

# Toward High School Biology: Helping Middle School Students Understand Chemical Reactions and Conservation of Mass in Nonliving and Living Systems

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

Project 2061, American Association for the Advancement of Science, Washington, DC 20005

## ABSTRACT

Modern biology has become increasingly molecular in nature, requiring students to understand basic chemical concepts. Studies show, however, that many students fail to grasp ideas about atom rearrangement and conservation during chemical reactions or the application of these ideas to biological systems. To help provide students with a better foundation, we used research-based design principles and collaborated in the development of a curricular intervention that applies chemistry ideas to living and nonliving contexts. Six eighth grade teachers and their students participated in a test of the unit during the Spring of 2013. Two of the teachers had used an earlier version of the unit the previous spring. The other four teachers were randomly assigned either to implement the unit or to continue teaching the same content using existing materials. Pre- and posttests were administered, and the data were analyzed using Rasch modeling and hierarchical linear modeling. The results showed that, when controlling for pretest score, gender, language, and ethnicity, students who used the curricular intervention performed better on the posttest than the students using existing materials. Additionally, students who participated in the intervention held fewer misconceptions. These results demonstrate the unit's promise in improving students' understanding of the targeted ideas.

## INTRODUCTION

While all of science has become more interdisciplinary in nature, modern biology is perhaps the most integrative. Cutting-edge research in biology requires knowledge and expertise from physics, chemistry, computer science, engineering, and mathematics to pursue answers to some of society's most challenging problems of food, energy, health, and climate (National Research Council [NRC], 2009). A more integrated approach to science teaching and learning is also central to recommendations in the NRC's (2012) *A Framework for K–12 Science Education (Framework)* and in the *Next Generation Science Standards (NGSS)*; NGSS Lead States, 2013), which is based on the *Framework*. Envisioning a progression of learning across three dimensions: 1) disciplinary core ideas, 2) science and engineering practices, and 3) crosscutting concepts. These documents call for students to develop an integrated understanding of core ideas from physical, life, and earth sciences and engineering and ideas that transcend disciplinary boundaries and be able to use them along with science and engineering practices to make sense of phenomena and solve problems across these disciplines.

Integration of the three dimensions of science learning presents multiple challenges to curriculum developers and teachers alike, not least of which are the persistent problems that many students have in applying basic physical science ideas about chemical reactions to phenomena that occur in life science contexts (Anderson *et al.*, 1990; Marmaroti and Galanopoulou, 2006). Students' misconceptions about these phenomena have been well documented in the science education research literature, as reviewed by Driver *et al.* (1985), Driver *et al.* (1994), Andersson (1990), and

Hannah Sevia, *Monitoring Editor*

Submitted March 4, 2016; Revised June 6, 2016;  
Accepted June 11, 2016

CBE Life Sci Educ December 1, 2016 15:ar74

DOI:10.1187/cbe.16-03-0112

Conflict of interest statement: The authors of this paper were involved in the development of the *Toward High School Biology* unit referenced in this study. No promotion of this curriculum to the exclusion of other similar products should be construed.

\*Address correspondence to: Cari F. Herrmann-Abell (cabell@aaas.org).

© 2016 C. F. Herrmann-Abell *et al.* CBE—Life Sciences Education © 2016 The American Society for Cell Biology. This article is distributed by The American Society for Cell Biology under license from the author(s). It is available to the public under an Attribution–Noncommercial–Share Alike 3.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/3.0>).

“ASCB®” and “The American Society for Cell Biology®” are registered trademarks of The American Society for Cell Biology.

Krnel *et al.* (1998) and confirmed in more recent assessment research by the American Association for the Advancement of Science (AAAS, 2015) and others.

Currently available curricular materials are unlikely to help students develop the chemical foundation that is needed for building new knowledge in biology. In studies of textbook quality across more than a decade (e.g., Morse, 2001; Kesidou and Roseman, 2002; Stern and Roseman, 2004; AAAS, 2005), researchers found that most textbooks pay insufficient attention to students' prior knowledge and misunderstandings, use representations that reinforce common misconceptions, and present too few phenomena to connect abstract ideas to the real world. Most textbooks in these studies offer little guidance for helping students make sense of their experiences with phenomena or understand that a seemingly diverse array of phenomena can be explained by a small number of interrelated ideas.

In a follow-up review of materials published after its 2005 evaluation of high school biology textbooks, AAAS found little evidence of improvement in more recent biology materials (Roseman and Klein, unpublished data). While some newer textbooks provide teachers with a list of possible misconceptions, these are often general in nature, rarely linked to specific phenomena students could experience, and not accompanied by questions or activities designed to help teachers probe their students' ideas more deeply or to help students overcome their misconceptions. In addition, AAAS found other problems that make it difficult for students to construct a coherent understanding of essential chemistry concepts and apply them to life science phenomena:

- Most materials are organized around topics, such as physical or chemical change, chemical equations, conservation of mass, and balancing chemical equations in physical science and photosynthesis, cellular respiration, and chemical digestion in life science, rather than organizing ideas according to a coherent content story line that is developed from the perspective of novice learners. While the topic-based organization may make sense to scientists who already have a conceptual framework for organizing new information, it may not help novices organize new information or recognize how ideas fit together. Presenting details in the context of seemingly isolated topics may make it difficult for students to relate those details to the fundamental ideas. Studies have shown that such seductive detail negatively impacts learning (e.g., Harp and Mayer, 1998; Sanchez and Wiley, 2006).
- Materials rarely integrate physical and life sciences examples when presenting chemical reactions. The only life science chemical reactions consistently shown are the equations for photosynthesis and cellular respiration. These are often used as examples of balanced equations or to emphasize the role of energy. No attempt is made to help students understand the role of chemical reactions in producing substances needed to build body structures for the growth of organisms.
- Most importantly, materials rarely, if ever, engage students in using ideas and practices to make sense of phenomena or challenge their misconceptions. For example, students are not asked to examine data that provide evidence that new substances are produced during chemical reactions, to model atom rearrangement, or to explain why atom

rearrangement results in the production of substances with properties different from those of the starting substances.

With the increased student expectations called for in the NGSS, it is more important than ever for curriculum developers to address problems like these. To do so, developers must begin to draw not only on the recommendations of the new standards themselves but also on findings from multiple areas of research that can inform the curriculum development process (Clements, 2007). According to Clements, "describing and categorizing possible research bases for curriculum development and evaluation is a necessary first step" toward counteracting an approach to curricular design that has emphasized market research above all else (p. 36).

As recommended in Clements' curriculum research framework (CRF), developers can now draw on more than two decades of cognitive science research to gain insights into how people learn across content areas (NRC, 2000; Pashler *et al.*, 2007; Deans for Impact, 2015) and to distill fundamental cognitive principles for guiding the development of curricular materials in every subject area. For science learning in particular, research on specific misconceptions, as noted earlier, can shed additional light on likely student difficulties. Theoretical and empirical work on learning progressions can provide guidance on integrating content from different science domains with appropriate science practices and on sequencing activities over time (AAAS, 2001, 2007; Corcoran *et al.*, 2009; NRC, 2012). Finally, a curriculum development framework, such as that proposed by Clements (2007) can help developers coordinate the multiple lines of research that contribute to and are undertaken as part of the curriculum development process.

To respond to the need for a more effective approach to helping middle school students build a strong conceptual foundation for high school biology, our team of researchers and curriculum developers designed a 6-week replacement unit that takes advantage of existing knowledge and addresses the vision of NGSS. We hypothesized that students would improve their understanding when given opportunities to 1) observe and analyze data related to physical and life sciences phenomena that explicitly target the ideas to be learned and address specific misconceptions and 2) interpret and explain those phenomena in light of relevant science ideas and crosscutting concepts about atom rearrangement and conservation. A field test conducted in year 3 of the development project used a randomized control trial to compare outcomes for students who used the *Toward High School Biology (THSB)* unit with outcomes for a matched comparison group. This paper reports on those results and provides data to answer the following questions:

1. To what extent does the *THSB* unit improve students' understanding of the targeted science ideas when compared with the business-as-usual curriculum?
2. To what extent and in what ways does the *THSB* unit decrease students' misconceptions related to the targeted science ideas?

We also discuss the curricular design principles that guided the development of the unit, consider implications of the project's findings that may have wider applications for other developers, and point to promising directions for further research.

## A THEORETICAL FRAMEWORK FOR CURRICULAR DESIGN

Clements (2007) contends that the isolation of curriculum development and educational research is one reason curricular materials have not improved and describes a multiphase CRF that couples the two. According to Clements' CRF, initial curricular design is based on the coherence of the subject matter and on both general and subject matter-specific research on how students learn, which Clements refers to as "a priori foundations" and "learning model." The initial design is then subjected to multiple rounds of testing in which data collected on classroom feasibility and student learning informs revisions, which Clements refers to as "evaluation." The process exemplifies the design research approach proposed by Brown (1992) and Collins (1992) as a way to carry out research to test and refine educational designs based on prior research.

### A Priori Foundations

The first three phases of the CRF focus on 1) a subject matter a priori foundation that guides the selection of the content to be covered by the unit, 2) a general a priori foundation that draws on learning theories to establish general goals and directions for the unit, and 3) a pedagogical a priori foundation that informs the development of unit activities.

The subject matter covered by the *THSB* unit consists of science ideas that are central to the domains of physical and life sciences, based on their inclusion in AAAS's *Science for All Americans* (1989) and *Benchmarks for Science Literacy* (1993, 2009) and, more recently, the NRC's *Framework* (2012) and the NGSS (NGSS Lead States, 2013). In these documents, the selection of domain-specific ideas was based on the consensus of scientists across disciplines about the knowledge that would be most important for making sense of the natural world and serve as a lasting foundation on which to build more knowledge over a lifetime. The ideas targeted in the middle school unit are age-appropriate steps, based on their inclusion for eighth grade students in both *Benchmarks* and NGSS. In addition, the *Framework* and NGSS emphasize the importance of having students use science practices as they make sense of phenomena relevant to the science ideas targeted.

Regarding the general a priori foundation, the design of the *THSB* unit was influenced by constructivist and metacognitive theories of learning. Constructivist theory posits that students' existing knowledge is used to build new knowledge (e.g., Piaget, 1954; Vygotsky, 1978). Each chapter in the *THSB* unit starts with phenomena in both nonliving and living systems that are directly observable and then engages students in using Lego and ball-and-stick models to make sense of their observations in terms of invisible events, namely atom rearrangement and conservation. Metacognitive theory posits that student engagement increases with increasing knowledge of and ability to monitor their learning (Brown, 1975; Flavell, 1979). Each lesson in the *THSB* unit elicits students' initial ideas and actively engages students in reflecting on how their ideas have changed.

The design of the *THSB* unit's pedagogical strategies were grounded in conceptual change, situated cognition, and knowledge-transfer theories of instruction. Conceptual change is the theory that knowledge develops over time within a specific domain as students reconcile naïve ideas having limited explanatory power with more robust scientific ideas (Posner *et al.*,

1982). The *THSB* unit engages students in observing and making sense of phenomena that challenge common misconceptions about where new substances come from and why changes in mass do not violate conservation principles. Knowledge transfer relates to the idea that 1) concepts transfer when students are given opportunities to apply them in multiple contexts and 2) generalizations (two or more concepts stated in a relationship that experts have found to possess explanatory power) can help organize information within and across contexts (Perkins and Salomon, 1988). *THSB* organizes and sequences lessons to present a coherent content story line of interrelated ideas about changes in matter in physical and life sciences, expresses the ideas making up the story line so as to reinforce connections among related ideas, engages students in generalizing across contexts before being introduced to the "correct" science ideas and then in using the science ideas to explain familiar and then novel phenomena. Situated cognition is learning to use ideas in context through cognitive apprenticeship with a master (Brown *et al.*, 1989; Collins *et al.*, 1989). This includes strategies such as having an expert explicitly model the use of knowledge, coaching novices as they practice using the knowledge, scaffolding students' practice, having students make their thinking visible to expose and clarify thinking, and decreasing support as students' become more competent in their use of knowledge. *THSB* engages students in analyzing data to provide evidence for the production of new substances during chemical reactions; modeling atom rearrangement and conservation to make sense of why new substances are formed; and reasoning from evidence, science ideas, and modeling activities to explain phenomena.

