the educative features, but not always and not always as expected (McNeill, 2008; McNeill & Krajcik, 2008; Schneider & Krajcik, 2002). Most of these studies were small in size, were largely descriptive in nature, and lacked reliable measures of what teachers learned. In the only study dealing directly with NGSS, Bismack, Arias, Davis, and Palinscar (2015) enhanced a fourth-grade unit to increase support for three science practices (recording observations, making scientific predictions, and making evidence-based claims). Findings indicated that the educative features improved students' observations and predictions but not their evidence-based claims. Teachers' comments suggested that they were uncertain "how to make and engage … students in this practice" (p. 838), and the authors highlighted the need for more extensive research on the design of educative features to support this and other science practices. This study focused explicitly on the science practices dimension in isolation rather than integrating it with the other two NGSS dimensions of core disciplinary ideas and crosscutting concepts.

Measures of teacher knowledge

The field has found it difficult to develop both valid and authentic measures that adequately capture all that teachers must know and be able to do in the classroom. At one end of the spectrum, for example, Garet, Porter, Desimone, Birman, and Yoon (2001) used self-reports on a Likert scale survey to measure teachers' knowledge and skills, and Rowan, Correnti, and Miller (2002) used items from large-scale surveys to measure content knowledge and PCK in mathematics and writing. They found the survey more successful in measuring such knowledge in the context of mathematics than writing and more successful in measuring mathematics content knowledge than mathematics PCK.

At the other end of the spectrum, researchers have used more authentic tasks to measure teacher knowledge, including evaluating teachers' ability to analyze videotapes of teaching or to analyze curriculum materials or other artifacts of instruction. For example, Kersting (2008) used teachers' ability to analyze videotapes of mathematics teaching as a proxy for their teaching knowledge, and Santagata, Zannoni, and Stigler (2007) used a video-based program on lesson analysis in a preservice program for mathematics teachers. They measured preservice teachers' ability to analyze a new lesson using the lesson analysis framework taught in the program. Schwarz, Gunckel, Smith, Covitt, Bae et al. (2008) explored the use of curriculum analysis—specifically Project 2061's curriculum materials analysis criteria (AAAS, 2005)—in preservice elementary teacher education. Teachers were given six activities about light and shadows and asked to (a) assess strengths and weaknesses of each, (b) write a lesson sequence that briefly described the activities they would use/adapt, and (c) provide a rationale for their sequence. Pre-and posttest results showed an increase in the use of criteria related to content alignment and attention to some of the instructional criteria, but the study also found that teachers had their own criteria (e.g., making science fun) and were generally negative about the use of the Project 2061 criteria.

Other studies have used teachers' ability to attend to their students' thinking as a measure of their knowledge. Jacobs, Lamb, and Philipp (2010) developed a framework that included (a) attending to children's strategies, (b) interpreting children's mathematical understandings, and (c) deciding how to respond on the basis of children's understanding. Teacher knowledge was measured by having teachers observe a video excerpt of an elementary mathematics classroom and prepare a written report that responded to prompts for the three categories in the framework.

Overview of the THSB unit

Developed by a team of scientists, education researchers, and curriculum developers at AAAS and BSCS, the THSB curriculum intervention is an 8-week replacement unit for eighth-grade science classes. It includes (a) a print student edition (SE) workbook, (b) a print teacher edition (TE), (c) Web-based teacher resources, and (d) a 3-day introductory face-to-face PD workshop. The team used an iterative design process that involved the initial conceptualization and writing of the unit; multiple rounds of testing the unit in classrooms; and multiple revisions of the unit based on student performance data, curriculum analysis, and feedback from teachers. Over the course of the 5-year development project, more than 1,000 students and 12 teachers were involved in various phases of the design and testing of the curriculum intervention.

The overarching goal of the THSB unit is to help students make sense of phenomena related to plant and animal growth using NGSS core ideas about chemical reactions in nonliving and living systems; the crosscutting concept of matter conservation across physical and life science; and science practices of data analysis, modeling, explanation, and communication.

The THSB unit consists of 19 lessons organized into four chapters. Chapter 1 develops the idea that substances with different properties form during chemical reactions because the atoms that make up molecules of the starting substances rearrange to form the molecules of the ending substances. Chapter 2 develops the idea that the total mass stays the same during chemical reactions because the number of each type of atom stays the same and the mass of each atom stays the same regardless of differences in their arrangement. Chapter 3 applies ideas of atom rearrangement and conservation to plant growth and repair, and Chapter 4 applies these ideas to animal growth and repair. Table 1 shows how the phenomena that students experience in each chapter relate to the chapter's conceptual focus.

Research question 1: Can the design heuristics proposed by Davis and Krajcik (2005) be adapted and used to inform the design of educative materials for NGSS?

Methods

Elsewhere, we and our colleagues have described initial efforts of the THSB development team to design and test the promise of educative materials in supporting teachers in their use of the THSB unit (Flanagan, Herrmann-Abell, & Roseman, 2013; Kruse, Howes, Carlson, Roth, & Bourdelat-Parks, 2013). The team drew on theories of teacher learning and the role of educative materials articulated by Ball and Cohen (1996) and Davis and Krajcik (2005) and on AAAS's (2005) own criteria for evaluating the quality of curriculum materials, including the support that materials provide for teachers. We then distilled from these sources an initial set of principles that was THSB-specific and focused on the design of supports to help teachers develop their (a) knowledge of the content and practices that make up THSB's storyline at the unit, chapter, and lesson levels; (b) knowledge of the pedagogical purposes of each component of the unit

Table 1. Instructional phenomena in the Toward High School Biology unit.

Conceptual focus	What students experience
1. New substances form during chemical reactions because atoms rearrange to form new molecules	Observing, modeling, and explaining how different ending substances can form from starting substances when • Iron is exposed to air • Vinegar is mixed with baking soda • Hexamethylenediamine is mixed with adipic acid
2. Mass is conserved in chemical reactions because atoms are conserved	Observing, modeling, and explaining how the measured mass of a system can change even though atoms are not created or destroyed (same reactions used in Chapter 1)
3. Plant growth involves chemical reactions in which atoms rearrange and are conserved	Observing, modeling, and explaining how plants produce carbohydrates for growth from substances in their environment and increase in mass without violating conservation principles when • Algae produce $^{14}C_6H_{12}O_6$ from $^{14}CO_2$ • Algae produce $^{18}O_2$ not $C_6H_{12}{}^{18}O_6$ from $H_2{}^{18}O$ • Cress plants make more ^{14}C-cellulose without herbicide than with it
4. Animal growth involves chemical reactions in which atoms rearrange and are conserved	Observing, modeling, and explaining how animals produce proteins for growth that are different from what they eat and increase in mass without violating conservation principles when • A snake eats only eggs but can replace its shed skin • Humans eat muscles but can make tendons • Herring eat ^{14}C-labeled brine shrimp and make ^{14}C-labeled body structures

(e.g., each phenomenon, model, and activity); (c) knowledge of teaching strategies that help to make the unit storyline visible to students and that help teachers elicit, listen to, and respond to students' ideas and questions; and (d) ability to take an analytical approach toward planning and implementing the unit and monitoring their students' progress (Kruse et al., 2013). All four components of the THSB unit—the SE, TE, online resources, and PD—were designed to contribute in different ways to teachers' understanding of and skill in using the unit. For the purposes of the study reported on here, however, we focus mainly on the educative features of the SE, TE, and online resources.

Following the release of the NRC *Framework* and NGSS in 2012 and 2013, the development team improved the unit's alignment with the three-dimensional vision of the new standards and considered what additional supports teachers would need to understand and implement it. In doing so, the team focused on aspects of NGSS that were central to the THSB unit, were doable within 8 weeks of instructional time, and would illustrate what it means to use ideas and practices to make sense of phenomena.

To take into account the implications of NGSS for the design of the THSB unit's educative features, the development team first identified critical aspects of NGSS that were not addressed explicitly in the design heuristics proposed by Davis and Krajcik (2005) but were essential for teachers to understand in order to help their students achieve the goals of NGSS. Given the broad scope of the changes called for in NGSS and the busy lives of teachers, the development team tried first to ensure that the "'base' curriculum materials were accurate, complete, and coherent in terms of content and effective in terms of pedagogy—with good representations of the content, a clear purpose for learning it, and multiple opportunities for students to explain their ideas" (Davis & Krajcik, 2005, p. 3). In doing so, the team confronted what Davis and Krajcik referred to as "tensions in determining an appropriate amount of guidance and prescription" (p. 9). Our own observations during pilot testing had indicated that several teachers used only the SE while teaching, so we opted to incorporate a good deal of instructional support for integrating core disciplinary ideas, science practices, and crosscutting concepts directly into the SE. Additional educative features—particularly support for science practices—were then developed and used by teachers during field tests of the unit and revised based on their feedback during debriefing sessions.

Adapting the design heuristic for SMK

Davis and Krajcik's (2005) design heuristic for SMK calls for materials to, among other things, support teachers in "developing factual and conceptual knowledge of science content … at a level beyond the level of understanding required by the students" (p. 12). It also calls for materials to help teachers see how the "scientific ideas relate to real-world phenomena and to the activities in the unit …" (p. 12). We interpreted this heuristic to include knowledge of NGSS disciplinary core ideas and crosscutting concepts and how they are organized into a coherent story, which is consistent with best practices gleaned from an analysis of science lessons from five countries in the Third International Mathematics and Science Study (TIMSS) Video Study (Roth et al., 2006). To design specific lessons and activities for a curriculum unit situated within a specific context (i.e., developing a molecular explanation for growth and repair in living things), it was necessary to extract from the disciplinary core ideas and crosscutting concepts a set of contextualized science ideas that would provide a coherent content storyline for the THSB unit.

Organizing SMK around a coherent content storyline is critical for both students and teachers so that both come to see the big picture that encompasses the science ideas and the contribution of each activity to its development. Figure 1 presents a map of the THSB science ideas showing how they were sequenced, starting with disciplinary core ideas and culminating in ideas at the top of the map that draw on both disciplinary core ideas and crosscutting concepts. Also included in this SMK heuristic is information on the contribution of each lesson to the development of the content storyline. Each numbered box in Figure 1 identifies not only the relevant NGSS code but also the THSB lesson(s) in which the science idea is targeted.

In designing curriculum for NGSS, it was also important to consider what teachers would need to know about the role of crosscutting concepts (in this case the idea that atom conservation explains mass conservation) as science ideas themselves and as a distinct dimension of science learning that serves to unite core ideas across the disciplines to explain phenomena and answer questions (NGSS Lead States, 2013). Although the power of certain ideas that cut across disciplines was recognized in *Science for All Americans*, which describes such ideas as "tools for thinking about how the world works" (AAAS, 1989, p. 19), and in the NRC (2012) *Framework*, which describes crosscutting concepts as providing "an organizational framework for connecting knowledge" (p. 83), neither provides guidance about how connections should be made or how to support teachers in guiding their students to make them. In developing the THSB unit, we interpreted crosscutting concepts as tools for thinking across a wider range of phenomena than could be explained with particular core ideas. Additional supports would be needed to help teachers appreciate this explanatory value of the crosscutting concepts.

Adapting design heuristics for PCK for science topics

The heuristics proposed by Davis and Krajcik (2005) for this category focus on supporting teachers in engaging their students with phenomena, in using scientific representations, and in addressing students' ideas, all of which are relevant to realizing the three-dimensional vision of NGSS.

Phenomena. Making sense of phenomena is at the heart of NGSS, and students are expected to engage in science practices and use disciplinary core ideas and crosscutting concepts to explain them. NGSS also emphasizes the interconnectedness of science with the goal of helping students see how a set of core ideas can be used to explain different phenomena across multiple science disciplines (NRC, 2012). With this in mind, we elaborated on the Davis and Krajcik (2005) design heuristics to include educative features that provide teachers with a rationale for the inclusion of the chosen phenomena along with support in carrying out activities designed to help students see the explanatory power of science ideas, how they fit together, and how they can apply to a range of physical and life science phenomena.

To capture the integrated learning across disciplines that is called for in NGSS, we distinguished two categories of phenomena: those that have the potential to be generative of the targeted science ideas and those that are not likely to be generative of the ideas but can be explained by them once the

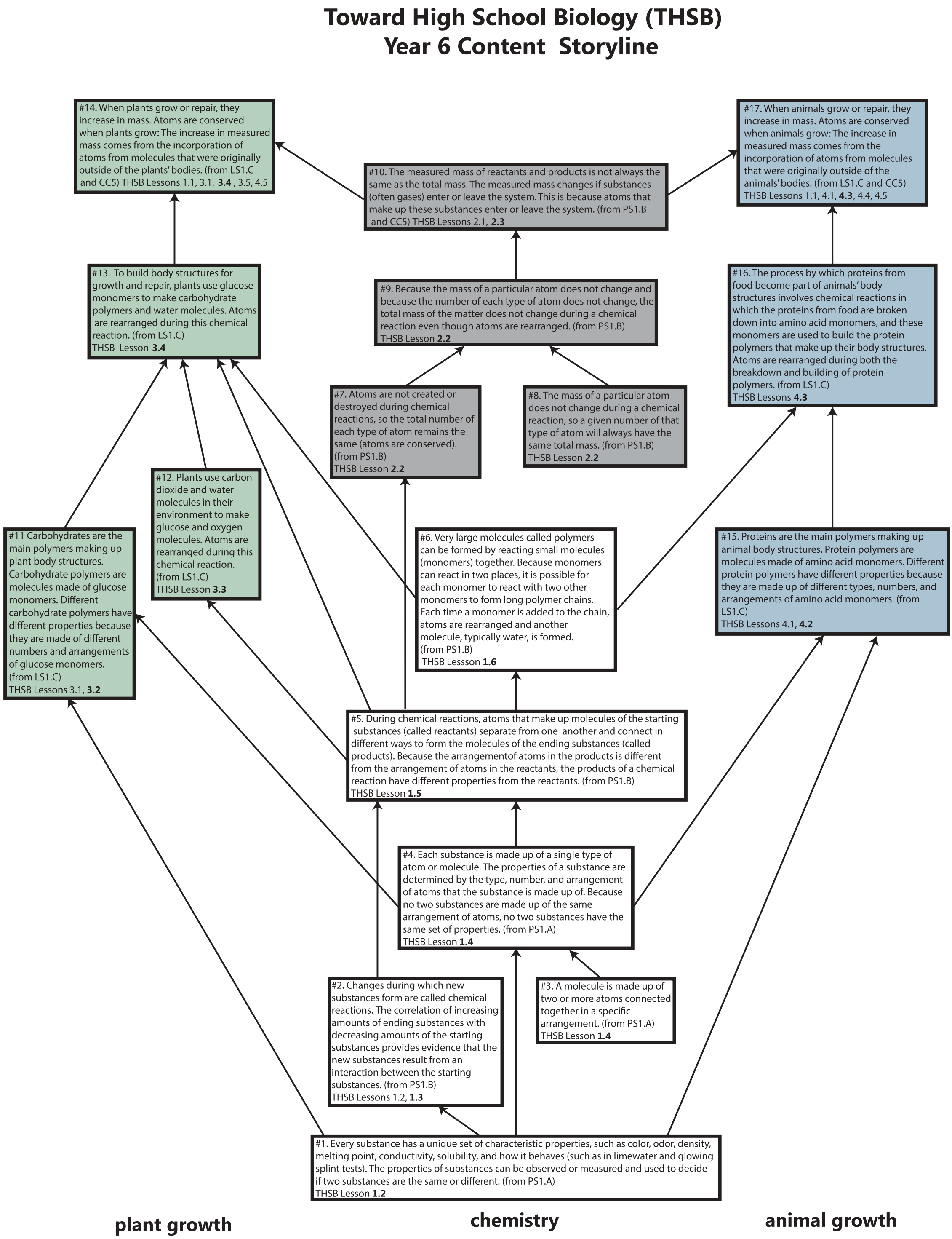

Figure 1. The Content Storyline Map displays the sequence and interconnections of ideas targeted in the THSB unit. THSB = *Toward High School Biology*; CC = crosscutting concepts in NGSS; LS = life sciences core disciplinary idea in NGSS; PS = physical sciences core disciplinary idea in NGSS.

ideas are understood. For example, physical science phenomena that can be carried out with pure substances are more likely to be generative of the ideas that new substances can be produced from starting substances (which can be directly observed) and that the molecules making up the new substances can be made by rearranging the atoms making up the molecules of starting substances (which can be modeled). The need for phenomena to be generative was considered in our decision to include chemical reactions in nonliving and living systems in the same unit and to start with chemical reactions in nonliving systems, and we wanted teachers to understand the rationale for selecting and sequencing these phenomena.

Representations. This Davis and Krajcik (2005) design heuristic focuses on helping teachers select and adapt appropriate representations for their students and identifying features of a representation that are the most salient for instruction. Educative curriculum materials can also support teachers in becoming aware of and making use of a wide variety of representations—illustrations, tables and graphs, diagrams, physical models, and simulations—that can help to make phenomena accessible to a wide range of students (Kesidou & Roseman, 2002; Roseman et al., 2010). We found support for this adaptation of the heuristic in the EQuIP rubric (see Supplemental Material [Online Resource (OR) 1]), which recognizes the role that representations, among other things, can play in NGSS-focused materials as tools for differentiating instruction for students who struggle as well as for those who are more advanced (Achieve, Inc., 2016).

Students' ideas. Educative curriculum materials can help teachers understand, identify, and deal with intuitive and non-normative science ideas that many of their students bring to the classroom (Driver, Squires, Rushworth, & Wood-Robinson, 1994). Indeed, teachers' awareness of students' misconceptions has been shown to correlate with student learning gains (Sadler, Sonnert, Coyle, Cook-Smith, & Miller, 2013). Studies have shown that teachers typically do not focus on their students' ideas (Sherin & van Es, 2005), so materials and accompanying PD need to help them do so. Again, we drew on the EQuIP rubric (see Supplemental Material [OR 1]) for NGSS-specific guidance. The rubric has eight criteria in three categories that include the need for instructional supports to help teachers identify students' ideas and for assessment rubrics and scoring guidelines to help teachers plan instruction and give feedback to their students.

Adapting design heuristics for PCK for scientific inquiry

The NRC *Framework* and NGSS clarify scientific inquiry with a set of eight science practices that relate closely to the design heuristics for PCK for scientific inquiry in Davis and Krajcik (2005). Though practices are listed separately, NGSS makes clear that they work together to help students make sense of phenomena and that reflecting on their use can help students understand how the practices contribute to the development of scientific knowledge. The THSB unit focuses intensely on three science practices—data analysis, modeling, and explanation. Each is developed in the context of making sense of specific phenomena in nonliving and living systems in terms of ideas about atom rearrangement and conservation during chemical reactions.

In adapting this design heuristic for use with materials that support NGSS, we tried to address the knowledge that teachers would need in order to help foster their students' integrated science learning across the three dimensions of disciplinary core ideas, science practices, and crosscutting concepts. Alignment with the NGSS vision would, at the least, require that (a) students engage in explaining phenomena; (b) students' explanations would drive their learning; and (c) their explanations would require the use of all three dimensions—disciplinary core ideas, science practices, and crosscutting concepts. To achieve this integration, teachers would need activities that were designed explicitly for this purpose and additional supports to help them see how the three dimensions work together to deepen students' understanding.

Using the EQuIP rubric to guide three-dimensional design

In 2014 the first version of the EQuIP rubric was released by Achieve, Inc., to provide criteria for analyzing the extent to which lessons and units are designed for NGSS (Achieve, Inc., 2014). The development team used these criteria and the revised versions of the criteria released in subsequent years (Achieve, Inc., 2016) as an additional check on the extent to which the THSB unit's educative features were likely to support teachers in achieving NGSS's vision of three-dimensional learning.

Results

Based on the adapted heuristics described previously, we designed the THSB unit itself, along with specific features just for teachers, to educate teachers about the unit and how to implement it in their classrooms. The design of the unit itself was intended to build teachers' awareness of NGSS's vision for three-dimensional learning and how the THSB unit supports that vision.

Educative features in THSB to promote teachers' SMK

THSB teacher materials include several components to clarify the ideas targeted, how they are organized into a coherent content storyline, and the contribution of specific lessons and activities to student learning. The TE unit introduction lists the specific NGSS disciplinary core ideas, crosscutting concepts, and science practices that are targeted (see Supplemental Material [pp. iii–v in OR 2]); shows how they are organized into a coherent storyline (Supplemental Material [p. viii in OR 2]); and provides a rationale for the inclusion and exclusion of core ideas, crosscutting concepts, and vocabulary (see Supplemental Material [pp. vii–xi in OR 2]).

To deepen teachers' understanding of the relevant content, each chapter in the TE includes a Background Knowledge for Teachers section that provides essential information about the science ideas targeted in the upcoming lessons. The Chapter Overviews (see Supplemental Material [OR 3] for an example) and Lesson Guides (see Supplemental Material [OR 4] for an example) describe the contribution of each lesson and activity to the progression of the content storyline, including how the crosscutting concept is developed in each chapter.

Educative features in THSB to promote teachers' PCK for science topics

Phenomena. Chapters 1 and 2 of the THSB SE start with a set of specific phenomena that are generative of each of the science ideas targeted, and the TE helps teachers understand the role that each phenomenon plays in advancing the content storyline. The Unit Overview describes how observations and data are used to provide evidence for the science ideas and/or their explanatory power (see Supplemental Material [p. vi in OR 2]); the Chapter Overview describes the rationale for the specific phenomena included (see Supplemental Material [pp. 1b–1d in OR 3]); and each Lesson Overview describes the purpose of each activity and the intended observations students should make, either through direct observations or inferences from data (see Supplemental Material [pp. 44c–44d in OR 4 and p. 60c in OR 5]).

Representations. The THSB unit uses a variety of representations to make atoms, molecules, and chemical reactions concrete to students, including conventional models (e.g., space-filling and ball-and-stick models) and unconventional models (e.g., LEGO® bricks and flattened two-dimensional versions of ball-and-stick models). Chemical reactions are represented by chemical reaction mats, as shown in Figure 2, which display molecular models of starting and ending substances. Modeling activities help students see that it is possible to build the product molecules just by rearranging the atoms of the reactant molecules. Students also see how it is possible for measured mass to increase even though total mass remains the same, an observation they will use when explaining why the growth of mushrooms on a fallen dead tree does not violate conservation principles.

The TE Unit Overview (see Supplemental Material [pp. iii–vii in OR 2]) supports teachers in understanding the role of models in the THSB unit and the intended observations students should make when using them. To help ensure that the modeling activities are used to explain the phenomena in

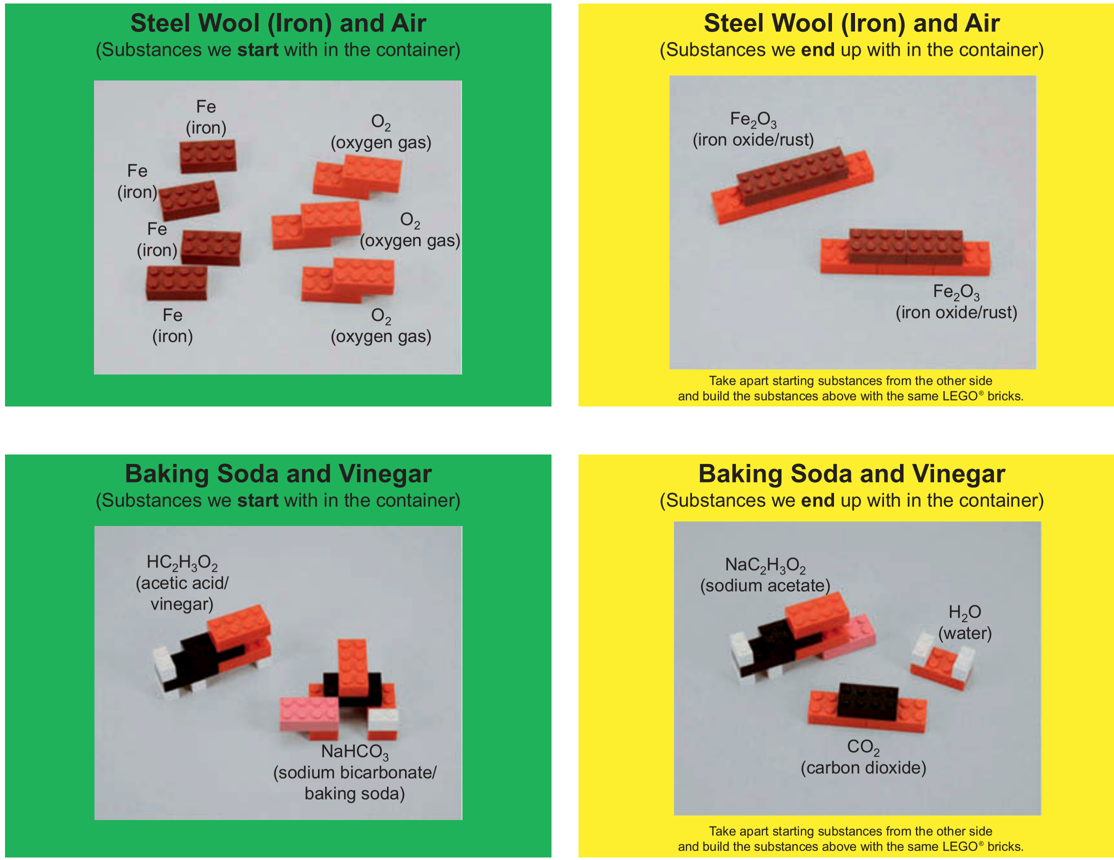

Figure 2. Examples of chemical reaction mats.