### Learning Model

This category focuses on the domain-specific models of learning. Activities within the unit are consistent with empirically based models of students' thinking and learning, and the sequence of activities is based on learning trajectories. The *THSB* lessons and activities are sequenced to progress 1) from less to more sophisticated data analysis, 2) from using simple to more complex models and modeling tasks, and 3) from contexts involving simpler to more complex systems. Data analysis in *THSB* begins with direct observations about the production of substances with different properties from the starting substances, then progresses to observations of changes in measured properties, and then to inferences from patterns in data. Modeling in *THSB* starts with modeling chemical reactions involving simple molecules that can easily be modeled with Legos and progresses to those producing complex molecules needed to build plant and animal body structures that can only be modeled with sophisticated ball-and-stick models. *THSB* starts with chemical reactions involving pure substances in simple physical systems in which the production of new substances from starting substances can be directly observed and then progresses to chemical reactions occurring in the bodies of living organisms in which inferences from radioactive-labeling experiments are required to document the production of new substances from starting substances.

### Evaluation

This category involves the collection of empirical evidence to evaluate the curriculum. It includes market research, formative

research, and summative research. The *THSB* project focused on formative and summative research. In the first year of the project, the curricular materials were tested with small groups of students and whole classes, with the research staff coteaching. Classroom feasibility tests were conducted in year 2 to see whether teachers could carry out and manage the activities (Herrmann-Abell *et al.*, 2013; Kruse *et al.*, 2013). The results of these tests were used to inform revisions to the unit. The summative research includes the small-scale randomized control trial described in this paper. Future studies would involve large-scale summative research as described in phase 10 (Clements, 2007, pp. 52–53).

### Design Principles for the Development of *THSB*

Four design principles emerged from the learning research. First, related to the subject matter and general a priori foundations, the unit should present a coherent set of science ideas and make explicit the connections among them. Second, related to the general and pedagogical a priori foundations, the unit should take account of students' existing science knowledge and misconceptions. Third and fourth, related to the pedagogical a priori foundation and learning model, the activities in the unit should provide students with experiences with relevant science phenomena and provide support for students' interpretation and explanation of the phenomena. The domain-specific research basis for these design principles is summarized below.

**Design Principle 1: Present a Coherent Set of Science Ideas and Connections among Them.** Because science knowledge is richly interconnected, understanding one idea often depends on understanding a set of prerequisite and related ideas. Nevertheless, research has indicated that, for many students, science knowledge exists as bits and pieces of often naïve and conflicting ideas (Clough and Driver, 1986; diSessa, 1988; diSessa *et al.*, 2002; Demastes *et al.*, 1996; Clark and Linn, 2003) and that students have difficulties in making necessary connections between their observations and relevant science ideas (Bagno and Eylon, 1997) and in making the kinds of inferences needed to fill in gaps found in science texts that lack coherence (Best *et al.*, 2005). DiSessa (1988) described how physics students' problem-solving strategies involved piecemeal explanations rather than a coherent theory (pp. 56–60). And work by Chi and Slotta (1993) and Bagno and Eylon (1997) has suggested that students do not spontaneously make connections such as those involved in relating phenomena to relevant scientific ideas; such connections must be brought out explicitly during instruction. Researchers at AAAS identified essential aspects of coherence in science textbooks that include 1) focusing on a set of interrelated and age-appropriate scientific ideas and making the connections among them explicit, 2) clarifying the ideas and connections with effective representations, 3) illustrating the power of the ideas in explaining phenomena, and 4) avoiding the use of unnecessary technical terms or details that are likely to distract students from the main story (Roseman *et al.*, 2010). Several research and development teams applied this lens to look systematically at both the structure and narrative of their materials and how the individual pieces fit together (Heller, 2001; Krajcik *et al.*, 2008). Work by Roth *et al.* (2009) provides evidence for the importance of a coherent story line

for student learning and identifies six key aspects: 1) establishing the learning goal, 2) selecting and sequencing activities based on relevant phenomena and representations that support the learning goal, 3) linking science ideas to the activities, 4) connecting science ideas within and across lessons, 5) adapting learning experiences to students' contributions, and 6) presenting accurate and age-appropriate science content.

**Design Principle 2: Take Account of Students' Existing Science Knowledge and Beliefs.** According to Anderson *et al.* (1990), "students' difficulties in understanding biological processes are rooted in misunderstandings about concepts in the physical sciences, such as conservation of matter and energy, the nature of energy, and atomic-molecular theory [that] were not addressed in instruction" (p. 775). Even those students who appear to grasp the fundamentals of middle school chemistry often exhibit difficulty applying molecular principles to living organisms (DeBoer *et al.*, 2009; Mohan *et al.*, 2009). For many students, the basic life functions of plants and animals seem unrelated to the inert molecular examples given in chemistry class. For example, in an assessment of middle school students' understanding of photosynthesis, Marmaroti and Galanopoulos (2006) found that a great majority of students do not realize that photosynthesis is a chemical reaction. Even when students are able to make the link, past assessment research has shown that students' misconceptions about concepts in the physical sciences, such as atomic-molecular theory and conservation of matter and energy, limit their ability to develop a coherent understanding of the molecular basis of biology (Anderson *et al.*, 1990). More recent assessment studies have shown that misconceptions related to these concepts are prevalent at both the middle and high school levels (Karataş *et al.*, 2013; AAAS, 2015). Table 1 provides a list of the most commonly held misconceptions related to chemical reactions, conservation of mass, and biological growth and, when applicable, the percentage of students selecting distractors aligned to the misconceptions as their answer choices during the AAAS Project 2061 assessment study (DeBoer *et al.*, 2009; AAAS, 2015).

Research has also shown the strength and persistence of many misconceptions about specific biology concepts. For example, fewer than 20% of a national sample of ~3000 middle school students correctly answered items testing the link between matter transformation and growth, and performance on these items did not significantly improve for high school graduates (DeBoer *et al.*, 2009). Additional research has shown that even undergraduate biology majors are likely to hold the same misconceptions as younger students. Coley and Tanner (2015) report that 93% of undergraduate biology majors and 98% of nonmajors agreed with at least one of several biology misconception statements, with nearly half (49%) of biology majors agreeing with the statement that "Plants get their food from the soil." And research conducted by Hartley *et al.* (2011) suggests that the difficulty college students have with biochemical accounts of processes in living systems stems from their long-standing inability to apply appropriate explanatory science principles at levels other than the organismal. Such problems begin well before college and, as Hartley and colleagues argue, are often exacerbated by science textbooks and instruction that fail to support "principle-based scientific reasoning" (p. 65).



**TABLE 1. Commonly held student misconceptions used as distractors during the AAAS Project 2061 assessment study and the percentage of students selecting them**

Misconception	Grades 6–8	Grades 9–12
The atoms of the reactants of a chemical reaction are transformed into other atoms (Andersson, 1986; DeBoer <i>et al.</i> , 2009).	44%	36%
When mold grows in a closed system, the mass of the system must have increased (DeBoer <i>et al.</i> , 2009).	56%	50%
Mass increases during chemical reactions because new atoms are created (DeBoer <i>et al.</i> , 2009).	46%	33%
Mass decreases during chemical reactions because atoms are destroyed (DeBoer <i>et al.</i> , 2009).	39%	32%
Food is either used for energy or eliminated as waste; it is not used to build or repair body parts (Smith and Anderson, 1986; Leach <i>et al.</i> , 1992).	60%	69%
Most of a plant's mass comes from minerals that it takes in from the soil, not from CO <sub>2</sub> from the air (Vaz <i>et al.</i> , 1997).	54%	58%
Cell division alone can account for plant and animal growth (Krüger <i>et al.</i> , 2006; Riemeier and Gropengießer, 2008). <sup>a</sup>	N/A	N/A

<sup>a</sup>Items in the AAAS Project 2061 assessment study did not include distractors that targeted the cell division misconception.

Considerable evidence exists that students' understanding improves when curriculum and instruction take account of and build on students' existing ideas (Lee *et al.*, 1993; Lehrer and Chazan, 1998; White and Frederickson, 1998).

**Design Principle 3: Provide Experiences with Relevant Science Phenomena.** Understanding science means being able to make sense of a wide variety of events and processes in the natural world (phenomena) in terms of a small number of interrelated principles (science ideas). Appropriate phenomena can help students see where science ideas come from and enhance students' sense of the usefulness of those ideas (Champagne *et al.*, 1985; Strike and Posner, 1985; Anderson and Smith, 1987).

Curricular materials can support teaching and learning of science ideas by enabling students to experience phenomena that have been carefully chosen to illustrate a range of relevant real-world events and processes in different contexts. For the *THSB* unit, that means experiencing related phenomena in the context of both the life and physical sciences. Because most students are novice learners in science topics and can learn more readily about things that are tangible and accessible to their senses (Boulanger, 1981; Wise and Okey, 1983; Kyle *et al.*, 1988), the phenomena should, whenever possible, be directly observable by students or require only simple inferences from data.

Phenomena can also play an important role in helping students view science ideas as useful. Curricular materials can begin to support teaching and learning of a coherent set of ideas by including a set of phenomena that have been carefully chosen to efficiently illustrate a range of relevant real-world events and processes. Efficiency can come from focusing on explaining those phenomena that are particularly problematic for students. Surprising phenomena that contradict students' predictions can be helpful in motivating students to consider differences between their own ideas and scientific ideas. Appropriate phenomena can help students see where science ideas come from and enhance their sense of the usefulness of those scientific ideas (Champagne *et al.*, 1985; Strike and Posner, 1985; Anderson and Smith, 1987).

Middle school science ideas often involve phenomena that are difficult to observe and ideas that are abstract. Well-chosen representations—illustrations, tables and graphs, diagrams, models, and simulations—can help to make these phenomena

accessible to students (Champagne *et al.*, 1985; Strike and Posner, 1985; Feltovich *et al.*, 1989). When phenomena involve events that occur at too small a scale to be seen, representations can be used to enlarge the phenomena for students; when phenomena involve events that occur over too short or too long a time frame, representations can be used to slow down or speed up the processes. The representations should be comprehensible and accurately depict salient aspects of the phenomena. Because middle and high school students tend to see models as actual copies of reality (Grosslight *et al.*, 1991), representations should correspond to the real thing as closely as possible, and students should be asked to consider which aspects of the real thing are being represented and which are not (Thagard, 1992). Multiple representations can increase the likelihood of engaging a range of students (Ainsworth, 1999) and may be useful in helping students to distinguish critical attributes of the phenomena or processes being represented from irrelevant attributes of the representation. Indeed, instruction using multiple modes of linked representations was shown to be more effective in promoting students' understanding of the particulate nature of matter than was comparable instruction without multiple representations (Adadan *et al.*, 2009).

**Design Principle 4: Support Students' Interpretation and Explanation of the Science Phenomena.** Constructing explanations of phenomena is considered to be an important learning goal for its own sake as well as a means by which students can improve their understanding. It is one of the science practices articulated in the NRC's *Framework* and in *NGSS*:

Asking students to demonstrate their own understanding of the implications of a scientific idea by developing their own explanations of phenomena, whether based on observations they have made or models they have developed, engages them in an essential part of the process by which conceptual change can occur. (NRC, 2012, pp. 68–69)

But having students experience phenomena and representations on their own is rarely sufficient to promote understanding of fundamental science ideas. Students need to be actively engaged in thinking about the phenomena and representations and interpreting them in light of basic principles of science

(Driver, 1983; Anderson and Smith, 1984). With guidance, students are able to consider the strengths and limitations of representations as models of the real world, consider how their own ideas compare in explanatory power to scientific ideas, connect new ideas to what they already know, and link related ideas into a coherent story.