Phenomena, Data, or Models	Intended Observations	Purpose	Rationale and/or Notes
Activity 1 Space-filling models representing a system of hydrogen gas and oxygen gas molecules chemically reacting to form water molecules (and leftover reactant molecules)	• When representing chemical reactions with models, some of the reactants may be used up, and some may be left over. • The number of molecules is not necessarily conserved during chemical reactions. • The number of each type of atom is conserved, hence the total mass of those atoms, during chemical reactions.	Students explain why the total mass is conserved during chemical reactions. An ideal explanation should include the ideas that (a) atoms aren't created or destroyed, though the number of molecules may change, and (b) the mass of a particular atom does not change, so the same number of that type of atom will always have the same mass.	The activity challenges the common idea that conservation of matter applies to both conserving atoms and conserving molecules. The activity provides students with the first opportunity to examine a "system" of reactants and products in which some of the reactants may be used up and some may be left over.
Activity 2 LEGO® models of chemical reactions in opened containers (systems): • Baking soda and vinegar • Steel wool (iron) and air (oxygen)	• In a sealed container, the number of each type of LEGO® brick is the same before and after the chemical reaction. • In a sealed container, the total mass of the LEGO® bricks is the same before and after the chemical reaction. • When LEGO® models of product molecules leave an opened container, the number of atoms decreases and the measured mass decreases. • When LEGO® models of reactant molecules enter an opened container and form new products, the number of atoms increases and the measured mass increases.	Students explain decreases/ increases in the measured mass when a chemical reaction occurs in an open system using atomic-level ideas. An ideal explanation includes the ideas that (a) atoms aren't created or destroyed; (b) the mass of a particular atom does not change, so the same number or type of atom will always have the same mass; (c) therefore total mass is conserved, and (d) if the measured mass of a system changes, it is because atoms have entered or left the system.	The activity provides students with another opportunity to examine a "system" of reactants and products in which some of the reactants may be used up and some may be left over. In the case of rusting, it challenges the common idea that rusted iron weighs less than un-rusted iron.
Pulling It Together The formation of patina on the Statue of Liberty, specifically the production of a layer of copper carbonate on the copper exterior of the statue.	(Evidence provided to students) A layer of copper carbonate forms on the statue when copper interacts with oxygen, carbon dioxide, and water vapor in the air.	Students predict the change in measured mass in an open system after the chemical reaction occurs and support their prediction using science ideas about conservation.	The context challenges a common idea that oxidation (e.g., rusting, tarnishing, patina) causes decreases in mass. Patina provides an analogy for photosynthesis in plants, helping students understand that invisible reactants can yield solid products, thereby leading to an increase in the mass of an open system.

Figure 3. Lesson Guide chart. © American Association for the Advancement of Science, Toward High School Biology: Summer 2016.

terms of the science ideas, the Lesson Guides describe the intended observations students should make, as shown in Figure 3 from a lesson targeting ideas about conservation of mass and measured mass.

Students' ideas. The THSB unit provides a variety of opportunities for students to express their initial ideas, engage in activities that build on or challenge them, contrast their ideas with scientific ideas, and consider how their ideas have changed after they have experienced the activities. For example, after interpreting data from radioactive labeling experiments that provide evidence that plants use the carbon and oxygen atoms from carbon dioxide to make glucose and use glucose to make cellulose, students examine Van Helmont's willow tree data, critique his (invalid) conclusion that most of the increased mass comes from water, and then construct a valid explanation based on the data they have examined (see Supplemental Material [pp. 136–138 in OR 6]).

The THSB unit supports teachers in identifying, building on, challenging, and monitoring students' ideas throughout the unit. Every TE Chapter Overview alerts teachers to commonly held student ideas, their manifestations in student work, and the role of specific activities in addressing them (see Supplemental Material [pp. 1d–1e in OR 3]). In addition, the THSB unit includes embedded assessment tasks at the end of each chapter (see Table 2), and online resources provide rubrics and scoring guidelines for each task along with examples of how misconceptions might be expressed in student explanations (see Supplemental Material [pp. 2–3 and 5–6 in OR 9]).

Educative features in THSB to promote teachers' PCK for scientific inquiry

Analyzing and interpreting data. The THSB unit provides many opportunities for students to collect, analyze, and interpret data. Students determine properties of substances through direct observation (e.g., color, state of matter at room temperature, electrical conductivity, water solubility, ability to form fibers, behavior in splint and limewater tests), identify patterns in their observations

Table 2. Embedded assessments in the *Toward High School Biology* unit.

Chapter	Focus	Embedded assessment
1	Atom rearrangement explains the production of new substances during chemical reactions	a. Hydrogen peroxide is a clear, colorless liquid. When you apply it to a cut or scrape on your skin, you will see bubbles and you may hear a "fizzing" sound. Figure 1.3 uses LEGO® models to represent hydrogen peroxide on the left. On the right are models showing what happens to hydrogen peroxide when you apply it to a cut or scrape on your skin. Do you think a chemical reaction occurs when hydrogen peroxide is put on a wound? Explain why you do or do not think so in terms of atoms and molecules. Your explanation should use science ideas, evidence, and models to support your claim. (Use the table for your notes.) b. As a pot of water is heated on a stove, bubbles of gas form in the liquid. Figure 1.4 uses LEGO® models to represent the molecules of water in the pot on the left. On the right are models of the molecules inside the bubbles that form. Do you think a chemical reaction occurs when water is heated on a stove? Explain why you do or do not think so, using science ideas, evidence, and models to support your claim. (Use the table for your notes.)
2	Atom conservation explains mass conservation and changes in measured mass during chemical reactions in open systems	The Statue of Liberty is made up of copper (Cu atoms). But the statue doesn't have the shiny, orange-brown appearance of copper. Instead, it is green. Why? After being exposed for many years to oxygen, carbon dioxide, and water vapor in the air, a thin layer of green copper carbonate ($CuCO_3$ molecules) formed on the copper statue. Answer the following questions and be sure to use ideas about atoms in your explanation, and to write an explanation that meets the Explanation Quality Criteria. (a) Is the change a chemical reaction? Explain. (b) Do you think the Statue of Liberty has less mass, the same mass, or more mass now than when it was first made? Explain.
3	Atom rearrangement and conservation explains the production of new substances for building plant body structures	The paper on the next page describes an experiment carried out by a scientist named J. B. Van Helmont that was published nearly four centuries ago (1662). (a) Read the paper, underline Dr. Van Helmont's conclusion, and circle his summary of his evidence. (b) Does Van Helmont's evidence support his conclusion? Why or why not? Does his explanation meet the elements of the Explanation Quality Criteria? (c) Write a better explanation for where most of the mass of a dry willow tree comes from.
4	Atom rearrangement and conservation explains the production of new substances for building animal body structures	When you get a cut on your hand, your body builds scar tissue to seal up the cut. Scar tissue is made up of the protein *collagen*. Explain how this can happen. Be sure your explanation meets the Explanation Quality Criteria.

(e.g., correlations between increasing amounts of ending substances and decreasing amounts of starting substances), and supplement direct observations with information in data tables from published scientific studies (e.g., data from radioactive labeling experiments). Students use their observations and the data as evidence that new substances were produced, that the ending substances were actually produced from the starting substances, and that atoms making up the molecules of starting substances were incorporated into the molecules of the ending substances (e.g., algae incorporate ^{14}C atoms from carbon dioxide into glucose, cress plants incorporate ^{14}C atoms from glucose into cellulose [see Supplemental Material (OR 8)], young herring fish incorporate ^{14}C atoms from brine shrimp proteins into their own body proteins). Carefully sequenced questions guide students in interpreting data and in using their interpretations as evidence to support claims. The TE provides ideal student responses to those questions (see Supplemental Material [p. 119 of OR 8]) and includes in each Lesson Guide a chart summarizing activities in which students analyze and interpret data to provide evidence for phenomena; clarifying the purpose of each activity; and listing intended observations, inferences from data, and conclusions that students should make (see Figure 3). The Closure notes for teachers at the end of each lesson provide the main points that should emerge from a class discussion about the data analysis tasks (see Supplemental Material [p. 125a of OR 8]).

Modeling. The THSB unit engages students in modeling activities that help them make sense of observations and data, namely, that molecules making up the ending substances can be made by rearranging the atoms making up molecules of the starting substances. For example, students use ball-and-stick models to show how a fibrous solid (made up of nylon polymers) might be made from two clear colorless liquids (each made up of a single monomer) and how algae might incorporate ^{14}C atoms from carbon dioxide into glucose, cress plants might incorporate ^{14}C atoms from glucose into cellulose, and young herring fish might incorporate ^{14}C atoms from brine shrimp proteins into their own body proteins. As with data analysis, carefully sequenced questions guide students in observing and interpreting their modeling activities.

The teacher resources support teachers in helping students carry out the modeling activities and observe, interpret, and draw the intended conclusions. Online how-to videos of the more complicated modeling activities help teachers become more familiar with handling the models and provide examples of how the models should be used during the activities. The TE includes notes on how to encourage students to use models to make sense of novel phenomena and how students should use ideas from modeling activities to support their explanations.

Constructing explanations. The THSB unit engages students in constructing scientific explanations of phenomena, which is one of the key scientific practices in NGSS. Because of the complexity of this practice and the difficulties students have with it, THSB provides support consistent with the cognitive apprenticeship model (Brown, Collins, & Duguid, 1989; Collins, Brown, & Newman, 1989). Early in the unit, students are provided with an example of a valid explanation, and the essential elements of an explanation—claim, relevant science ideas, and evidence that will be linked to the claim using science ideas—are defined. The early lessons also establish Explanation Quality Criteria that students can use in judging an explanation's quality and introduce a template to help them organize their thinking and writing around the essential elements. After students experience modeling activities, a new element is added to the explanation template—models and modeling as tools for thinking about how something might work—and from then on students are expected to include models in their reasoning (see Supplemental Material [p. 51 in OR 4]). As the unit proceeds, students are reminded to use ideas about atoms in their explanations and to be sure that their explanations meet the unit's Explanation Quality Criteria. The explanation template is faded in the SE but is available in the TE if the teacher feels students still need it. At the end of the unit, the scaffolding is almost completely removed, with the exception of a reminder that students can use science ideas from anywhere in the unit in their explanations.

To support teachers in helping students construct evidence-based explanations, the TE provides ideal responses for each explanation task and for each element of the explanation template. The TE also provides discussion notes (see Supplemental Material [p. 58a in OR 4]) to help teachers guide their students in using the science ideas and considering the role that models play in their explanations. In addition, as described below, online teacher resources provide rubrics for scoring the explanation tasks that are recommended for use as embedded assessments.

Educative features to promote three-dimensional teaching and learning

Figure 4 presents data from an analysis of the unit using Version 3.0 of the EQuIP rubric criteria to demonstrate how the THSB unit meets the criteria for integrating the three dimensions of NGSS (Category 1 criteria), supporting three-dimensional teaching and learning (Category 2 criteria), and monitoring student progress in all three dimensions (Category 3 criteria).

Category I: NGSS 3D Design (lessons and units): *The lesson/unit is designed so students make sense of phenomena and/or design solutions to problems by engaging in student performances that integrate the three dimensions of the NGSS.*

Lesson *and* Unit Criteria Lessons and units designed for the NGSS include clear and compelling evidence of the following:	**Specific evidence from materials** (what happened/where did it happen, with references to evidence cited elsewhere in this paper and in the Online Resources that accompany it) **and reviewers' reasoning** (how/why this is evidence)
A. **Explaining Phenomena/Designing Solutions:** Making sense of phenomena and/or designing solutions to a problem drive student learning. i. Student questions and prior experiences related to the phenomenon or problem motivate sense making and/or problem solving. ii. The focus of the lesson is to support students in making sense of phenomena and/or designing solutions to problems. iii. When engineering is a learning focus, it is integrated with developing disciplinary core ideas (DCIs) from physical, life, and/or earth and space sciences.	• Table 1 (this paper) lists phenomena in student edition (SE) that students make sense of using science ideas (drawn from DCIs and CCCs) and practices. • Figure 1 (this paper) shows the progression of development of the science ideas in the unit. • Each lesson begins with an introduction that links the lesson to prior lessons and engages students with the lesson's Key Question that they will revisit at the end of the lesson (e.g., pp. 45 and 56 in OR 4).
B. **Three Dimensions:** Builds understanding of multiple grade-appropriate elements of the science and engineering practices (SEPs), DCIs, and crosscutting concepts (CCCs) *that are deliberately selected to aid student sense making of phenomena and/or designing of solutions.*	Document evidence and reasoning, and evaluate whether or not there is sufficient evidence of quality for each dimension separately
i. Provides opportunities to *develop and use* specific elements of the SEP(s).	• Pp. 51–55 in OR 4 illustrate part of the development of the explanation practice, including reasoning from evidence, science ideas, and models. • Pp. 57–60 in OR 4 and pp. 136–139 in OR 6 show examples of tasks that engage students in explaining phenomena not encountered during instruction. • OR 10 shows the set of opportunities students have to develop and use data analysis, modeling, explanation, and communication practices to make sense of phenomena.
ii. Provides opportunities to *develop and use* specific elements of the DCI(s).	• Pp. 51–55 in OR 4 illustrate part of the development of the use of science ideas in explaining phenomena. • Pp. 57–60 in OR 4 and pp. 136–139 in OR 6 show examples of tasks that engage students in using science ideas to explain phenomena not encountered during instruction. • OR 10 shows the set of opportunities students have to develop and use science ideas derived from DCIs to make sense of phenomena.

Figure 4. An analysis of the THSB unit using *Educators Evaluating the Quality of Instructional Products Rubric for Lessons and Units (EQUIP): Science* (Version 3.0). THSB = *Toward High School Biology*; NGSS = Next Generation Science Standards; 3D = three-dimensional; ELA = English language arts; OR = Online Resources [see Supplemental Material].

iii. Provides opportunities to *develop and use* specific elements of the CCC(s). **Evidence needs to be at the *element level* of the dimensions (see rubric introduction for a description of what is meant by "element")**	• OR 6, p. 138, illustrates the use of the CCC of mass conservation (Science Idea 14) in explaining a phenomenon. Similar examples in THSB Chapters 2 and 4 are not shown. • OR 10 shows the set of opportunities students have to develop and use science ideas derived from CCCs to make sense of phenomena.
C. **Integrating the Three Dimensions:** Student sense making of phenomena and/or designing of solutions requires student performances that integrate elements of the SEPs, CCCs, and DCIs.	• In addition to the instructional examples described in Table 1 (this paper), pp. 57–60 in OR 4 and pp. 136–139 in OR 6 provide other examples of explanation tasks that require both science practices and science ideas based on DCIs. • In addition to the instructional examples described in Table 1, p. 138 of OR 6 illustrates an explanation task that requires both science practices and science ideas based on the CCC of mass conservation. • Additional examples in THSB Chapters 2, 3, and 4 that require students to use the CCC of matter conservation in their explanations are described in the text following Figure 4 (this paper). • OR 10 shows the set of performances that require students to integrate science ideas (from DCIs and/or CCCs) and practices.
D. **Unit Coherence:** Lessons fit together to target a set of performance expectations. i. Each lesson builds on prior lessons by addressing questions raised in those lessons; cultivating new questions that build on what students figured out; or cultivating new questions from related phenomena, problems, and prior student experiences. ii. The lessons help students develop toward proficiency in a targeted set of performance expectations.	• Figure 1 (this paper) shows the progression of development of the science ideas over the unit and the chapter and lesson(s) where each is targeted. • Each lesson begins with an introduction that links the lesson to prior lessons and engages students with the lesson's Key Question, which they will revisit at the end of the lesson (e.g., pp. 45 and 56 in OR 4). • Teacher Edition (TE) for each lesson includes suggestions for how teachers can facilitate a discussion about what students learned in the lesson and what they will learn in the next lesson (e.g., see Closure and Link on p. 60a in OR 4, p. 132a in OR 7, and p. 125a in OR 8). • TE for each lesson lists science ideas and practices targeted, purpose of each activity, and observations students are intended to make (e.g., pp. 44b–44d in OR 4, pp. 60b–60c in OR 5, pp. 132b–132c in OR 6, pp. 125b–125c in OR 7, and pp. 115b–115d in OR 8). • Lessons help students develop proficiency in performance expectations listed in OR 2, p. v, and additional performances that reflect different combinations of the targeted science ideas and practices (see OR 10).
E. **Multiple Science Domains:** *When appropriate,* links are made across the science domains of life science, physical science, and Earth and space science. i. DCIs from different disciplines are used together to explain phenomena. ii. The usefulness of CCCs to make sense of phenomena or design solutions to problems *across science domains* is highlighted.	• As described in Table 2 (this paper) and illustrated on p. 56 in OR 4 and on p. 125 in OR 8, students use science ideas about atom rearrangement and conservation to explain physical and life science phenomena in embedded assessments at the end of each chapter. These tasks are in addition to the physical and life science phenomena students use ideas about atom rearrangement to explain during instruction (see Table 1 in this paper). • Students are also asked to compare phenomena in physical science and life science, such as how photosynthesis is similar to the formation of rust (p. 124 in OR 8). • Although students have numerous opportunities to use conservation ideas across physical and life science phenomena, they are not asked to step back and reflect on the power of conservation ideas in making sense of phenomena across disciplines. It could be made explicit in either SE or TE by asking students to do so.
F. **Math and ELA:** Provides grade-appropriate connection(s) to the Common Core State Standards in Mathematics and/or English Language Arts & Literacy in History/Social Studies, Science and Technical Subjects.	The THSB unit engages students in several tasks related to the Common Core State Standards in ELA for Science and Technical Subjects specified below (National Governors Association Center for Best Practice & Council of Chief State School Officers, 2010): • "Determine the central ideas or conclusions of a text; provide an accurate summary of the text distinct from prior knowledge or opinions": Students have several opportunities to determine the conclusions from simplified versions or scientific publications (see p. 6 of OR 10 for description and pp. 136–137 in OR 6 and pp. 127–128 in OR 7 for examples). • "Determine the meaning of symbols, key terms, and other domain-specific words and phrases as they are used

Figure 4. (Continued).

	in a specific scientific or technical context relevant to grades 6–8 texts and topics": Students interpret symbols (e.g., p. 117 in OR 8 and pp. 134–135 in OR 6) and representations of experimental findings (e.g., p. 137 in OR 6 and p. 119 in OR 8). • "Integrate quantitative or technical information expressed in words in a text with a version of that information expressed visually (e.g., in a flowchart, diagram, model, graph, or table)": Students have numerous opportunities to relate models of chemical reactions to word equations (e.g., see Figure 2 in this paper for models of reactants and products of the chemical reactions between iron and oxygen and between baking soda and vinegar, pp. 46–48 of OR 4 for an example of how students were helped to translate their modeling of the chemical reactions into words, and pp. 51–55 of OR 4 for an example of how students were helped to incorporate ideas from their modeling experiences into an explanation).

Category II: NGSS Instructional Supports (lessons and units): *The lesson/unit supports 3D teaching and learning for ALL students by placing the lesson in a sequence of learning for all three dimensions and providing support for teachers to engage all students.*

Lesson *and* Unit Criteria Lessons and units designed for the NGSS include clear and compelling evidence of the following:	**Specific evidence from materials and reviewers' reasoning**
A. **Relevance and Authenticity:** Engages students in authentic and meaningful scenarios that reflect the practice of science and engineering as experienced in the real world. i. Students experience phenomena or design problems as directly as possible (firsthand or through media representations). ii. Includes suggestions for how to connect instruction to the students' home, neighborhood, community and/or culture as appropriate. iii. Provides opportunities for students to connect their explanation of a phenomenon and/or their design solution to a problem to questions from their own experience.	All of the following phenomena were shown to be engaging and memorable to a wide range of students: • Phenomena described in Table 1 (this paper) include firsthand experiences (e.g., vinegar + baking soda, iron rusting) and videos (e.g., nylon formation, egg-eating snake) as well as data from radioactive labeling experiments when firsthand experiences are not possible (but that students model using a yellow sticker on a black ball to visualize the tracking of a carbon atom from brine shrimp to fish body proteins). • Additional videos, not shown, include spider silk formation, time-lapse photos of corn growing, time-lapse photos of the growth of a German shepherd puppy over 7 months, and time-lapse photos of the growth of a baby girl over 13 years. Students are also asked to explain phenomena with which they may have had personal experience (e.g., why bubbles form when hydrogen peroxide is put on a wound, why the mass of the Statue of Liberty should increase over time, and why the growth of mushrooms on a fallen dead tree does not violate conservation principles; see Table 2).
B. **Student Ideas:** Provides opportunities for students to express, clarify, justify, interpret, and represent their ideas and respond to peer and teacher feedback orally and/or in written form as appropriate.	• The SE provides numerous opportunities for students to express and clarify their ideas in writing (see, e.g., Key Questions on p. 45 of OR 3 and p. 61 of OR 5 and Pulling It Together Questions, e.g., p. 140 of OR 6), represent their ideas (e.g., p. 68 of OR 5 and p. 125 of OR 8), and justify their ideas (e.g., see Table 2 in this paper). • The TE describes commonly held student ideas and misconceptions relevant to each chapter and how they may be manifest in student explanations (e.g., pp. 1d–1e of OR 3). • The SE engages students in challenging misconceptions expressed in text (e.g., p. 131, Pulling It Together Question 2 of OR 7). • Embedded assessments at the end of each chapter (see Table 2 in this paper) provide opportunities for students to explain phenomena and obtain feedback from peers and/or the teacher (described in the Methods section for Research Question 2 in this paper).

Figure 4. (Continued).