To foster this process, instruction should be organized around a range of structured tasks that are designed to help students relate the phenomena and representations to the science ideas, reconcile their own ideas with the science ideas, and use the science ideas to explain other relevant phenomena (Eaton *et al.*, 1984; Minstrell, 1984; Osborne and Freyberg, 1985; McDermott, 1991; Roth, 1991). Carefully chosen and sequenced questions can be particularly powerful in supporting students' sense making (Anderson and Smith, 1987; Anderson and Roth, 1989; Arons, 1990). Activities that make students' thinking about experiences and ideas overt to themselves, to the teacher, and to other students can allow those ideas to be examined, questioned, and shaped (Needham, 1987; Clement, 1993; Linn and Burbules, 1993; Glaser, 1994; Flick, 1995; Roth, 1996). Work by Sandoval (2003) and Sandoval and Reiser (2004) points to the interaction between conceptual learning and the practice of explanation and demonstrates the need for scaffolding, such as an explanatory framework, to guide students in developing more complete and evidence-based explanations and in evaluating the quality of their explanations.

To verify that these design principles were indeed manifest in the *THSB* unit and that it was well aligned with the vision of *NGSS*, we analyzed the unit at multiple stages using 1) criteria developed by AAAS Project 2061 for analyzing the alignment of curricular materials to standards and the quality of their instructional support (AAAS, 2005) and 2) the Educators Evaluating the Quality of Instructional Products (EQuIP) rubric for evaluating the fit of science materials to *NGSS* (Achieve, 2014). Data from these analyses provided important formative feedback to the curricular design over several cycles of development, classroom trials, and revisions (Roseman *et al.*, 2013, 2015, 2016).

## METHODS

Guided by the curricular design principles described above, the research team developed the *THSB* unit, revised it based on findings from pilot tests in diverse classrooms, and then compared it with the "business-as-usual" curriculum in a randomized control trial (RCT). The RCT was part of a field test of the unit in the Spring of 2013 in two districts in the Mid-Atlantic United States. This paper reports on the results from a study of one of those districts.

In year 1 of the effort, a "backward-design" strategy was used to develop an initial draft of the student materials (Wiggins and McTighe, 2005). Following an iterative design process, the draft student materials were pilot tested by researchers in a small number of schools, and data from the pilot test were used to revise the student materials and to develop teacher resources and professional development. In year 2, the revised materials and formal professional development were implemented by classroom teachers in six schools. The purpose of this round of testing was to examine student learning gains and the feasibility of implementing the curricular materials in a range of classrooms, using data collected to inform revisions. In

year 3 of the development process, the curriculum and professional development materials were revised to address issues that surfaced during pilot testing.

## Content Focus of the *THSB* Unit

The *THSB* unit tested in year 3 consists of 20 lessons organized into four chapters. As recommended in the *NRC Framework* and *NGSS*, the *THSB* unit addresses core disciplinary ideas, science practices, and crosscutting concepts in the context of making sense of relevant phenomena in nonliving and living systems. The overarching goal of the *THSB* unit is for students to use ideas about what happens to atoms and molecules during chemical reactions to explain growth and repair in living things.

Chapter 1 develops the central concept that, during chemical reactions, the atoms that make up the starting substances rearrange to form molecules of the new substances with different properties (ideas 1 and 2 in Table 2). Chapter 2 develops the concept that, regardless of how atoms are rearranged during a chemical reaction, the number of each type of atom stays the same and the mass of each atom stays the same; therefore, the total mass stays the same (idea 3 in Table 2). Chapter 3 applies the concepts of atom rearrangement and conservation to animal growth and repair (idea 4 in Table 2), and chapter 4 applies these concepts to plant growth and repair (idea 5 in Table 2).

**Disciplinary Core Ideas and Crosscutting Concepts.** The science ideas to which the *THSB* unit and assessments were aligned are shown in Table 2. These statements were adapted from grade band endpoints for grade 8 articulated in sections PS1.A, PS1.B, and LS1.C from the *NRC's Framework* (see Supplemental Table S1). The crosscutting concept Energy and Matter: Flows, Cycles, and Conservation is addressed by the unit as well, specifically the concept that "matter is conserved

**TABLE 2. Science ideas targeted by the *THSB* unit**

1. Pure substances are made from a single type of atom or molecule; each pure substance has characteristic properties that can be used to identify it. (from PS1.A)
2. Many substances react chemically in characteristic ways. In a chemical reaction, the atoms that make up the molecules of the original substances are regrouped into different molecules, and these new substances have different properties from those of the starting substances. (from PS1.B)
3. The total number of each type of atom is conserved during chemical reactions, and thus the mass does not change. If the measured mass changes, it is because atoms have entered or left the system. (from PS1.B)
4. Animals obtain food from eating plants or eating other animals. Within individual organisms, food moves through a series of chemical reactions in which the molecules that make up food are broken down and the atoms are rearranged to form new molecules to support growth. (from LS1.C)
5. Plants make glucose from CO<sub>2</sub> from the atmosphere and water through a chemical reaction that releases oxygen. Within individual organisms, glucose molecules undergo chemical reactions in which the atoms that make up the glucose molecules are rearranged to form new molecules to support growth. (from LS1.C)

because atoms are conserved in physical and chemical processes” (NGSS Lead States, 2013, appendix G). All of these science ideas are also found in the grade 7 or grade 8 state science curriculum for the state in which this study was conducted.

Traditionally, these life and physical sciences ideas are taught separately, with the life science being taught in the seventh grade and the physical science in the eighth grade. The *THSB* unit takes a different approach, treating ideas about chemical reactions and conservation of matter in both living and nonliving contexts together and taught by the same teacher, so that connections among the ideas can be readily made. This integration of physical and life sciences content is consistent with recommendations of the NGSS in general and in particular with its crosscutting concepts related to matter and energy.

**Science Practices.** The *THSB* unit focuses on five of the eight science practices recommended by the NRC’s *Framework*: 1) analyzing and interpreting data; 2) developing and using models; 3) constructing explanations; 4) engaging in argument from evidence; and 5) obtaining, evaluating, and communicating information (NRC, 2012).

#### Design of the *THSB* Unit

The following section discusses the ways in which the *THSB* unit embodies the curricular design principles described above.

**Design Principle 1: Present A Coherent Set of Science Ideas and Connections among Them.** The science ideas listed in Table 2 were broken down into smaller ideas that unfold in a coherent content story line as the *THSB* unit progresses (Roseman *et al.*, 2013). Figure 1 shows a map of these ideas. The map is similar to maps in the *Atlas of Science Literacy* (AAAS, 2001, 2007) in the way that it lists ideas in text boxes, represents connections among ideas with arrows, and displays how more sophisticated ideas (at the top of map) might develop from less sophisticated ideas (at the bottom of map). However, the map in Figure 1 differs from *Atlas* maps in its purpose and, therefore, its curricular specificity. The map in Figure 1 includes only the ideas and arrows that indicate connections that are relevant to the *THSB* unit. For example, science ideas 12–16 are targeted in chapter 3 of the unit after students have encountered prerequisite science ideas 1, 3, 6, and 7 in chapter 1. Furthermore, the prerequisite ideas shown in the map are somewhat curriculum specific; other curricular materials might reverse the order of ideas in some cases.

**Design Principle 2: Take Account of Students’ Existing Knowledge and Beliefs.** Many of the phenomena and modeling activities included in the unit were selected because they contradict commonly held student misconceptions presented in Table 1. For example, to challenge the misconception that food is either used for energy or eliminated as waste, students examine data showing that 20% of radioactively labeled carbon atoms from the brine shrimp they eat (food) become incorporated into the bodies of young herring fish that eat the food. And to challenge misconceptions that atoms are created, destroyed, or changed into other atoms during chemical reactions, students build Lego models of reactant molecules

and then rearrange the same bricks to form models of the product molecules.

The set of misconceptions targeted were chosen because past research has documented their prevalence with middle school students and their persistence into high school. The list of misconceptions was also included in the teacher edition of the *THSB* unit so that teachers were made aware of them, how they might be manifested in student work, and the activities designed to address them. Each lesson asked students to respond to a key question that gave the students the opportunity to present their initial ideas and activate their thinking and provided the teacher with information about students’ naïve ideas and potential learning difficulties. At the end of each lesson, students revisit the key question and reflect on how their thinking has changed. The teacher edition also provided teachers with follow-up questions they could use to probe and challenge student ideas during small-group and whole-class discussions.

**Design Principle 3: Provide Experiences with Relevant Science Phenomena.** The *THSB* unit includes a wide range of phenomena that students observe and make sense of throughout the unit. Tasks and questions within each lesson are designed to ensure that students make the intended observations, guide students to relate the instances they observed to the generalizations in the science ideas, and then apply the science idea to novel contexts.

The unit starts with phenomena in which the production of substances with different properties can be directly observed or at least require minimal inferences from data and moves toward phenomena that require more sophisticated inferences from data. Because the link between reactants and products is not obvious in living systems, in which hundreds of reactions are occurring simultaneously, the developers also took advantage of the rich scientific literature using radioactively labeled atoms to determine the products of a chemical reaction. After analyzing the data to provide evidence for the reactants and products, students use model-based reasoning to make sense of how the products could have been produced from the atoms making up the reactant molecules.

All phenomena were initially tested with students for engagement and comprehensibility and then to see whether students could use the practices to make sense of the phenomena. It is important to note that, although the *THSB* unit introduces students to a variety of phenomena involving chemical reactions—some of them quite complex—the purpose of these phenomena is to illustrate specific science ideas and their application across physical and life sciences contexts. Students are not presented with and are not expected to learn every detail about every reaction. The *THSB* unit specifies in the student and teacher materials exactly what its goals for student learning are, and these are the goals that students are held accountable for in the unit’s assessments.

The set of phenomena used to develop the science ideas is shown in Table 3. The iron and oxygen and hexamethylenediamine and adipic acid phenomena are physical science phenomena that are introduced in the first two chapters and then used as analogies to life science phenomena such as growth in the last two chapters. The vinegar and baking soda phenomenon is used to start students thinking about the role of gases in chemical reactions.