C. **Building Progressions:** Identifies and builds on students' prior learning <u>in all three dimensions</u>, including providing the following support to teachers: i. Explicitly identifying prior student learning expected for all three dimensions. ii. Clearly explaining how the prior learning will be built upon.	• The content storyline (see Figure 1 in this paper) shows how science ideas (DCIs and CCCs) develop over the unit. • The evidence tables provided to teachers during professional development (OR 10) show how practices and science ideas are integrated over the unit as students analyze data, model invisible mechanisms, and construct explanations of phenomena.
D. **Scientific Accuracy:** Uses scientifically accurate and grade-appropriate scientific information, phenomena, and representations to support students' 3D learning.	• The unit was developed and extensively reviewed for accuracy by doctoral-level biologists, chemists, and biochemists and for grade appropriateness by classroom teachers and science education researchers.
E. **Differentiated Instruction:** Provides guidance for teachers to support differentiated instruction by including: i. Appropriate reading, writing, listening, and/or speaking alternatives (e.g., translations, picture support, graphic organizers, etc.) for students who are English language learners, have special needs, or read well below the grade level. ii. Extra support (e.g., phenomena, representations, tasks) for students who are struggling to meet the targeted expectations. iii. Extensions for students with high interest or who have already met the performance expectations to develop deeper understanding of the practices, DCIs, and CCCs.	THSB development took account of student needs in a variety of ways: • Using simple language whenever possible and restricting the use of technical terms to those needed to communicate about phenomena (see OR 2, p. xi–xii). • Using multiple representations that were tested for comprehensibility with a diverse population of students (see OR 2, pp. vi–vii). • Using videos of phenomena that could not be observed firsthand to help students who struggled to construct images from words. However, we did not attempt to develop and test additional support for differentiating instruction. Now that the unit has achieved its goals, based on analysis and empirical studies (Herrmann-Abell et al., 2016), we would like to investigate how to support students with learning difficulties and students with exceptional interest and ability.
F. **Teacher Support for Unit Coherence:** Supports teachers in facilitating coherent student learning experiences over time by: i. Providing strategies for linking student engagement across lessons (e.g., cultivating new student questions at the end of a lesson in a way that leads to future lessons, helping students connect related problems and phenomena across lessons, etc.). ii. Providing strategies for ensuring student sense making and/or problem solving is linked to learning in all three dimensions.	• The TE provides the rationale for the chapter and lesson sequence (OR 2, pp. v–vi) and for the sequence of activities within lessons, including the role of Pulling It Together questions in connecting each lesson to prior and subsequent lessons (OR 2, pp. xii–xiii). These specific questions go beyond merely providing generic strategies that teachers have to figure out how to apply to specific instances. • Explanation tables (e.g., OR 4, pp. 51–55) require students to use science ideas (both DCIs and CCCs) and practices such as evidence from data and ideas from using models in explaining phenomena.

Figure 4. (Continued).

G. **Scaffolded differentiation over time:** Provides supports to help students engage in the practices as needed and gradually adjusts supports over time so that students are increasingly responsible for making sense of phenomena and/or designing solutions to problems.	As described in this paper's Results section for Research Question 1, the unit applies the cognitive apprenticeship approach to the explanation practice by • Modeling a valid explanation (e.g., see pp. 52–53 in OR 4). • Providing criteria students can use to evaluate the quality of an explanation, including their own (e.g., see p. 51 in OR 4). • Providing a template to guide the development of students' own explanations (e.g., see pp. 52–53 in OR 4). • Asking students to use the criteria to evaluate their own explanations (e.g., see p. 55 in OR 4). • Giving students numerous opportunities to explain phenomena (e.g., see pp. 57–60 in OR 4 and pp. 137–139 in OR 6). • Gradually withdrawing the scaffold, including the structured table and reminders to include a discussion of what is happening to atoms and molecules in their explanation (e.g., see p. 125 in OR 8).

Category III: Monitoring NGSS Student Progress (lessons and units): *The lesson/unit supports monitoring student progress in all three dimensions of the NGSS as students make sense of phenomena and/or design solutions to problems*.

Lesson *and* Unit Criteria Lessons and units designed for the NGSS include clear and compelling evidence of the following:	**Specific evidence from materials and reviewers' reasoning**
A. **Monitoring 3D student performances:** Elicits direct, observable evidence of 3D learning; students are using practices with core ideas and CCCs to make sense of phenomena and/or to design solutions.	• Pre- and posttests provide opportunities to assess the quality of student explanations and the reduction in their misconceptions (Herrmann-Abell, Koppal, & Roseman, 2016). • Embedded assessments (see Table 2 in this paper) enable teachers to monitor students' use of science ideas, evidence, and reasoning from modeling tasks to explain phenomena.
B. **Formative**: Embeds formative assessment processes throughout that evaluate student learning to inform instruction.	• Embedded assessments (see Table 2 in this paper) enable teachers to monitor students' use of science ideas, evidence, and reasoning from modeling tasks to explain phenomena.
C. **Scoring guidance:** Includes aligned rubrics and scoring guidelines that provide guidance for interpreting student performance along the three dimensions to support teachers in (a) planning instruction and (b) providing ongoing feedback to students.	• Scoring rubrics, including elements of an ideal response and relevant misconceptions to watch for (e.g., pp. 2–3 and 5–6 in OR 9), provide guidance in evaluating how well student explanations reason from evidence, science ideas, and modeling activities.
D. **Unbiased tasks/items:** Assesses student proficiency using methods, vocabulary, representations, and examples that are accessible and unbiased for all students.	• Assessment items on the pre- and posttests were reviewed using a procedure that evaluated the items' comprehensibility, evaluated the appropriateness of the task contexts, and eliminated test-wiseness issues (DeBoer et al., 2008). • Pilot and field testing revealed no significant bias based on gender, ethnicity, or English language status.
E. **Coherent Assessment system:** Includes pre-, formative, summative, and self-assessment measures that assess 3D learning.	• The pretest, embedded assessments, and posttest all include items that assess students' use of the targeted practices with science ideas based on core ideas and CCCs. The current unit lacks explicit self-assessment tasks, although students are asked to evaluate their own explanations using the Explanation Quality Criteria.
F. **Opportunity to learn:** Provides multiple opportunities for students to demonstrate performance of practices connected with their understanding of DCIs and CCCs and receive feedback.	• Pulling It Together questions at the end of each lesson (see Table 2 in this paper and OR 9) provide opportunities throughout the unit for students to demonstrate performance of 3D understanding.

Figure 4. (Continued).

In addition to the evidence cited in Figure 4, the THSB unit provides many other opportunities for students and teachers to achieve the NGSS goal of integrating core disciplinary ideas, science practices, and crosscutting concepts to explain a range of phenomena in physical and life science settings. For example, Chapter 2 science ideas about atom conservation in chemical reactions, modeling and explanation practices, and the crosscutting concept of atom conservation explaining mass conservation help students make sense of observations of changes in measured mass in open containers, such as when baking soda reacts with vinegar, iron reacts with oxygen, a log burns in a fireplace, and copper in the Statue of Liberty reacts with substances in the atmosphere.

Similarly, in Chapters 3 and 4, students develop explanations of plant and animal growth, both of which involve increases in measured mass, by making sense of data about mass increases in a growing willow tree, the incorporation of ^{14}C from glucose into cellulose in an experiment that tests the effect of an herbicide on the growth of cress plants, and the incorporation of ^{14}C from brine shrimp proteins into herring fish body proteins in an experiment investigating possible food sources for fish farming. Lesson Guides in the TE identify the ideas and practices that are integrated into each lesson (see Supplemental Material [p. 60b in OR 5]).

Embedded assessment tasks at the end of each chapter tap all three dimensions of science learning and serve as examples of what it means to integrate those dimensions in teaching and learning. The assessment tasks require students to use (a) science ideas, including those derived from both disciplinary core ideas and crosscutting concepts (see the Content Storyline Map in Figure 1); (b) evidence gathered by students as they engage in the practices of observation and data analysis; and (c) reasoning from models and modeling activities to explain a phenomenon. For example, in one task students are asked to explain why the growth of mushrooms on a fallen dead tree does not violate the concept of matter conservation. Suggested answers and rubrics for scoring tasks such as these help teachers see how disciplinary core ideas, practices, and crosscutting concepts can be integrated into assessments (see Supplemental Material [OR 9]).

Face-to-face PD extended the discussion of three-dimensional learning in the THSB unit in several ways. In sessions for experienced THSB teachers, for example, teachers (a) examined evidence tables that summarize how the unit integrates each science idea with each practice (see Supplemental Material [OR 10]); (b) considered the range of phenomena across the unit's four chapters that students observe, model, and explain; and (c) used the rubrics to score samples of student work.

Having a set of appropriate criteria for designing educative features and ensuring that the materials meet those criteria are essential first steps in increasing the knowledge and skills teachers need to implement the NGSS vision. It is also essential to find out whether the educative features actually do increase teachers' knowledge and skills. This requires valid measures of teacher knowledge and practice that could be used in future studies of what teachers learn from educative curriculum materials.

Research question 2: Can the same design heuristics also guide the development of measures to evaluate teacher knowledge and practice?

The first iteration of a measure was an online test of teachers' knowledge of science ideas, relevant student misconceptions, and phenomena that was administered before PD, after PD, and after teachers taught the unit (Flanagan et al., 2013). Although the data from the teacher test were useful in informing revisions to the unit, the variability in teacher performance across contexts indicated that the assessment was not yet a valid measure of teacher knowledge. We also wanted to measure teachers' knowledge using a more authentic task—one that they would view as being useful to their work. We focused on using teachers' analysis and synthesis of their students' responses to the unit's embedded assessments as a possible measure of teacher knowledge.

Table 3. Study participants.

Teacher characteristic	*M*	*n*
Race		
Black		1
White		5
Gender		
Female		5
Male		1
Average years of teaching experience		
Physical science	9.3	
Life science	7.4	
Education		
Undergraduate degrees		5 (biology) 1 (education)
Graduate degrees		1 (education)
Experience with *Toward High School Biology* unit		
1 year		2
2 years		2
3 years		2

Methods

Participants

During the 2015 implementation of the unit, six middle school teachers from a mid-Atlantic suburban school district participated in a study involving the embedded student assessments included at the end of each chapter in the THSB unit (see Table 2). Table 3 provides information about the participating teachers, including their prior experience with the THSB unit. Teachers had participated in 3 days of PD prior to their first year using the unit.

The teachers participated in an additional 1 day of PD that updated them on revisions to the unit and why they were made and provided them with an opportunity to examine and discuss evidence of the unit's alignment with NGSS (see Supplemental Material [OR 10]). Teachers analyzed and scored samples of student work and considered how the THSB unit is designed to achieve NGSS's three-dimensional vision. For the study, teachers agreed to (a) use rubrics provided by the research team to score embedded assessment tasks involving written explanations for 15 of their students representing the range of performance in all of their classes; (b) summarize their findings across the sample, noting ideas understood by most students and ideas posing difficulty and what the implications might be for subsequent lessons; (c) e-mail their reports to the research team along with scanned copies of the 15 students' work within a day or two of when students completed each embedded assessment; and (d) provide feedback to students either by having the class evaluate sample explanations or by having students self-score and revise their own explanations. Figure 5 shows excerpts from the scoring table teachers used to record their judgments about the embedded assessment task that is reported on in this study (the actual worksheet is shown in the Supplemental Material [OR 9]).

Teachers were told to spend about an hour on each embedded assessment task and were compensated financially for their time. By scoring each task immediately after students completed it, teachers would be able to use the results of their analyses to inform their instruction. The research team would use the results of the teachers' analyses to measure their SMK and PCK.

Indicators of teacher knowledge and their use in judging teacher knowledge

From the teachers' reports on their students' embedded assessments, we identified potential indicators of teachers' SMK, PCK for science ideas and topics, and PCK for science inquiry that seemed to us to be most relevant to the NGSS science practice of constructing explanations (see Table 4). For example, would teachers notice if their students applied a science idea incorrectly, failed to cite evidence to support claims, or had difficulty linking evidence to claims? Would teachers notice if their students were better at

Question 1a: Hydrogen peroxide is a clear, colorless liquid. When you apply it to a cut or scrape on your skin, you will see bubbles and you may hear a “fizzing” sound. Figure 1.3 uses LEGO models to represent hydrogen peroxide on the left. On the right are models showing what happens to hydrogen peroxide when you apply it to a cut or scrape on your skin. Do you think a chemical reaction occurs when hydrogen peroxide is put on a wound? Explain your answer, using evidence, science ideas, and models as needed to support your answer (use the table for your notes).																
		Student ID Codes:														
Elements of an Ideal Response																
Claim	A chemical reaction occurs.															
States Science Idea	If atoms rearrange to form new molecules then a chemical reaction occurs (Science Idea #5).															
Uses Evidence	The LEGO models in Figure 6.1 show that H_2O and O_2 have different arrangements of atoms (are different molecules) from H_2O_2.															
Uses Models	The LEGO models show that H_2O and O_2 molecules could have been formed from rearranging the atoms of H_2O_2 during a chemical reaction.															
Uses Science Idea	O and H atoms must disconnect and join in different ways to form H_2O and O_2 from H_2O_2.															
Misconceptions																
The ending substances are made up of the same molecules as starting substances. (confusing atoms and molecules)																
Substances change their properties during chemical changes, but their molecules stay the same.																
The atoms and molecules of the starting substances are completely independent of/unrelated to the ending substances.																
Gas has formed; gas formation always indicates a chemical reaction has occurred.																
The number of models increased, so atoms cannot have just rearranged. (confusing atoms and molecules, difficulties with conservation of matter)																

Figure 5. Scoring rubrics for *Toward High School Biology* Chapter 1 embedded assessment task. Teachers scored each student’s explanation, assigning up to 2 points for each element and noting whether the explanation reflected any indicators of difficulty. ID = identification.