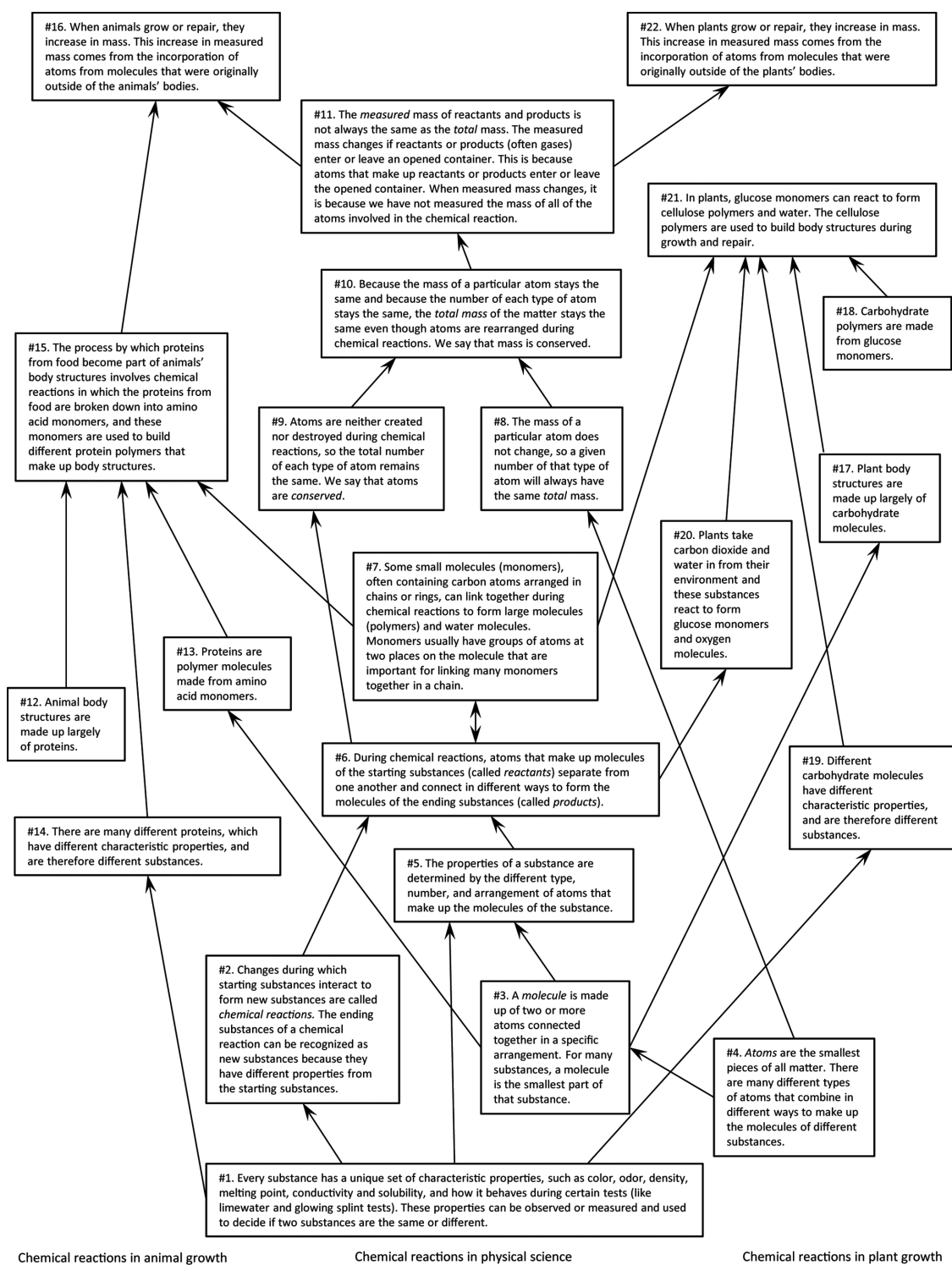


FIGURE 1. Map of science ideas targeted by the THSB unit.

The lessons within each chapter of the unit also engage students in modeling underlying molecular events, particularly atom rearrangement and conservation during chemical reactions, using Lego bricks and ball-and-stick models, and in relating these to conventional two-dimensional representations such as space-filling models and chemical and structural

formulas. Using and relating multiple models helps students appreciate critical attributes of molecules and to make abstract ideas about atom rearrangement and conservation concrete. For example, using Lego bricks to model the formation of rust in an open container allows students to see that atoms are conserved even though the measured mass increases, which



TABLE 3. Key phenomena for each *THSB* chapter

Chapter number and science ideas	Students observe, model, and explain these phenomena
1. New substances form during chemical reactions because atoms rearrange to form new molecules.	Why substances with different properties form when: <ul style="list-style-type: none"> <li>• Vinegar is mixed with baking soda</li> <li>• Iron is exposed to oxygen in the air</li> <li>• Hexamethylenediamine is mixed with adipic acid</li> </ul>
2. Mass is conserved in chemical reactions because atoms are conserved.	Why the measured mass of a system can change even though atoms are not created or destroyed when: <ul style="list-style-type: none"> <li>• Vinegar is mixed with baking soda</li> <li>• Iron is exposed to oxygen in the air</li> <li>• Hexamethylenediamine is mixed with adipic acid</li> </ul>
3. Animals build body structures for growth through chemical reactions, during which atoms rearrange and are conserved.	How animals produce proteins for growth of their body structures that are different from what they eat when: <ul style="list-style-type: none"> <li>• Egg-eating snake eats only eggs but can replace its shed skin</li> <li>• Humans eat muscles but can also make tendons</li> <li>• Herring fish eat <math>^{14}\text{C}</math>-labeled brine shrimp and make <math>^{14}\text{C}</math>-labeled body structures (mostly muscle)</li> </ul>
4. Plants build body structures for growth through chemical reactions, during which atoms rearrange and are conserved.	How plants produce carbohydrates for growth of their body structures that are different from substances they take in from their environment when: <ul style="list-style-type: none"> <li>• Algae produce <math>^{14}\text{C}</math>-glucose from <math>^{14}\text{C}</math>-carbon dioxide and they produce <math>^{18}\text{O}</math>-oxygen (not <math>^{18}\text{O}</math>-glucose) from <math>^{18}\text{O}</math>-water</li> <li>• Mouse-ear cress plants make more <math>^{14}\text{C}</math>-cellulose from <math>^{14}\text{C}</math>-glucose when grown without herbicide than with it</li> </ul>

prepares students for making sense of increases in mass that accompany biological growth. Using the same models across chemical reactions in living and nonliving systems highlights the crosscutting concept of atom rearrangement and conservation underlying phenomena in both disciplines.

**Design Principle 4: Support Students' Interpretation and Explanation of the Science Phenomena.** The *THSB* unit provides scaffolds that familiarize students with using evidence and reasoning from science ideas and models to explain phenomena and justify their explanations. Initially, students are introduced to the parts of a complete explanation, which were adapted from the claim, evidence, reasoning framework (McNeill and Krajcik, 2012), and they examine examples of explanations. In the next explanation activity, students use a table to help them organize their thinking and writing and to remind them about the necessary elements. The scaffolds fade as the unit progresses, and by the end of the unit, students are expected to write an explanation without a table or reminders about the elements.

#### Business-as-Usual Curriculum

Teachers in the comparison group followed the business-as-usual curriculum for the school district in which the study took place. The district's middle school science program is aligned with the state curriculum in science, which expects students to develop an understanding of the science ideas included in Table 2 by grade 8. Teachers were expected to choose and teach 6 weeks of activities that are aligned with the science ideas and practices. At the end of the unit, comparison-group teachers provided the research team with a summary of the activities their students completed that aligned with the science ideas included in Table 2 and, when possible, copies of those lessons. One teacher relied heavily on activities from a physical

science textbook published in 2001. The other teacher designed her own worksheets and labs.

An analysis of the teachers' summaries showed that, while the activities were topically aligned to the science ideas, the instructional strategies used differed considerably from the strategies used in the *THSB* unit. First, while students in the comparison group did use molecule kits and make Lewis dot models of metal ions, the modeling activities were not used to make sense of phenomena. Second, while students did observe examples of chemical reactions, the focus was mainly on writing balanced chemical equations rather than making sense of observations about mass conservation and changes in measured mass. Third, the business-as-usual curriculum made few explicit connections between physical and life sciences ideas and phenomena. While the eighth grade chemical reactions unit engages students in balancing equations in both physical and life sciences contexts (e.g., iron rusting and the reaction between baking soda and vinegar in physical science, photosynthesis and cellular respiration in life science), the equation balancing does not contribute to explaining any phenomena. Finally, lessons in the business-as-usual curriculum are not organized around a coherent story line; each lesson is on a different topic with no connections made to previous lessons.

#### RCT Study Design

**Research Setting.** Six teachers from six schools in a suburban district participated in the study in the Spring of 2013. Two of the teachers had participated in the year 2 pilot test of the curricular intervention (Spring of 2012) and were returning in year 3 to implement the revised unit with their classes. The classes of these teachers comprise what we will refer to as the "experienced group." Four teachers were new to the project and were matched in pairs based on school characteristics such as

eighth grade state test scores in math and science and student demographic variables such as ethnicity.

One teacher in each pair was randomly assigned to use the intervention with all of his or her classes (from this point forward referred to as the “novice group”), and all of the classes of the other teacher were assigned to the comparison group. Treatment assignment within each pair of schools was done randomly by Abt Associates, who used the Stata statistical software package (StataCorp, 2009) to assign random numbers to each school. Within each pair, the school with the smaller number was assigned to the novice group. In both the experienced and novice groups, the *THSB* unit replaced the students’ usual curricular materials, and the unit’s lessons were taught by the classroom teacher after the teacher participated in 3 days of face-to-face professional development. Regarding completion of the unit, five out of the nine classes in the experienced group completed all of the lessons compared with none of the classes in the novice group. The average number of completed chapters out of four and the range of completion are shown in Table 4 for both groups. The students in the comparison group used the business-as-usual curriculum, which targets the same science ideas shown in Table 2, as described earlier.

**Participants.** A total of 594 students participated in the study, but the data reported here are from the 574 students who completed both the pretest and the posttest and responded to at least 25% of the items on both tests. Student demographic data indicated that 55% of the students were male and 45% were female; ~9% of the students stated that English was not their primary language; and ~43% of the students were white, 16% were African American, 26% were Asian, 8% were Hispanic, and 7% were two or more ethnicities. A breakdown of the demographic data

by group is presented in Table 4 and a breakdown by class is presented in Supplemental Table S2. Data on students’ socioeconomic status were not made available by the school district.

**Student Content Knowledge Test.** To determine how students’ understanding of the targeted learning goals changed as a result of instruction using either the *THSB* unit or the school district curriculum, we administered a test before and after instruction. The items on the test were developed using a procedure designed to ensure the items’ match to the targeted ideas and their overall effectiveness as accurate measures of what students do and do not know about those ideas (DeBoer et al., 2007, 2008a,b). Each item was aligned to one or two of the targeted science ideas or crosscutting concepts listed in Table 2, and item distractors were designed to probe for relevant student misconceptions (Sadler, 1998). As part of the item development procedure, the items were pilot tested with 532 students from another county in the same state in which the study was conducted. A Rasch analysis of the pilot test data was performed, and the item separation reliability was 0.96. The pilot-test data were used to inform revisions to the items and the selection of the items for the final pre/posttest.

The final student content knowledge test included 36 items, which were a mix of distractor-driven multiple-choice items and two-tiered items. Three of the items required students to interpret models of atoms and molecules and four of the items required students to analyze data about substances’ characteristic properties. There were four two-tiered items that consisted of a multiple-choice item followed by two open-response questions. The open-response questions asked the students to explain in writing why they thought the answer choices they selected were correct and why they thought the other answer choices were not correct.

All of the multiple-choice items were scored dichotomously. A rubric was developed for each of the two-tiered items. The students’ written explanations for why they selected or rejected the answer choices were evaluated together against the ideal response included in the rubric. Each response was rated by two researchers, and any disagreements were resolved by consulting a third researcher. When Krippendorff’s alpha (Krippendorff, 2004) was calculated as an estimate of interrater reliability, results before reconciliation showed reliabilities of 0.71, 0.83, 0.83, and 0.92. The students received one score for these two-tiered items that summed their scores on the multiple-choice part and the written explanation.

**Description of Rasch Modeling.** Student-level scale scores were created using Rasch modeling (Liu and Boone, 2006; Boone et al., 2014; Bond and Fox, 2007). The “partial credit” model was used because the test included both dichotomous and polytomous items (Masters, 1982). When the data fit the Rasch model, the student scale scores and item difficulties are expressed on the same interval scale, are mutually independent, and are measured in the unit of logarithm called log odds or logits, which can vary from  $-\infty$  to  $+\infty$ . Winsteps Rasch measurement software was used to estimate student scale scores and item difficulties (Linacre, 2013). The control variable ISGROUPS was set to zero, which indicates that each item has its own response structure. The average item difficulty was set at zero.