Table 4. Indictors of teacher knowledge.

Type of knowledge	Indicator
Subject matter knowledge	
Scientific ideas	(1) Teacher correctly identifies correct/incorrect science ideas in their students' responses (based on how well teacher scores match researcher scores on elements involving stating science ideas).
	(2) Teacher's summary correctly paraphrases the science idea (note which science ideas are correctly paraphrased) and gives evidence-based example of students correctly using the science idea.
Coherence	(3) Teacher's summary notes a specific idea/activity beyond the current lesson where identified students' difficulties with ideas could cause problems (mention of specific practice counts as evidence for PCK for that practice).
	(4) Teacher's summary notes a specific idea/activity from an earlier lesson that was needed but not learned well enough for students to build on.
Pedagogical content knowledge	
Student ideas/ misconceptions	(5) Teacher correctly identifies examples of misconceptions in their students' responses.
	(6) Teacher's summary gives evidence-based example(s) of student misconceptions or difficulties with ideas in summary report (e.g., confusing atoms/molecules, thinking the mass of the willow tree comes only from water, overestimating the contribution of water to the increase in the willow tree's mass, confusion between total mass and measured mass in terms of what is conserved).
Pedagogical content knowledge for science practices	
Evidence	(7) Teacher correctly scores the evidence component of students' explanations.
	(8) Teacher's summary mentions citing evidence as a strength or difficulty in student explanations.
Models	(9) Teacher correctly scores the models component of students' explanations.
	(10) Teacher's summary mentions reasoning from models (could be mentioned when describing student difficulties).
Using science ideas	(11) Teacher correctly scores the using science ideas component of students' explanations.
	(12) Teacher's summary mentions using science ideas (could be mentioned when describing student difficulties).

Note. Indicators 1, 5, 7, 9, and 11 are quantitative measures, and Indicators 2, 3, 4, 6, 8, 10, and 12 are qualitative measures (based on the extent of the match of teacher judgments to researcher judgments).

stating the science ideas than at using them to link evidence to claims? Would teachers notice if any students stated that models "proved" the claims rather than merely giving them ideas about underlying molecular causes of the phenomena observed? Each component of teacher knowledge (except coherence) was evaluated in terms of a quantitative first indicator based on the percentage of times the teachers' scores of student responses matched the researchers' scores and a qualitative second indicator based on researchers' analyses of the teachers' written responses in their summaries (see Supplemental Material [p. 7 in OR 9]). The indicators were designed to measure these aspects of teacher knowledge that seemed to us to be most relevant to NGSS.

Knowledge of scientific ideas. People with a sophisticated understanding of science content are able to simplify the science ideas without distorting them. Therefore, we looked at how well teachers were able to identify science ideas—often paraphrased—in their students' responses to the embedded assessment tasks (Indicator 1) and how well teachers themselves paraphrased the ideas in their own summary reports (Indicator 2).

Coherence. A coherent understanding includes knowing the connections among the science ideas and between the science ideas and lesson activities (Roseman et al., 2010). After identifying difficulties exhibited in their students' responses, teachers were asked to identify difficulties students had in the current lesson that could cause them problems in subsequent lessons or chapters (Indicator 3) or difficulties they carried over from earlier lessons that were continuing to cause them problems (Indicator 4). For example, students who were still confusing atoms and molecules at the end of Lesson 1.5 might have difficulty making sense of the chemical reaction involved in nylon formation during Lesson 1.6 and similar reactions involved in plant and animal growth in Chapters 3 and 4. If students were not yet using ideas about atom rearrangement to account for the formation of new substances, they could have problems explaining mass conservation and changes in measured mass in

Chapters 2, 3, and 4. If teachers were able to foresee the implications of these problems for later lessons, it could indicate that they appreciated how ideas are connected to one another and to the unit's activities.

PCK: Student ideas/misconceptions. Attending to students' ideas starts with knowing relevant student misconceptions and how they might be manifest in students' oral and written work. Teachers were asked whether their students' responses included any of the misconceptions listed in the chart in Figure 5, along with any other misconceptions they could identify. Teachers' knowledge of students' ideas was measured by their ability to recognize misconceptions in the samples of their students' work (Indicator 5) and their ability to characterize and give an example of the misconception or difficulty (Indicator 6).

PCK: Science practices. Consistent with NGSS's vision for three-dimensional learning, teachers' knowledge of science practices was measured in the context of the specific phenomena students were asked to explain in the embedded assessment tasks and the science ideas they were expected to use in their explanations. Teachers' knowledge of the practice of analyzing and interpreting data was judged by their ability to correctly score the evidence component of students' explanations, which required teachers to decide whether the evidence students cited was relevant to and supported the claim (Indicator 7) and to comment in their summary reports on whether students did or did not cite evidence in their explanations (Indicator 8).

Teachers' knowledge of the practice of using models as a tool for reasoning about phenomena—particularly those involving atoms and molecules—was judged by their ability to determine whether students' use of models was consistent with the science ideas they cited in their explanations and whether students' use of models helped to show that the claim in their explanation was reasonable (Indicator 9). Teachers also had to comment in their summary reports on whether their students did or did not use models appropriately in their explanations (Indicator 10). Last, teachers' knowledge of the practice of using science ideas to link evidence to claims, another aspect of valid reasoning, required teachers to decide whether students applied the general principle in the science ideas to the specific claim and evidence (Indicator 11) and to comment in their summary reports on whether their students did or did not do so (Indicator 12).

Scoring

To arrive at a score for the quantitative first indicator of the teacher knowledge components (except for coherence) described previously (Indicators 1, 5, 7, 9, and 11 in Table 4), researchers worked in pairs to independently score 30 sample student explanations, five from each teacher. Using the same rubric as the teachers (see Figure 5), the researchers scored each student's explanation as 0, 1, or 2 based on whether the student (a) provided a correct answer, (b) stated the science idea, (c) used the science idea, (d) provided relevant evidence, and (e) used models in his or her explanation. The researchers also noted whether the student responses exhibited any of the likely misconceptions. The researchers then reconciled their judgments and compared their judgments to the teachers' judgments. Agreement was reached when researcher and teacher judgments were within 1 point of the two possible points for that component of the student answer. For misconceptions, teachers who called out a misconception directly on a student's paper but failed to note it in the study table were given credit for having identified the misconception. Researcher and teacher judgments that differed by more than 1 point were considered to not match. For example, if the researcher did not think that a student used the science idea in his or her explanation (a score of 0), but the teacher thought that the student did use the science idea (a score of 2), this would not be considered a match. The percentage of times the teachers' and researchers' scores matched was calculated. Teachers received credit for the first indicator if their judgments matched the researcher judgments for at least 80% of student responses.

To arrive at a score for the qualitative second indicator of the knowledge components described previously, and for both of the coherence indicators (Indicators 2, 3, 4, 6, 8, and 12 in Table 4), researchers again worked in pairs and independently read the teachers' responses to Questions 1–3 in their summary reports (see Supplemental Material [p. 7 of OR 9]). The researchers identified examples that could count as evidence of teacher knowledge using the indicators in Table 4 and then reconciled any differences in judgments. Teachers received credit for the qualitative indicators if they provided at least one example.

An overall rating was then assigned to each knowledge component. If the teacher met both indicators for a particular component, we concluded that there was robust evidence that the teacher had that knowledge. If only one indicator was met, we concluded that there was only some evidence that the teacher had the knowledge. If neither indicator was met, there was no evidence that the teacher had the knowledge.

Results

Table 5 presents scores for each teacher on each indicator of teacher knowledge. Table 6 presents a few examples teachers provided in their summary reports that were counted as evidence for the qualitative indicators.

Table 5. Scores on indicators of teacher knowledge for THSB teachers.

	Teacher (Years using THSB)					
Indicator	A (2)	B (3)	C (1)	D (1)	E (2)	F (3)
Subject matter knowledge						
Scientific ideas	Robust	Robust[a]	Robust[a]	Robust	Robust	Robust[a]
(1) Teacher correctly identifies as correct/incorrect the statement of science ideas in their students' written responses	1 (100%)	1 (90%)	1 (80%)	1 (100%)	1 (90%)	1 (100%)
(2) Teacher's summary correctly paraphrases the science idea and gives evidence-based examples of students correctly using the idea (note which science ideas are correctly paraphrased)	1	1	1	1	1	1
Coherence	None	Some	Some	Some	None	Some
(3) Teacher's summary notes a specific idea/activity beyond the current lesson where identified students' difficulties with ideas could cause problems	0	1	1	1	0	1
(4) Teacher's summary notes a specific idea/activity from an earlier lesson that was needed but not learned well enough for students to build on	0	0	0	0	0	0
Pedagogical content knowledge: Student ideas/misconceptions	Robust	Some	Some	Some	Some	Robust
(5) Teacher correctly identifies examples of misconceptions in their students' written responses	1 (90%)	1 (80%)	0 (70%)	1 (80%)	1 (80%)	1 (100%)
(6) Teacher's summary gives evidence-based example(s) of student misconceptions or difficulties with ideas in summary report (e.g., confusing atoms/molecules, thinking a substance's properties can change)	1	0	1	0	0	1
Pedagogical content knowledge for science practices: Explanations						
Using evidence	Some	Some	Some	Robust	Robust	Some
(7) Teacher correctly scores the evidence component of students' written explanations	1 (100%)	1 (90%)	1 (90%)	1 (90%)	1 (100%)	1 (100%)
(8) Teacher's summary mentions citing evidence as a strength or difficulty in student explanations	0	0	0	1	1	0
Using models	Some	Some	None	Robust	Some	Some
(9) Teacher correctly scores the models component of students' written explanations	1 (100%)	1 (90%)	0 (50%)	1 (80%)	1 (100%)	1 (100%)
(10) Teacher's summary mentions reasoning from models as a strength or difficulty in student explanations	0	0	0	1	0	0
Using scientific ideas	Robust	Robust	Robust	None	Robust	Some
(11) Teacher correctly scores the using science ideas component of students' written explanations	1 (100%)	1 (90%)	1 (80%)	0 (70%)	1 (100%)	1 (100%)
(12) Teacher's summary mentions using science ideas as a strength or difficulty in student explanations	1	1	1	0	1	0
Total score (out of 12)	8	8	7	8	8	8

Note. THSB = *Toward High School Biology*.
[a]Teacher notes that students have cited/used alternative science ideas (Science Ideas 2 and 4 vs. Science Idea 5).

Table 6. Evidence from teachers' reports on students' explanations.

Type of knowledge	Teacher and example provided
Subject matter knowledge	
Scientific ideas	Teacher E: "Almost all students understand Science Idea #5, that when the atoms in a molecule rearrange that a chemical reaction has occurred. I say this because almost all students correctly identified the chemical vs. non-chemical reactions. They also stated SI#5 in their answers. For example, student #1315610004 says that the boiling water is not a chemical reaction 'because all of the molecules are the same' and student #1315610022 stated that, 'for the substance to change it needs to have the atoms rearrange.'"
Coherence	Teacher B: "My students will most likely have difficulty in understanding the chemical reaction that occurs between monomers to create polymers. I will need to walk them through the formation of water so they can understand that something new is made. They will think that the polymer is not something new because only the atoms at the end come off."
	Teacher D: "About ¾ of my students are having a hard time conceptualizing monomers and polymers" and "Science Idea #6 [which applies the idea that atom rearrangement is why new substances are produced during chemical reactions to the specific case of polymer formation] will be intimidating to them, but as we move forward I see it being a building block for better understanding [of plant and animal growth]."
	Teacher C: "Science Idea #4 about the type, number and arrangement being a different substance seems to be weakly understood at this point. With Chapter 2 and the counting of the number and type of atoms this should be strongly enhanced."
Pedagogical content knowledge	
Student ideas	Confusing atoms and molecules:
	Teacher A: "The original substance, hydrogen peroxide, had a chemical formula of H_2O_2. After the reaction, the molecules have been separated, as the new chemical formulas are O and H_2O" (1315521054).
	"Chemical reactions have taken place if the molecules separate (science idea #5). The molecules did not separate, as shown in the model, so a chemical reaction didn't take place" (1315521054).
	A gas always indicates a chemical reaction has occurred:
	Teacher A: "Also, when touching the wound, the hydrogen peroxide starts to bubble and fizz which indicates a chemical reaction" (1315521001).
	Difficulty understanding atom rearrangement:
	Teacher F: "For the students who are struggling with Science Idea #5, they are also the ones that are unsure of what exactly is happening during the reaction. One of the students, in their response, seemed to think that there was an increase, or 'replication,' of atoms in the H_2O_2 question. This was also the student who thought that water heating did represent a chemical reaction because of the change in state, from liquid to gas."
Pedagogical content knowledge for science practices: Explanation	
Using evidence	Teacher E: "The idea that might be causing problems is SI#2 [substances with different properties form during chemical reactions]. Several students listed this idea as evidence. This is a problem because the properties of the substances were never talked about in the prompts. Students just used prior knowledge, rather than the text, to answer the question."
Using models	Teacher D: "Determining the difference between evidence and models is particularly challenging."
Using science ideas	Teacher E: "The difficulty I see some of my students having is connecting the Science Ideas to the explanations."
	Teacher A: "The majority of students accurately paraphrased Idea #5 and used it to link evidence to answer."
	Teacher B: "8/15 students were able to use the idea [#5] to help explain their answer."

Note. SI = Science Idea.

Subject matter knowledge

On the scientific ideas indicators, all of the teachers' judgments about their students' science ideas matched researcher judgments for at least 80% of their students' explanations (Indicator 1), and all of the teachers were able to paraphrase the ideas and cite evidence showing that their students understood the ideas (Indicator 2; see Tables 4 and 5). Therefore, there was robust evidence that the teachers had an understanding of the science ideas. Moreover, Teachers B, C, and F reported on an alternative student explanation for why a chemical reaction does or does not take place when hydrogen peroxide is applied to a wound or when water is heated. Students justified the claim that a chemical reaction occurs in the first case because H_2O and O_2 are new substances and have different atomic compositions from H_2O_2 and that a chemical reaction does not occur in the second case because the ending substance H_2O is the same as the starting substance. The three teachers scored their students' responses consistent with this valid alternative approach and noted that they did so in their reports. On the coherence indicators, four teachers provided some evidence of understanding implications of student difficulties for subsequent lessons. For example, Teachers B and D described difficulties their students could have in applying ideas about atom rearrangement (encountered in the context of small molecules in the current lesson) to the formation of polymers from monomers in subsequent lessons (see Table 6). Teacher C noted a difficulty her students still had with an idea from a previous lesson (the idea that substances are different because they have different arrangements of atoms) and described activities in subsequent lessons that might address the difficulty (see Table 6). None of the teachers provided more than a single example, and two of the teachers provided no examples.

PCK of student ideas/misconceptions

Two teachers provided robust evidence of understanding their students' ideas, and four teachers provided some evidence (see Table 5). One teacher did not receive credit for Indicator 5, but 70% of her judgments about students' explanations agreed with researchers' judgments. Three teachers who received credit for Indicator 6 provided specific evidence that their students confused atoms and molecules. As shown in Table 6, Teacher A also included an example of a student's response indicating the student held the misconception that the production of a gas was always associated with a chemical reaction. Teacher F provided additional insight into the relationship between students' lack of understanding of atom rearrangement (Science Idea 5) and difficulties distinguishing between the hydrogen peroxide and the heating water phenomena. Teacher F was one of the teachers who gave credit for her students' alternative approach to explaining the hydrogen peroxide phenomenon but noted that those students were not using atom rearrangement in their explanations.

PCK of science practices

Most of the teachers' responses showed some evidence of meeting the PCK of science practices indicators. Overall, the teachers scored highest on the using science ideas in explanations indicators (see Table 5). For example, Teacher A noted that his students were able to use the science ideas in explanations (see Table 6), which was consistent with researcher judgments. Teachers E and B noted that their students had more difficulty using the science ideas than merely stating them. For the indicators on using evidence in explanations, teachers' reports provided less evidence that they were attending to their students' use or lack of use of evidence in their explanations. Although all teachers met Indicator 7, only two teachers also met Indicator 8. Teachers provided the least evidence for the using models in explanations indicators. All but one met Indicator 9, but only one teacher also met Indicator 10. This indicates that the teachers' responses to the summary report questions lacked an evaluation of their students' strengths and difficulties with reasoning from models.

Discussion

Research Question 1

Our adaptation of the Davis and Krajcik (2005) design heuristics can be used to focus curriculum developers' thinking on the kinds of educative features that teachers will need to help them implement materials that support NGSS. Over time, as more curriculum materials are designed with NGSS in mind, it will be important to examine the nature of their educative features, the extent to which they are used, and their effectiveness in increasing teachers' knowledge and skills with regard to NGSS.

Given the magnitude of the changes called for in NGSS, the pace at which states and districts are adopting the new standards, and what research has to say about how teachers make use of educative materials, it is essential to prioritize what teachers most need to know and be able to do with regard to NGSS and what is practical within the context of a specific curriculum material. Our adaptation of the design heuristics and the educative features we developed for the THSB unit do not address every change called for in NGSS. No 8-week unit would be able to accomplish that. Indeed, NGSS calls for seven conceptual shifts, each one involving a multitude of changes on the part of teachers, schools, curriculum and assessment developers, higher education, and more. Instead, we chose to focus on helping teachers understand at a deep level what it means to integrate NGSS disciplinary core ideas, crosscutting concepts, and science practices and to guide their students in drawing on these three dimensions to make sense of a specific set of phenomena. And as noted earlier, we decided to build most of the support for three-dimensional learning into the SE itself so that, for example, the scaffolding provided in the SE for helping students engage in the practice of explaining phenomena by drawing on evidence and science ideas also illustrated for teachers how the three dimensions can work together to promote understanding. Other developers will make different choices about which aspects of NGSS are most relevant to their materials, which aspects will require the most support, and how best to provide that support to teachers. Our use of the EQuIP rubric criteria to analyze the THSB unit at various stages of development provided valuable insights into the extent to which the unit was achieving the goal of three-dimensional design. It is important to note, however, that understanding the criteria and applying them correctly and consistently is not a trivial undertaking. Much more clarification of the criteria as well as empirical results to support them will be needed in order for the EQuIP rubric to have a significant impact on the design of materials.

In developing the THSB unit, we were fortunate to work with a number of very talented, dedicated, and extremely busy teachers who tried out various iterations of student and teacher materials. We benefited from their feedback and from opportunities to observe how they used the materials in their classrooms. Our work confirms what Davis and Krajcik (2005) theorized: that "teacher learning will best be promoted by a set of complementary approaches, not by a single one" (p. 4). Some of the teachers in our study preferred to use the TE in the classroom, whereas many others worked only with the SE. Some referred to the online resources regularly, whereas others did not. All, of course, participated in the face-to-face PD and used the student materials with fidelity. Determining the most effective mix of educative features will require much more research. In the meantime, providing multiple approaches to supporting teachers in implementing NGSS seems to make sense, particularly to address the specific needs of teachers at different grade levels. According to Banilower, Nelson, Trygstad, Smith, and Smith (2013), for example, elementary teachers rarely have access to science curriculum materials that are educative in any sense.

Developers of new materials for NGSS have many challenges, chief among them helping teachers understand and use the new standards. Our adaptation of the Davis and Krajcik (2005) design heuristics, our use of the EQuIP rubric criteria, and the educative features developed for the THSB unit can serve as starting points for this effort.

Research question 2

How teachers responded to the embedded assessment tasks provided considerable and different kinds of information on teachers' SMK, PCK of student ideas/misconceptions, and PCK for science practices. The first indicator for each type of knowledge provided information about the extent to which teachers could use their knowledge to judge the different components of their students' explanations. The second indicator for each type of knowledge provided information on teachers' inclination to notice and comment on strengths and weaknesses of the various components of their students' knowledge and explanation writing. Differences in teachers' scores may also reflect differences in their writing ability and the lack of examples of what reports should look like. The time required for teachers to score their students' explanations and then complete a written report may have been another factor: 2 hr rather than 1 hr per task is a more realistic expectation. Furthermore, the teachers were teaching in different schools with different histories of implementing the Common Core English Language Arts standards, so teachers and their students varied in their experience with constructing explanations and arguments in the context of language arts.

For the coherence component of SMK, teachers had less opportunity to receive a robust score because, unlike their scoring of the other components, they were not asked to quantitatively score students' understanding of the storyline. Also, the fourth indicator may have been less relevant for a task that occurs early in the unit. We might have elicited richer responses from teachers if we had asked them to look at the Content Storyline Map (see Supplemental Material [p. viii of OR 2]) and describe ideas and lessons in which problems they had observed with their students could have an impact later in the unit. Or we might have presented teachers with a hypothetical student difficulty and asked them to use the Content Storyline Map to identify lessons in which the student might have problems. Providing a common example for teachers to address could also level the playing field for teachers whose own students exhibited few misconceptions. Given the success of THSB in reducing student misconceptions (Herrmann-Abell et al., 2016), these strategies are worth considering.

This study was designed to examine the usefulness of the educative features heuristics for designing materials for and tools to assess teacher learning, not to measure the effects of the educative features on teacher learning. However, data from previous work suggest that teachers involved in this study have learned a great deal from using the THSB unit. Using an online test to measure science content knowledge and knowledge of students' ideas, Flanagan et al. (2013) reported that teachers' knowledge of atoms/molecules and their inclination to use atom rearrangement and conservation to explain phenomena increased after PD and further increased after they used the unit. Teachers' knowledge of students' misconceptions, however, was minimal before PD and showed no increase. Differences between teacher knowledge reported in Flanagan et al. and in the study reported here may be due to (a) considerably more support in the current iteration of the THSB for improving teachers' knowledge and practice, leading to higher quality student explanations (Herrmann-Abell & Roseman, 2016); (b) wider adoption of NGSS, leading to an increase in teachers' attention to science practices, especially data analysis and explanation; and (c) most teachers' knowledge increasing as their experience with the THSB unit has increased.

The research literature has pointed to the difficulty of developing measures of teacher knowledge that are both valid and authentic. What is more, our review of the literature found no instruments that were designed to measure NGSS-related teacher knowledge. The measure developed for this study, therefore, may be the first to be applied to a curriculum material that is designed explicitly to align with NGSS and the first to operationalize teacher knowledge of the NGSS vision of three-dimensional learning in an authentic way, that is, in the context of teachers' analysis and synthesis of their students' written responses to embedded assessments. The high level of compliance, which is not always the case with more traditional measures, might be because the teachers did not perceive the measure as a test of their knowledge but rather as a task that would provide useful insights into their students' thinking. Our study, unlike others, focused specifically on science and on a curriculum material rather than on video (see, e.g., Kerstin, 2008) and on in-service teachers with experience using the THSB unit rather than on preservice teachers (see, e.g., Schwarz et al., 2008).

Although more work is needed to refine and validate the measure, even in its preliminary form it can shed light on teacher knowledge. Further studies are also needed to determine whether the teacher knowledge we observed resulted from teachers' use of the educative features (or components of those features) in the THSB unit. The existence of validated measures based on the one described here would make such studies possible.