**TABLE 4. Summary of class and student-level variables**

	Comparison	<i>THSB</i> novice	<i>THSB</i> experienced
Number of classes	9	10	9
Gifted and talented classes	67%	50%	44%
Number of students	196	194	184
Average pretest score (logits)	-0.15	-0.45	-0.70
Chapters of <i>THSB</i> completed			
Range	0	2.8–3.3	3.2–4
Mean	0	2.9	3.7
Gender			
Male	56%	55%	55%
Female	44%	45%	45%
Ethnicity			
White	45%	41%	42%
Asian	27%	29%	22%
Black	14%	11%	23%
Hispanic	9%	10%	6%
Two or more ethnicities	6%	9%	7%
Primary language			
English	89%	92%	93%
Other	11%	8%	7%

**Measuring Change Using Rasch Modeling.** In this paper, we apply the stacking method to the pretest and posttest data (Wright, 2003). Stacking allowed us to create two scale scores per person: a pretest score and a posttest score. The stacked analysis was done by first preparing a data file that contained two rows of data per student. One row contains their responses during the pretest and the second row contains their responses during the posttest. This analysis results in two measures per student: a pretest scale score and a posttest scale score. The difference between these scale scores represents the change in the students' understanding as a result of instruction.

**Hierarchical Linear Modeling.** Once the pretest and posttest scale scores were created using Rasch modeling, the posttest scale scores were modeled as outcome measures in two-level hierarchical linear models (HLMs) with students at level 1 and classes at level 2. Classes were used at level 2 instead of teachers, because we had evidence that student posttest scores varied between classes of the same teacher. Student-level variables included pretest scale scores, gender, ethnicity, and language. Class-level variables included whether or not the class was part of the novice group of a matched pair, whether or not the class was part of the experienced group, and whether or not the class was designated as a gifted and talented class. Using an intent-to-treat approach, all classes were included in the analyses regardless of whether or not they completed all of the lessons in the *THSB* unit.

A fully unconditional model containing only the posttest outcome variable and no independent variables, except an intercept, was estimated first. This was followed by a conditional model in which pre-test scale score, gender, language, and ethnicity were included as controls and modeled as fixed effects. HLM 7 software was used in this study (Raudenbush *et al.*, 2011). The method of estimation was restricted maximum likelihood. Effect sizes were calculated by dividing the coefficient by the square root of the pooled student-level unadjusted SD.

## RESULTS

### Rasch Fit

The fit statistics presented in Table 5 show how accurately the stacked field-test data fit the Rasch model. The separation indices and corresponding reliabilities were 16.66 and 1.00 for the items and 2.67 and 0.88 for the students. Both of the separation indices are considered acceptable—that is, greater than 2, according to Wright and Stone (2004). Additionally, the SEs for the items and students were small (see Table 5). The infit and outfit mean-square values for the majority of the items and students were within the acceptable range of 0.7–1.3 for multiple-choice tests (Bond and Fox, 2007). Because the fit

statistics are within the acceptable ranges, we conclude that the data have a good fit to the Rasch model.

### Fully Unconditional HLM

A fully unconditional HLM with no independent variables at either level was run to calculate the intraclass correlation coefficient. The results of the model are shown in Table 6. The intraclass correlation coefficient represents the proportion of variance in posttest scores that could be the result of class characteristics, such as the curriculum used. In this case, almost half (48%) of the variance in posttest score could be the function of class characteristics. Therefore, the proportion of the variance in posttest scores that exists at the individual level is 52%. A chi-square test indicated that posttest scores varied significantly between classes ( $\chi^2 = 482.34, p < 0.001$ ).

### Conditional HLM

The mixed-model for the conditional HLM is

$$\begin{aligned} \text{POSTTEST}_{ij} = & \gamma_{00} + \gamma_{01} * \text{NOVICE}_j + \gamma_{02} * \text{EXPERIENCED}_j + \\ & \gamma_{03} * \text{GT}_j + \gamma_{10} * \text{PRETEST}_{ij} + \gamma_{20} * \text{FEMALE}_{ij} + \gamma_{30} * \text{BLACK}_{ij} + \\ & \gamma_{40} * \text{HISPANIC}_{ij} + \gamma_{50} * \text{ASIAN}_{ij} + \gamma_{60} * \text{2ORMORE}_{ij} + \\ & \gamma_{70} * \text{ENGLISH}_{ij} + u_{0j} + r_{ij} \end{aligned}$$

where  $\text{POSTTEST}_{ij}$  and  $\text{PRETEST}_{ij}$  are the post- and pretest scale scores for the student  $i$  within class  $j$ , respectively. GT is a dummy variable indicating whether or not the class is designated as a gifted and talented class. Two dummy variables were created for the instruction used in the class; NOVICE is a dummy variable indicating whether or not the teacher of the class was a first-year implementer of the *THSB* unit; EXPERIENCED is a dummy variable indicating whether or not the teacher of the class was an experienced implementer of the *THSB* unit. The comparison group, which was using the business-as-usual curriculum, was used as a reference group. FEMALE is a dummy variable indicating the gender of student  $i$  in class  $j$  (female = 1; male = 0). Four dummy variables were created for ethnicity (BLACK, HISPANIC, ASIAN, and 2ORMORE), and white was used as a reference group. ENGLISH is a dummy variable indicating whether or not English is the primary language of student  $i$  in class  $j$  (English = 1; other language = 0). All of the student-level variables were grand-mean centered and all of the class-level variables were uncentered. The terms  $u_{0j}$  and  $r_{ij}$  are the error terms associated with the classes and students, respectively. The results of the conditional HLM are shown in Table 7.

According to the coefficients shown in Table 7, the average posttest score for students in the non-GT comparison group

TABLE 5. Rasch fit statistics for the stacked data

	Item			Person		
	Minimum	Maximum	Median	Minimum	Maximum	Median
SE	0.03	0.10	0.07	0.32	1.04	0.35
Infit mean-square	0.82	1.23	0.98	0.19	4.85	0.98
Outfit mean-square	0.70	1.53	0.97	0.05	9.90	0.93
Point-measure correlation coefficients	0.26	0.68	0.46	−0.28	0.78	0.46
Separation index (reliability)		16.66 (1.00)			2.67 (0.88)	

**TABLE 6. Fully unconditional HLM**

Variable	Value
Within-classroom variance ( $\sigma^2$ )	0.86
Between-classroom variance ( $\tau$ )	0.80
Between-classroom SD	0.95
Reliability ( $\lambda$ )	0.94
Intraclass correlation ( $\rho$ )	0.48

classes is  $-0.31$  logits (controlling for pretest score, gender, ethnicity, and primary language). The coefficients for the NOVICE and EXPERIENCED variables indicate that, on average, students in the non-GT classes in the novice group score  $0.92$  logits higher than students in the non-GT comparison classes, and students in the non-GT classes in the experienced group score  $1.19$  logits higher than students in the non-GT comparison classes. Therefore, the average posttest score for students in the non-GT novice classes is  $0.61$  logits and the average posttest score for students in the non-GT experienced group classes is  $0.88$  logits (controlling for pretest score, gender, ethnicity, and primary language). Compared with the business-as-usual curriculum, the effect size for the *THSB* unit being implemented for the first time (novice group) is  $0.84$ , and the effect size for the *THSB* unit being implemented by teachers with prior experience with *THSB* (experienced group) is  $1.10$ . Additionally, the model shows that being in a GT class increases the posttest score by  $0.41$  logits.

Because not all of the classes completed the entire *THSB* unit, a second conditional HLM model was run. In this model, the number of chapters completed was used as a level 2 variable instead of the NOVICE and EXPERIENCED dummy variables. The mixed model is

$$\begin{aligned} \text{POSTTEST}_{ij} = & \gamma_{00} + \gamma_{01} * \text{CHAPTER}_{ij} + \gamma_{02} * \text{GT}_{ij} + \\ & \gamma_{10} * \text{PRETEST}_{ij} + \gamma_{20} * \text{FEMALE}_{ij} + \gamma_{30} * \text{BLACK}_{ij} + \\ & \gamma_{40} * \text{HISPANIC}_{ij} + \gamma_{50} * \text{ASIAN}_{ij} + \gamma_{60} * \text{2ORMORE}_{ij} + \\ & \gamma_{70} * \text{ENGLISH}_{ij} + u_{0j} + r_{ij} \end{aligned}$$

**TABLE 7. Results from the conditional HLM**

Fixed effects	Coefficient	SE	t Ratio	Approximate df	p Value
Class-level variables					
Intercept, $\gamma_{00}$	$-0.31$	$0.14$	$-2.26$	$24$	$0.03$
Novice, $\gamma_{01}$	$0.92$	$0.15$	$3.13$	$24$	$<0.001$
Experienced, $\gamma_{02}$	$1.19$	$0.15$	$7.86$	$24$	$<0.001$
GT, $\gamma_{03}$	$0.41$	$0.13$	$3.13$	$24$	$0.005$
Individual-level variables					
Pretest, $\gamma_{10}$	$0.80$	$0.05$	$17.81$	$539$	$<0.001$
Female, $\gamma_{20}$	$0.04$	$0.06$	$0.79$	$539$	$0.43$
Black, $\gamma_{30}$	$-0.37$	$0.10$	$-3.90$	$539$	$<0.001$
Hispanic, $\gamma_{40}$	$-0.21$	$0.12$	$-1.75$	$539$	$0.08$
Asian, $\gamma_{50}$	$0.11$	$0.08$	$1.38$	$539$	$0.17$
Two or more, $\gamma_{60}$	$-0.12$	$0.12$	$-1.06$	$539$	$0.29$
English, $\gamma_{70}$	$0.23$	$0.12$	$1.99$	$539$	$0.05$
Random effects	SD	Variance	df	$\chi^2$	p Value
Intercept, $u_0$	$0.27$	$0.08$	$29$	$107.74$	$<0.001$
level-1, $r$	$0.68$	$0.46$			

All of the student-level variables were grand-mean centered, and all of the class-level variables were uncentered. The results of this conditional HLM are shown in Table 8.

According to the coefficients from the second conditional HLM model shown in Table 8, the average posttest score for students in the non-GT comparison group classes is  $-0.30$  logits (controlling for pretest score, gender, ethnicity, and primary language). The coefficients for the CHAPTER variable indicate that, on average, students in the non-GT classes using *THSB* score  $0.32$  logits higher than the students in the non-GT comparison classes for each chapter of *THSB* they complete. Therefore, the average score for non-GT students who completed two chapters of *THSB* is  $0.34$  logits. Non-GT students who completed three chapters of *THSB* score, on average,  $0.66$  logits, and non-GT students who completed all four chapters score, on average,  $0.98$  logits.

### Analyzing Distractors for Misconceptions

An analysis of students' selection of distractors in the pre- and posttest items was performed to gain insight into the effects the *THSB* unit had on students' misconceptions related to the targeted science ideas. As discussed earlier, past research has identified several misconceptions that students hold about chemical reactions and growth. Many of the distractors of the items on the pre- and posttest targeted these misconceptions. Looking at the frequency with which these distractors were selected provides more detailed information about how student thinking in each group changed after receiving instruction. The following summarizes results of distractor analyses focused on common student misconceptions.