Conclusions

We have demonstrated that the heuristics proposed by Davis and Krajcik (2005) for the design of educative curriculum materials can be adapted for use with materials that aim to achieve the NGSS vision of three-dimensional science education. Analysis of drafts of the unit using Project 2061's research-based textbook evaluation criteria (AAAS, 2005) and EQuIP's adaptation of them (Achieve, Inc., 2016) provided additional guidance on what the final product should look like (Roseman, Fortus, Krajcik, & Reiser, 2015; Roseman, Herrmann-Abell, & Kruse, 2016). Our findings from the use of the embedded assessments suggest that the same design heuristics can also guide the development of measures to evaluate teacher knowledge and practice.

Acknowledgments

We would like to thank Project 2061 staff members Ana Cordova, Jean Flanagan, Martin Fernandez, Bernard Koch, and Caitlin Klein for their work on the Toward High School Biology project. We are also grateful for the many contributions of the staff at BSCS, who worked in partnership with AAAS during the first three years of the project (2010 through 2013); these include Janet Carlson, Rhiannon Baxter, Brooke Bourdélat-Parks, Elaine Howes, Rebecca Kruse, Stacey Luce, Chris Moraine, Kathleen Roth, and Kerry Skaradzinski. We would also like to thank the many excellent teachers and their students in Colorado, the District of Columbia, Maryland, and Massachusetts who participated in pilot and field testing the unit. During the final phase of the project, Rebecca Kruse (currently at the National Science Foundation) worked with Project 2061 staff in scoring student responses to the embedded assessment tasks.

Funding

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through grant R305A100714 to the American Association for the Advancement of Science. The opinions expressed are those of the authors and do not represent views of the institute or the U.S. Department of Education. The work of contributor Rebecca Kruse (2014 to 2016) was funded by the National Science Foundation Independent Research/ Development Program. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the National Science Foundation.

ORCID

Mary Koppal http://orcid.org/0000-0003-3988-423X

References

Achieve, Inc. (2014). *Educators evaluating the quality of instructional products (EQuIP) rubric for lessons and units: Science* (Version 2.0). Retrieved from http://www.nextgenscience.org/sites/ngss/files/EQuIP%20Rubric%20for%20Science%20052714.pdf

Achieve, Inc. (2015). *Next Generation Science Standards: Primary evaluation of essential criteria for alignment.* Retrieved from http://www.nextgenscience.org/sites/default/files/Draft_PEEC-Alignment%20May%202015.pdf

Achieve, Inc. (2016). *Educators evaluating the quality of instructional products (EQuIP) rubric for lessons and units: Science* (Version 3.0). Retrieved http://www.nextgenscience.org/resources/equip-rubric-lessons-units-science

American Association for the Advancement of Science. (1989). *Science for all Americans*. New York, NY: Oxford University Press.

American Association for the Advancement of Science. (2005). *Project 2061 curriculum materials analysis procedure.* Retrieved from http://www.project2061.org/publications/textbook/hsbio/report/crit-used.htm

American Association for the Advancement of Science. (In press). *Toward high school biology*. Arlington, VA: National Science Teachers Association Press.

Ball, D. L., & Cohen, D. K. (1996). Reform by the book: What is—or might be—the role of curriculum materials in teacher learning and instructional reform? *Educational Researcher*, *25*(9), 6–8, 14.

Banilower, E. R., Nelson, M. M., Trygstad, P. J., Smith, A. A., & Smith, P. S. (2013). *Instructional materials to support the Next Generation Science Standards: Results of a proof-of-concept study*. Retrieved from http://www.horizon-research.com/aim/wp-content/uploads/AIM-NARST-2013-paper.pdf

Beyer, C., & Davis, E. (2009). Investigating teacher learning supports in high school biology curricular programs to inform the design of educative curriculum materials. *Journal of Research in Science Teaching*, *46*, 977–998.

Bismack, A. S., Arias, A. M., Davis, E. A., & Palincsar, A. S. (2015). Examining student work for evidence of teacher uptake of educative curriculum materials. *Journal of Research in Science Teaching*, *52*(6), 816–846.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (2000). *How people learn: Brain, mind, experience, and school.* Washington, DC: National Academies Press.

Brown, J., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, *18*(4), 32–42.

Clements, D. H. (2007). Curriculum research: Toward a framework for "research-based curricula." *Journal for Research in Mathematics Education*, *38*(1), 35–70.

Collins, A., Brown, J. S., & Newman, S. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 453–494). Hillsdale, NJ: Erlbaum.

Davis, E. A., & Krajcik, J. (2005). Designing educative curriculum materials to promote teacher learning. *Educational Researcher*, *34*(3), 3–14.

DeBoer, G. E., Herrmann-Abell, C. F., Gogos, A., Michiels, A., Regan, T., & Wilson, P. (2008). Assessment linked to science learning goals: Probing student thinking through assessment. In J. Coffey, R. Douglas, & C. Stearns (Eds.), *Assessing student learning: Perspectives from research and practice* (pp. 231–252). Arlington, VA: NSTA Press.

Drake, C., Land, T. J., & Tyminski, A. M. (2014). Using educative curriculum materials to support the development of prospective teachers' knowledge. *Educational Researcher*, *43*(3), 154–162.

Driver, R., Squires, A., Rushworth, P., & Wood-Robinson, V. (1994). *Making sense of secondary science: Research into children's ideas*. New York, NY: Routledge.

Flanagan, J. C., Herrmann-Abell, C. F., & Roseman, J. E. (2013, April). *Developing and evaluating an eighth grade curriculum unit that links foundational chemistry to biological growth. Paper 5: Using teacher measures to evaluate the promise of the intervention*. Paper presented at the annual international conference of the National Association for Research in Science Teaching, Rio Grande, Puerto Rico.

Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal*, *38*, 915–945.

Herrmann-Abell, C. F., Koppal, M., & Roseman, J. E. (2016). Toward high school biology: Helping middle school students understand chemical reactions and conservation of mass in non-living and living systems. *CBE-Life Sciences Education*, *15*(4). doi:10.1187/cbe.16-03-0112

Herrmann-Abell, C. F., & Roseman, J. E. (2016). *Comparing the year 3 and year 4 versions of Toward High School Biology*. Retrieved from http://www.aaas.org/sites/default/files/THSB%20Research%20Summary%202016.pdf

Horizon Research, Inc. (2013). *2012 National Survey of Science and Mathematics Education: Highlights report*. Chapel Hill, NC: Author.

Jacobs, V., Lamb, L., & Philipp, R. (2010). Professional noticing of children's mathematical thinking. *Journal for Research in Mathematics Education*, *41*(2), 169–202.

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

Kersting, N. (2008). Using video clips of mathematics classroom instruction as item prompts to measure teachers' knowledge of teaching mathematics. *Educational and Psychological Measurement*, *68*, 845–861.

Kesidou, S., & Roseman, J. E. (2002). How well do middle school science programs measure up? Findings from Project 2061's curriculum review study. *Journal of Research in Science Teaching*, *39*, 522–549.

Kruse, R., Howes, E. V., Carlson, J., Roth, K., & Bourdélat-Parks, B. (2013, April). *Developing and evaluating an eighth grade curriculum unit that links foundational chemistry to biological growth. Paper 3: Designing professional development to support teaching*. Paper presented at the annual international conference of the National Association for Research in Science Teaching, Rio Grande, Puerto Rico.

Land, T. J., Tyminski, A. M., & Drake, C. (2015). Examining pre-service elementary mathematics teachers' reading of educative curriculum materials. *Teaching and Teacher Education*, *51*, 16–26.

McNeill, K. (2008). Teachers' use of curriculum to support students in writing scientific arguments to explain phenomena. *Science Education*, *93*(2), 233–268.

McNeill, K. L., & Krajcik, J. (2008). Scientific explanations: Characterizing and evaluating the effects of teachers' instructional practices on student learning. *Journal of Research in Science Teaching*, *45*(1), 53–78.

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common core state standards (English Language Arts Standards: Science and technical subjects, grade 6–8)*. Washington, DC: Authors. Retrieved from http://www.corestandards.org/ELA-Literacy/RST/6-8/

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

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

National Research Council. (2015). *Guide to implementing the Next Generation Science Standards*. Washington, DC: National Academies Press.

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

Nicol, C. C., & Crespo, S. M. (2006). Learning to teach with mathematics textbooks: How preservice teachers interpret and use curriculum materials. *Educational Studies in Mathematics*, *62*(3), 331–355.

Pashler, H., Bain, P., Bottge, B., Graesser, A., Koedinger, K., McDaniel, M., & Metcalfe, J. (2007). *Organizing instruction and study to improve student learning* (NCER 2007-2004). Retrieved from the Institute of Education Sciences website: http://ies.ed.gov/ncee/wwc/pdf/practiceguides/20072004.pdf

Reiser, B.J., (2013), September. *What professional development strategies are needed for successful implementation of the Next Generation Science Standards?* Paper presented at the Invitational Research Symposium on Science Assessment, Washington, DC.
Remillard, J. T. (2005). Examining key concepts in research on teachers' use of mathematics curricula. *Review of Educational Research*, *75*(2), 211–246.
Roseman, J. E., Fortus, D., Krajcik, J., & Reiser, B. J. (2015, April). *Curriculum materials for Next Generation Science Standards: What the science education research community can do*. Retrieved from http://www.aaas.org/sites/default/files/2015_Roseman_Fortus_Krajcik_Reiser_NARST.pdf
Roseman, J. E., Herrmann-Abell, C. F., & Kruse, R. (2016, April). *Integrating NGSS core ideas and practices: Supporting and studying teachers' implementation*. Poster presented at the annual meeting of the American Educational Research Association, Washington, DC.
Roseman, J. E., Stern, L., & Koppal, M. (2010). A method for analyzing the coherence of biology textbooks: A study of its application to the topic of matter and energy transformations in four textbooks. *Journal of Research in Science Teaching*, *47*(1), 47–70.
Roth, K. J., Druker, S. L., Garnier, H. E., Lemmens, M., Chen, C., Kawanaka, T., … Gallimore, R. (2006). *Teaching science in five countries: Results from the TIMSS 1999 Video Study* (NCES 2006-011). Washington, DC: Government Printing Office.
Rowan, B., Correnti, R., & Miller, R. J. (2002). *What large-scale, survey research tells us about teacher effects on student achievement: Insights from the prospectus study of elementary schools* (CPRE Research Report Series RR-051). Philadelphia, PA: University of Pennsylvania.
Sadler, P. M., Sonnert, G., Coyle, H. P., Cook-Smith, N., & Miller, J. L. (2013). The influence of teachers' knowledge on student learning in middle school physical science classrooms. *American Educational Research Journal*, *50*, 1020–1049.
Santagata, R., Zannoni, C., & Stigler, J. (2007). The role of lesson analysis in pre-service teacher education: An empirical investigation of teacher learning from a virtual video-based field experience. *Journal of Mathematics Teacher Education*, *10*(2), 123–140.
Schneider, R., & Krajcik, J. (2002). Supporting science teacher learning: The role of educative curriculum materials. *Journal of Science Teacher Education*, *13*(3), 221–245.
Schneider, R. M., Krajcik, J., & Marx, R. W. (2000). The role of educative curriculum materials in reforming science education. In B. J. Fishman & S. F. O'Connor-Divelbiss (Eds.), *Proceedings of the Fourth International Conference of the Learning Sciences* (pp. 54–61). Mahwah, NJ: Erlbaum.
Schwarz, C., Gunckel, K., Smith, E., Covitt, B., Bae, M., Enfield, M., & Tsurusaki, B. (2008). Helping elementary preservice teachers learn to use curriculum materials for effective science teaching. *Science Education*, *92*(2), 345–377.
Schwarz, C., Reiser, B., Davis, B., Kenyon, L., Acher, A., Fortus, D., … Krajcik, J. (2009). Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners. *Journal for Research in Science Teaching*, *46*, 632–654.
Sherin, M. G., & van Es, E. A. (2005). Using video to support teachers' ability to notice classroom interactions. *Journal of Technology and Teacher Education*, *13*, 475–491.
Stern, L., & Roseman, J. E. (2004). Can middle school science textbooks help students learn important ideas? Findings from Project 2061's curriculum evaluation study: Life science. *Journal of Research in Science Teaching 41*, 538–568.