**Misconception: Atoms Are Transmuted during Chemical Reactions.** As shown in Table 1, a common misconception about chemical reactions is that, during a reaction, the atoms that make up the reactants are transformed into different types of atoms (Andersson, 1986; DeBoer et al., 2009). Five items on the pre- and posttest included distractors that probed this misconception. One item included two distractors aligned to this misconception, and the other four items include one distractor



TABLE 8. Results from the conditional HLM with the number of chapters completed as a class-level variable

Fixed effects	Coefficient	SE	t Ratio	Approximate df	p Value
Class-level variables					
Intercept, $\gamma_{00}$	-0.30	0.12	-2.53	25	0.02
Chapter, $\gamma_{01}$	0.32	0.03	9.56	25	<0.001
GT, $\gamma_{03}$	0.38	0.12	3.21	25	0.004
Individual-level variables					
Pretest, $\gamma_{10}$	0.80	0.04	17.88	539	<0.001
Female, $\gamma_{20}$	0.04	0.06	0.77	539	0.44
Black, $\gamma_{30}$	-0.37	0.09	-3.90	539	<0.001
Hispanic, $\gamma_{40}$	-0.21	0.12	-1.75	539	0.08
Asian, $\gamma_{50}$	0.11	0.08	1.44	539	0.15
Two or more, $\gamma_{60}$	-0.13	0.12	-1.09	539	0.28
English, $\gamma_{70}$	0.24	0.12	2.03	539	0.04
Random effects	SD	Variance	df	$\chi^2$	p Value
Intercept, $u_0$	0.24	0.06	25	88.89	<0.001
level-1, $r$	0.68	0.46			

aligned to this misconception. Table 9 shows the frequency at which these distractors were selected on the pre- and posttests along with the overall percent correct on these items. On the pretest, this misconception was very popular; the selection of these distractors represented more than half of the incorrect responses on these items. Overall, these distractors were selected on the pretest 31% of the time by students in comparison classrooms and 33% of the time by students who used the *THSB* unit (in classrooms of both novice and experienced users;  $\chi^2(1) = 1.51$ , n.s.; Cramer's *V* effect size = 0.02). On the posttest, the percentages decreased to 23% for the comparison group students and 14% for students who used the unit. The posttest percentage for *THSB* users is significantly lower than the percentage for the comparison group ( $\chi^2(1) = 30.72$ ,  $p < 0.001$ ; Cramer's *V* effect size = 0.11).

**Misconception: Atoms Are Created during Chemical Reactions.** It is well known that students have difficulty predicting that mass will be conserved, especially for systems in which there appears to be an increase or decrease of "stuff" (Mitchell and Gunstone, 1984; Hesse and Anderson, 1992; Lee *et al.*, 1993; DeBoer *et al.*, 2009). Students may explain this apparent increase or decrease by saying that atoms can be created or destroyed during chemical reactions, a common misconception as shown in Table 1. Three items in the pre- and posttest each included one distractor aligned to the misconception that new atoms are created during chemical reactions; two items each included one distractor aligned to the misconception

that atoms are destroyed, and one item included three distractors that aligned to misconceptions about the destruction and/or the creation of atoms misconceptions. The results from these items are summarized in Table 10.

As indicated in Table 10, on the pretest, distractors involving atoms being created were selected 26% of the time by the students in the comparison group and 31% of the time by students who used the *THSB* unit (in classrooms of both novice and experienced users;  $\chi^2(1) = 5.19$ ,  $p < 0.05$ ; Cramer's *V* effect size = 0.05). On the posttest, the percentages dropped to 16% for the comparison group and 21% for the students using the *THSB* unit ( $\chi^2(1) = 7.43$ ,  $p < 0.01$ ; Cramer's *V* effect size = 0.06). Distractors involving matter being destroyed were selected 15% of the time on the pretest by the students in the comparison group and 20% of the time by students who used the *THSB* unit ( $\chi^2(1) = 6.77$ ,  $p < 0.05$ ; Cramer's *V* effect size = 0.07). On the posttest, the percentages dropped to 9% for the comparison group and 8% for the students using the unit ( $\chi^2(1) = 0.03$ , n.s.; Cramer's *V* effect size = 0.01).

**Misconception: Food Is Excreted as Waste and Does Not Become Part of the Body.** Past research has shown that some students do not think that any of the food animals eat becomes part of the animals' bodies; instead, they think the food is either eliminated as waste or used for energy (Smith and Anderson, 1986; Leach *et al.*, 1992). Two items tested the idea that some of the food becomes part of the body. One of these items included two distractors that included the misconception that

TABLE 9. Frequency of selecting the correct answer and distractors targeting the misconception that atoms are transmuted (based on six distractors in five items)

Answer choice <sup>a</sup>	Comparison				THSB			
	Pretest	Posttest	$\chi^2$ (p value)	Effect size	Pretest	Posttest	$\chi^2$ (p value)	Effect size
<i>Atoms are rearranged.</i>	47%	59%	24.95 ( $<0.001$ )	0.17	37%	65%	288.72 ( $<0.001$ )	0.28
<i>Atoms are changed into other atoms.</i>	31%	23%	14.31 ( $<0.001$ )	0.09	33%	14%	175.82 ( $<0.001$ )	0.22

<sup>a</sup>The correct answer choice is in italics.

**TABLE 10. Frequency of selecting the correct answer and distractors targeting the misconceptions that atoms can be created or destroyed (based on seven distractors in six items)**

Answer choice <sup>a</sup>	Comparison				THSB			
	Pretest	Posttest	$\chi^2$ (p value)	Effect size	Pretest	Posttest	$\chi^2$ (p value)	Effect size
<i>Atoms are neither created nor destroyed.</i>	52%	62%	23.21 ( $<0.001$ )	0.10	41%	55%	86.25 ( $<0.001$ )	0.14
Atoms are created.	26%	16%	24.45 ( $<0.001$ )	0.13	31%	21%	39.97 ( $<0.001$ )	0.12
Atoms are destroyed.	15%	9%	9.77 ( $<0.01$ )	0.09	20%	8%	65.59 ( $<0.001$ )	0.17

<sup>a</sup>The correct answer choice is in italics.

none of the food becomes part of the body and the other item included four distractors. The results of these items are shown in Table 11. On the pretest, the distractors were selected 34% of the time by the students in the comparison group and 37% of the time by students who used the THSB unit (in classrooms of both novice and experienced users;  $\chi^2(1) = 1.25$ , n.s.; Cramer's  $V$  effect size = 0.04). On the posttest, the percentages dropped to 22% for the comparison group and 11% for the students using the unit. The posttest percentage for the THSB users is significantly lower than the percentage for the comparison group ( $\chi^2(1) = 23.82$ ,  $p < 0.001$ ; Cramer's  $V$  effect size = 0.15).

**Misconception: Cell Division Alone Can Account for Growth.**

Some students think that living organisms grow merely because the cells that make up their bodies divide, not because the organisms take in additional matter that becomes part of their bodies (Krüger *et al.*, 2006; Riemeier and Gropengießer, 2008). Six items on the pre- and posttest each included one distractor aligned to this misconception. Table 12 shows the frequency at which these distractors were selected and the overall percent correct on the pre- and posttests. On the pretest, these distractors were selected 34% of the time by students in the comparison group and 28% of the time by students who used the THSB unit (in classrooms of both novice and experienced users;  $\chi^2(1) = 12.92$ ,  $p < 0.001$ ; Cramer's  $V$  effect size = 0.06). On the posttest,

the percentage decreased to 26% for the comparison group and 7% for the novice and experienced groups. The posttest percentage for the THSB users is significantly lower than the percentage for the comparison group ( $\chi^2(1) = 244.12$ ,  $p < 0.001$ ; Cramer's  $V$  effect size = 0.27).

**Misconception: Most of Plants' Mass Comes from Minerals.**

Studies have shown that students have difficulty accepting that most of the mass of a plant comes from  $\text{CO}_2$  in the air. They commonly believe that the mass comes from minerals in the soil (Vaz *et al.*, 1997), mostly because they think that gases have negligible mass (Mas *et al.*, 1987) and therefore cannot contribute significantly to the mass of a tree. There were four items that each included one distractor aligned to this misconception, and the results from these items are presented in Table 13. The table presents the answer choice selections for the novice and experienced groups separately, because the activities that targeted this misconception were part of the lessons that the novice group did not complete. The distractors were selected on the pretest 42% of the time by the comparison group, 50% of the time by the novice group, and 41% of the time by the experienced group ( $\chi^2(2) = 12.63$ ,  $p < 0.01$ , Cramer's  $V$  effect size = 0.08). On the posttest, the comparison and novice groups did not show a significant decrease in percentage of times these distractors were selected (see Table 13). The experienced

**TABLE 11. Frequency of selecting the correct answer and distractors targeting the misconception that food does not become part of the body (based on six distractors in two items)**

Answer choice <sup>a</sup>	Comparison				THSB			
	Pretest	Posttest	$\chi^2$ (p value)	Effect size	Pretest	Posttest	$\chi^2$ (p value)	Effect size
<i>Atoms from food become part of the body.</i>	60%	74%	15.82 ( $<0.001$ )	0.15	56%	87%	170.65 ( $<0.001$ )	0.34
Atoms from food do not become part of the body.	34%	22%	13.39 ( $<0.001$ )	0.14	37%	11%	144.30 ( $<0.001$ )	0.31

<sup>a</sup>The correct answer choice is in italics.**TABLE 12. Frequency of selecting the correct answer and distractors targeting the misconceptions that cell division alone can account for growth (based on six distractors in six items)**

Answer choice <sup>a</sup>	Comparison				THSB			
	Pretest	Posttest	$\chi^2$ (p value)	Effect size	Pretest	Posttest	$\chi^2$ (p value)	Effect size
<i>Incorporation of atoms from food accounts for growth.</i>	45%	54%	16.92 ( $<0.001$ )	0.09	43%	68%	271.67 ( $<0.001$ )	0.25
Cell division accounts for growth.	34%	26%	16.02 ( $<0.001$ )	0.09	28%	7%	349.38 ( $<0.001$ )	0.28

<sup>a</sup>The correct answer choice is in italics.

**TABLE 13.** Frequency of selecting the correct answer and distractors targeting the misconceptions that plants' mass comes from minerals (based on four distractors in four items)

Answer choice <sup>a</sup>	Comparison				THSB novice				THSB experienced			
	Pretest	Posttest	$\chi^2$ (p value)	Effect size	Pretest	Posttest	$\chi^2$ (p value)	Effect size	Pretest	Posttest	$\chi^2$ (p value)	Effect size
<i>Plants' mass does not come from minerals.</i>	32%	40%	10.82 ( $<0.01$ )	0.09	31%	39%	10.91 ( $<0.01$ )	0.09	30%	71%	244.17 ( $<0.001$ )	0.41
Plants' mass comes from minerals.	42%	39%	1.19 (n.s.)	0.03	50%	50%	0.01 (n.s.)	0.00	41%	16%	116.93 ( $<0.001$ )	0.29

<sup>a</sup>The correct answer choice is in italics.

group, however, showed a very large decrease from 41% on the pretest to 16% on the posttest.

## DISCUSSION

### Research Question 1

To evaluate the overall promise of the *THSB* unit in increasing students' understanding of the targeted science ideas, we modeled the posttest scale scores as outcomes in HLMs. The results of the conditional HLM models indicate that the *THSB* unit shows great promise in increasing students' understanding of ideas related to chemical reactions in living and nonliving systems. When dummy variables for the novice and experienced groups were used, the results show that the novice group significantly outperformed the comparison group on the posttest, and the experienced group significantly outperformed the novice group. The effect sizes for both the novice group and experienced group compared with the comparison group are considered to be large (i.e.,  $>0.80$ ; Cohen, 1988).

The larger effect size for the experienced group (1.10) may be due to teachers' increased familiarity and comfort with the *THSB* unit and to their increased completion rate. Feedback received from the teachers in the experienced group at the end of the study suggested that during the year 3 implementation of the *THSB* unit, they better understood the story line of the unit and the science content and were more comfortable facilitating students' use of the molecular models. More of the classes of the experienced group teachers completed the unit, and the HLM model using the number of chapters completed did show that higher posttest scores are associated with higher completion rates. As shown in Table 8, there is a 0.32 logit increase in posttest scores for each chapter completed.

These results provide evidence that the *THSB* unit has promise in increasing students' understanding of foundational science ideas and their ability to use those ideas to explain phenomena. They also suggest that the unit may require multiple implementations by a teacher before reaching its full potential.

### Research Question 2

In answer to the second research question about the *THSB* unit's promise in reducing students' misconceptions, finer-grained analyses of students' answer choice selections for specific sets of items revealed differences in the extent to which students in the different groups had changed their thinking after instruction. We also identified the design principles that guided the selection of phenomena and activities aimed at helping students overcome specific misconceptions.

**Atoms Are Rearranged Not Transmuted during Chemical Reactions.** On the pretest, the transmutation misconception that atoms are changed into different atoms was prevalent in all of the groups, and distractors aligned to this misconception were selected almost a third of the time. While both the comparison and *THSB* groups showed a decrease in the frequency of selection of the transmutation distractors from pre- to posttest, the *THSB* group showed a significantly larger decrease, as indicated by the larger effect size (see Table 9). This indicates that the *THSB* unit was more successful at reducing this misconception than the business-as-usual curriculum.

This success may be due to students' experiences with the modeling activities during which the unit pushes students to notice that the products of chemical reactions they model are made from only the atoms that made up the starting substances. The students had experiences with a variety of chemical reactions in both nonliving and living systems throughout the *THSB* unit (see *Design Principles 1* and *3*). For most of these reactions, students built models of the reactant molecules, rearranged the "atoms" to build models of the product molecules, and were asked to consider what happened to the numbers and types of "atoms." At no point during the modeling of the reaction did the students have to go back to the box of models and exchange "atoms" for other types. From this, students could infer that no "real" atoms changed into other types of atoms during these or any other chemical reactions.

**Atoms Are Not Created or Destroyed during Chemical Reactions.** From the distractor analysis of the misconceptions about atoms being created or destroyed during chemical reactions, the groups showed similar reductions in frequency of selection (see Table 10). In this case, the *THSB* unit and the business-as-usual curriculum were equally successful in reducing students' misconceptions about the creation and destruction of atoms.

The phenomena selected for the *THSB* unit purposefully included chemical reactions during which the amount of matter being measured either increases or decreases (see *Design Principle 2*). The modeling activities were used to show students that, even when the measured mass changed, the number of atoms did not change if they took account of atoms entering and/or leaving the system. Students modeled the observed changes in measured mass by weighing Lego models of reactants and products in closed and open systems and noticed that in neither case were any Legos created or destroyed. The unit also included scaffolding to support the students in constructing written explanations of these phenomena (see *Design Principle 4*).

The comparison group, which did not experience the modeling activities, had a similar reduction in misconceptions. It is possible that the assessment items probing this misconception could be correctly answered simply by knowing that “atoms cannot be created or destroyed.” All of these items used the terms “created” and “destroyed” in distractors, which students could eliminate if they had merely memorized the phrase. It is possible that these multiple-choice items were not sufficient probes of students’ conceptual understanding of conservation. However, a difference in the quality of student explanations was observed between treatment and comparison groups. Analysis of students’ written explanations from the two-tiered items showed that students in the *THSB* group were better able to use their understanding of conservation to explain novel phenomena. Almost a third of the students in the novice and experienced groups used ideas about atom rearrangement and conservation in their posttest explanations compared with only 1% of the comparison group. To illustrate the improvement in the quality of explanations of *THSB* users, we show sample pre- and posttest explanations from two students who used the *THSB* unit in Table 14. The item asked the students to predict how the mass of the sealed bag containing a piece of bread would change after mold grew on the bread, and then they were asked to explain their answers. The students who wrote the explanations in Table 14 selected the correct answer choice on both the pre- and posttests: the mass of the sealed bag containing the bread would not change after the mold grew. In each example, the student wrote a substance-level explanation on the pretest and an atomic-level on the posttest.

Nonetheless, a significant percentage of *THSB* students (21%) selected answer choices aligned to the misconception that atoms are created during chemical reactions on the posttest. It is possible that these students are still confused about the distinction between molecules, which are created during chemical reactions, and atoms, which are not. But it is also possible that these students are not yet reasoning with a mental model of atom rearrangement and conservation. As a result, we revised the questions in the conservation activities throughout the unit to press students to acknowledge that, although different molecules were created, atoms were not. We also added suggestions in the teacher edition to have students model their written responses to explanation tasks.

**Food Becomes Part of an Animal’s Body during Growth.** The students who experienced the *THSB* unit were less likely

than the comparison group to think that food an animal eats does not become part of the animal’s body (see Table 11). The misconception that atoms from food do not become part of an animal’s body was equally selected by both groups on the pretest. On the posttest, the misconception was only selected 11% of the time by the students who participated in the *THSB* unit but was still selected by 22% of the comparison group. The effect size for the comparison group is considered to be small (i.e., around 0.10) and the effect size for the *THSB* group is considered to be medium (i.e., around 0.30; Cohen, 1988). These results suggest that the *THSB* unit more effectively targeted this misconception than the comparison curriculum.

The larger reduction of this misconception in the treatment groups may be attributed to carefully sequenced activities that provided evidence from phenomena and reasoning from models to contradict this misconception (see *Design Principles 1, 2, and 3*). Students using the *THSB* unit observed the “growth” of nylon thread and modeled the polymerization reaction to prepare them to make sense of polymerization reactions required for animal and plant growth. Students then examined data on the composition of animal body parts that would serve as evidence that animal bodies are mostly made up of protein polymers that would have to be made for animals to grow. Students examined data showing that the proteins an animal eats have different properties and, hence, are different substances from the proteins making up animal body structures. To provide evidence that animals actually do convert proteins from food into proteins making up their body structures, students 1) examined data from radioactive-labeling experiments showing that herring fish incorporated 20% of  $^{14}\text{C}$  atoms from the brine shrimp they ate into their bodies and 2) modeled the processes of protein digestion and protein synthesis to explain how the incorporation could have occurred. Throughout these activities, students responded to questions to guide them to make the intended observations and link the substance-level observations to atomic/molecular events (see *Design Principle 3*).

**Growth Requires the Incorporation of New Atoms, Not Just Cell Division.** One of the more prevalent misconceptions about both animal and plant growth is the idea that cell division alone explains the growth of organisms. Students holding this misconception seem to view growth merely as “getting bigger,” not as increasing in mass. Students who do not link getting bigger to increasing in mass have no need to account for it. As a result, an explanation of growth that involves chemical reactions and

**TABLE 14. Sample explanations for the moldy bread item from students in the experienced group**

Pretest explanation	Posttest explanation
“The bread chemically changed to mold, but the mass did not change.”	“The bag is a closed container. The total and measured mass stay the same inside closed containers. The atoms that start in the plastic bag cannot change mass or escape. No new atoms can be created, so the mass stays the same.”
“I think the bag weighed the same because nothing could get in or out of the bag, so theoretically the weight should not change.”	“The bag and its contents weighed the same because in the closed container, nothing can get in or out. This means that atoms that make up the bread cannot slip out of the bag, and atoms outside cannot get in, so the weights won’t be changed. The mold absorbed molecules in the bread and, through chemical reactions, rearranged the atoms to incorporate them in the mold. Throughout the process, the number of total atoms in the bag stayed the same, so the measured mass of the bag will stay the same also.”



the incorporation of atoms is unnecessary. Table 12 shows that students who experienced the *THSB* unit selected distractors targeting this misconception <10% of the time. In the comparison group, these distractors were selected more than a quarter of the time after instruction. Based on the effect sizes shown in Table 12, the *THSB* unit had a medium effect (i.e., around 0.30) on this misconception, and the business-as-usual curriculum had a small effect (i.e., around 0.10; Cohen, 1988).

This finding suggests that the *THSB* unit was more effective at convincing the students that cell division alone, without the incorporation of additional atoms, cannot account for the increase in mass that accompanies growth. Modeling of phenomena such as iron rusting that involve mass conservation (in closed systems) and increases in measured mass (in open systems) helped students recognize that the only way for the mass of a system to increase is to add atoms from outside the system (see *Design Principle 3*). This idea is built upon in the chapters on animal and plant growth (see *Design Principle 1*). Students are guided to see that animals and plants are open systems and to recognize that the incorporation of new atoms into an organism's body is essential for an increase in mass and, therefore, growth to occur. The explanation scaffolding also supported students in interpreting and explaining growth phenomena (see *Design Principle 4*).

#### **Plants' Mass Increase Comes from CO<sub>2</sub> Not from Minerals.**

On the plant growth items, only the experienced group showed a decrease in the frequency of selection of the misconception that most of a plant's mass comes from minerals in the soil (see Table 13). The difference in performance on these items between the students of experienced and novice *THSB* users may have been caused by differences in the number of lessons teachers completed: whereas most of the experienced *THSB* users completed all of the lessons targeting plant growth ideas, none of the novice *THSB* users did. The effect size for the experienced group suggests that the unit had a medium effect (i.e., around 0.30) on this misconception (Cohen, 1988).

During the plant growth lessons, which come at the end of the *THSB* unit, students participated in activities that directly contradict the misconception that plants' mass comes from minerals while providing students with evidence for where the material that makes up plants does come from (CO<sub>2</sub> in the air; see *Design Principle 2*). The students were shown data from radioactive-labeling experiments that proved that 1) the carbon and oxygen atoms of glucose molecules in plants come from CO<sub>2</sub> molecules in the air and 2) plants can make cellulose from glucose. Then they modeled the chemical reactions involved, tracing the <sup>14</sup>C and <sup>18</sup>O atoms from CO<sub>2</sub> to glucose and then from glucose to cellulose.

#### **Study Limitations**

There are several limitations to the study. First, the number of participating teachers was small, and they were all from one relatively well-performing district. Therefore, we view this study as an evaluation of the unit's promise and not its efficacy. A larger RCT including a larger and more representative sample is needed to explore the unit's efficacy.

Additionally, not all of the classes completed the entire unit. In fact, none of the classes in the novice group and only five classes in the experienced group finished all of the lessons. It is possible that more improvement in understanding would have

been achieved if the students had been able to experience the unit in its entirety.

Another limitation is that the comparison group instruction devoted less time to the life science contexts and the science practices. This may inflate the effect size due to a difference in time spent on these areas.

Furthermore, although the comparison group teachers received a list of the science ideas, they did not receive professional development on aligning activities to those ideas, whereas the treatment teachers did. In future studies, the addition of professional development on alignment to the science ideas and practices targeted should be provided to both groups.

#### **Implications for Curriculum Development and Implementation**

Although the findings reported in this paper are specific to our experiences with the *THSB* unit, they likely have implications for the design and implementation of other science curriculum materials, particularly those being developed to support the NGSS vision and a more learner-centered approach to science education.

Alignment with the NGSS requires that curricular materials do much more than simply "cover" a set of specified ideas and skills. Some developers and publishers are attempting to modify their materials, while others are already making claims of alignment. To date, however, there has been little guidance available for understanding what it means to align with NGSS or to support students in achieving the NGSS performance expectations. The EQuIP rubric seeks to fill that gap (Achieve, 2014).

The study described in this paper may be one of the first to cite empirical evidence that a curricular material aligned to the NGSS as articulated in the EQuIP rubric has the potential to improve students' understanding of important science ideas and practices. Nevertheless, it is important to note that, while the EQuIP rubric provided formative input to the revision of the *THSB* unit, it was not used in the initial development of the unit, so this study does not address how effectively the rubric on its own would serve as a curricular design tool. As a result, our findings with regard to NGSS alignment suggest that the research-based design principles described in this paper and used to guide the development of the *THSB* unit may have wider applications by other curriculum developers seeking to align their materials to NGSS.

While there is general agreement that the NGSS vision is laudable, most educators acknowledge that it is also highly ambitious and will be challenging to implement. For example, one of the main goals of the NGSS effort was to focus on a smaller set of core ideas so that students could learn them more deeply and use them with a range of science practices to make sense of phenomena across disciplines. After working with many excellent teachers throughout the development and testing of the *THSB* unit, however, it is clear that helping students use core ideas and practices to make sense of phenomena will require much more instructional time than schools currently provide. Indeed, after analyzing data from the year 2 pilot test of the *THSB* unit, we found that, because of the difficulty of the ideas being taught, the time students needed to learn those ideas well and the need to improve the overall coherence and comprehensibility of the unit, it was necessary to streamline the unit (Roseman *et al.*, 2013). This required some

significant design trade-offs, such as focusing on only parts of core ideas and providing fewer phenomena as examples in order to provide students with the additional scaffolding and experience they needed to explain them well. Other developers will make other choices, of course, but the point is that curricular design in the era of NGSS must be more evidence-based than ever and attend closely to the realities of the classroom, such as the amount of instructional time available and the conceptual difficulties that many students are likely to have.

Finally, although this paper does not deal with the teacher support provided in the *THSB* unit, it is clear to us that materials designed to take the NGSS vision seriously will require much more extensive and ongoing support for teachers than has previously been provided. As noted earlier in this paper, teachers' prior experience with the *THSB* unit was a significant variable in predicting student performance. While this is likely to be true of any curriculum, truly addressing the three dimensions of learning called for in NGSS require teachers to take on content and instructional practices for which they have had little preparation. Providing teachers with adequate time to understand new materials and improve their skill in using them to best advantage in their classrooms is essential to the NGSS vision and to all improvements in science education.

## CONCLUSIONS

This paper reports on data from the year 3 field test of a new curricular unit, *Toward High School Biology*, which is designed to help students explain biological growth and repair in terms of atom rearrangement and conservation during chemical reactions. Guided by a set of research-based design principles, the unit was developed to improve on currently available materials and breaks new ground by engaging students in making sense of phenomena that occur in both nonliving and living systems using science ideas, crosscutting concepts, and science practices and supporting their ability to do so. This support includes carefully sequenced data analysis and modeling tasks and scaffolded questions that help students connect phenomena to a coherent set of science ideas, confront differences between their own ideas and science ideas, and relate the science ideas targeted in each lesson to other science ideas and phenomena. This approach aligns well with the three dimensions of learning recommended in the *NRC Framework* and *NGSS*.

A study was conducted to investigate the overall promise of the unit in increasing students' understanding of the science ideas and reducing their misconceptions. Three groups of students were compared during the study: 1) classes of teachers implementing the intervention for the first time (novice group), 2) classes of teachers who had implemented an earlier version of the intervention in the previous year (experienced group), and 3) classes of teachers using the school district curriculum that targets the same science ideas. Rasch modeling was used to create scale scores for both the pre- and posttests. These scale scores were then modeled as outcomes in a two-level HLM to investigate effects of the intervention controlling for pretest score, gender, language, and ethnicity. The results of the model showed a significantly positive correlation between using the *THSB* unit and posttest score. Large effect sizes were found for both the novice group and the experienced group. A distractor analysis showed that the unit was also successful in reducing

the prevalence of commonly held student misconceptions. In most cases, students who took the *THSB* unit were less likely to select misconceptions aligned to distractors than students in the comparison group. This suggests *THSB* was more successful at reducing the misconceptions than the business-as-usual curriculum. These results provide evidence of the promise of the *THSB* unit for increasing students' understanding of chemical reactions and conservation of mass in living and nonliving systems and for the unit's feasibility, which improves in the hands of experienced teachers.

## ACKNOWLEDGMENTS

We acknowledge our Biological Sciences Curriculum Study (BSCS) collaborators in the development of the student and teacher materials and professional development: Janet Carlson, Brook Bourdelat-Parks, Elaine Howes, Rebecca Kruse, Kathy Roth, Aleigh Raffelson, Kerry Skaradzinski, Rhiannon Baxter, Stacey Luce, and Chris Moraine. We thank Project 2061 staff members Jean Flanagan, for her contribution to the development of the curricular unit and the assessments, and Martin Fernandez, Bernard Koch, and Caitlin Klein for their help in scoring the written explanations. We also thank the pilot and field-test teachers for participating in our studies and providing helpful feedback. 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. Visit [www.aaas.org/sites/default/files/THSBSummaryBooklet-F.pdf](http://www.aaas.org/sites/default/files/THSBSummaryBooklet-F.pdf) for more information about the THBS unit. To request electronic copies of the student and teacher editions, contact Project 2061 at [project2061@aaas.org](mailto:project2061@aaas.org).

## REFERENCES

- Achieve (2014). The Educators Evaluating the Quality of Instructional Products (EQulP) Rubric for Science. [www.nextgenscience.org/sites/default/files/EQulP%20Rubric%20for%20Science%20v2.pdf](http://www.nextgenscience.org/sites/default/files/EQulP%20Rubric%20for%20Science%20v2.pdf) (accessed 22 November 2016).
- Adadan E, Irving KE, Trundle KC (2009). Impacts of multi-representational instruction on high school students' conceptual understandings of the particulate nature of matter. *Int J Sci Educ* 31, 1743–1775.
- Ainsworth S (1999). The functions of multiple representations. *Computers Educ* 33, 131–152.
- American Association for the Advancement of Science (AAAS) (1989). *Science for All Americans*, New York: Oxford University Press.
- AAAS (1993). *Benchmarks for Science Literacy*, New York: Oxford University Press.
- AAAS (2001). *Atlas of Science Literacy*, vol. 1, Washington, DC.
- AAAS (2005). *High School Biology Textbooks: A Benchmarks-Based Evaluation*. [www.project2061.org/publications/textbook/hsbio/report](http://www.project2061.org/publications/textbook/hsbio/report) (accessed 22 November 2016)
- AAAS (2007). *Atlas of Science Literacy*, vol. 2, Washington, DC.
- AAAS (2009). *Benchmarks Online*. [www.project2061.org/publications/bsl/online/index.php](http://www.project2061.org/publications/bsl/online/index.php) (accessed 22 November 2016)
- AAAS (2015). *AAAS Science Assessment*. <http://assessment.aaas.org>.
- Anderson CW, Roth KJ (1989). Teaching for meaningful and self-regulated learning of science. In: *Advances in Research on Teaching*, vol. 1, ed. J Brophy, Greenwich, CT: JAI, 265–310.
- Anderson CW, Sheldon T, Dubay J (1990). The effects of instruction on college nonmajors' conceptions of respiration and photosynthesis. *J Res Sci Teach* 27, 761–776.

- Anderson CW, Smith EL (1984). Children's preconceptions and content-area textbooks. In: *Comprehension Instruction: Perspectives and Suggestions*, ed. GG Duffy, LR Rochler, and J Mason, New York: Longman, 187–220.
- Anderson CW, Smith EL (1987). Teaching science. In: *The Educator's Handbook: A Research Perspective*, ed. V Richardson-Koehler, New York: Longman, 84–111.
- Andersson B (1986). Pupils' explanations of some aspects of chemical reactions. *Sci Educ* 70, 549–563.
- Andersson B (1990). Pupils' conceptions of matter and its transformations (age 12–16). In: *Relating Macroscopic Phenomena to Microscopic Particles*, ed. P Lijnse, P Licht, W de Vos, and AJ Waarlo, Utrecht, Netherlands: CD-p Press, 12–35.
- Arons A (1990). *A Guide to Introductory Physics Teaching*, New York: Wiley.
- Bagno E, Eylon B (1997). From problem solving to a knowledge structure: an example from the domain of electromagnetism. *Am J Phys* 65, 726–736.
- Best RM, Rowe M, Ozura Y, McNamara D (2005). Deep-level comprehension of science texts: the role of the reader and the test. *Top Lang Disord* 25, 65–83.
- Bond T, Fox C (2007). *Applying the Rasch Model*, Mahwah, NJ: Erlbaum.
- Boone WJ, Staver JR, Yale MS (2014). *Rasch Analysis in the Human Sciences*, Dordrecht, Netherlands: Springer.
- Boulanger FD (1981). Instruction and science learning: a quantitative synthesis. *J Res Sci Teach* 18, 311–327.
- Brown AL (1975). The development of memory: knowing, knowing about knowing, and knowing how to know. In: *Advances in Child Development and Behavior*, vol. 10, ed. HW Reese, New York: Academic, 103–152.
- Brown AL (1992). Design experiments: theoretical and methodological challenges in creating complex interventions. *J Learn Sci* 2, 141–178.
- Brown JS, Collins A, Duguid S (1989). Situated cognition and the culture of learning. *Educ Res* 18, 32–42.
- Champagne A, Gunstone R, Klopfer L (1985). Instructional consequences of students' knowledge about physical phenomena. In: *Cognitive Structure and Conceptual Change*, ed. L West and AL Pines, Orlando, FL: Academic, 61–90.
- Chi MTH, Slotta JD (1993). The ontological coherence of intuitive physics. *Cogn Instruct* 10, 249–260.
- Clark D, Linn M (2003). Scaffolding knowledge integration through curricular depth. *J Learn Sci* 12, 451–493.
- Clement J (1993). Model construction and criticism cycles in expert reasoning. In: *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society*, ed. Institute of Cognitive Science, University of Colorado–Boulder, Hillsdale, NJ: Erlbaum, 336–341.
- Clements D (2007). Curriculum research: toward a framework for “research based curricula.” *J Res Math Educ* 38, 35–70.
- Clough E, Driver R (1986). A study of consistency in the use of students' conceptual frameworks across different task contexts. *Sci Educ* 70, 473–496.
- Cohen J (1988). *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed., New York: Academic.
- Coley JD, Tanner K (2015). Relations between intuitive biological thinking and biological misconceptions in biology majors and nonmajors. *CBE Life Sci Educ* 14, ar8.
- Collins A (1992). Toward a design science of education. In: *New Directions in Educational Technology*, ed. E Scanlon and T O'Shea, Berlin: Springer, 15–22.
- Collins A, Brown J, Newman S (1989). Cognitive apprenticeship: teaching the crafts of reading, writing, and mathematics. In: *Knowing, Learning, and Instruction: Essays in Honor of Robert Glaser*, ed. LB Resnick, Hillsdale, NJ: Erlbaum, 453–494.
- Corcoran TB, Mosher FA, Rogat A (2009). *Learning Progressions in Science: An Evidence-Based Approach to Reform*, CPRE Research Reports, New York: Center on Continuous Instructional Improvement, Teachers College, Columbia University.
- Deans for Impact (2015). *The Science of Learning*, Austin, TX.
- DeBoer GE, Herrmann-Abell CF, Gogos A (2007). Assessment linked to science learning goals: probing student thinking during item development. Paper presented at the Annual Meeting of the National Association of Research in Science Teaching, held 15–18 April 2007, in New Orleans, LA.
- DeBoer GE, Herrmann-Abell CF, Gogos A, Michiels A, Regan T, Wilson P (2008a). Assessment linked to science learning goals: probing student thinking through assessment. In: *Assessing Student Learning: Perspectives from Research and Practice*, ed. J Coffey, R Douglas, and C Stearns, Arlington, VA: National Science Teachers Association, 231–252.
- DeBoer GE, Herrmann-Abell CF, Wertheim J, Roseman JE (2009). Assessment linked to middle school science learning goals: a report on field test results for four middle school science topics. Paper presented at the Annual Meeting of the National Association of Research in Science Teaching, held 17–21 April 2009, in Garden Grove, CA.
- DeBoer GE, Lee HS, Husic F (2008b). Assessing integrated understanding of science. In: *Coherent Science Education: Implications for Curriculum, Instruction, and Policy*, ed. Y Kali, MC Linn, and JE Roseman, New York: Columbia University Teachers College Press, 153–182.
- Demastes SS, Good RG, Peebles P (1996). Patterns of conceptual change in evolution. *J Res Sci Teach* 33, 407–431.
- diSessa AA (1988). Knowledge in pieces. In: *Constructivism in the Computer Age*, ed. G Forman and P Pufall, Hillsdale, NJ: Erlbaum, 49–70.
- diSessa AA, Elby A, Hammer D (2002). J's epistemological stance and strategies. In: *Intentional Conceptual Change*, ed. GM Sinatra and PR Pintrich, Mahwah, NJ: Erlbaum, 237–290.
- Driver R (1983). *The Pupil as Scientist?*, Milton Keynes, UK: Open University Press.
- Driver R, Guesne E, Tiberghien A (1985). *Children's Ideas in Science*, Milton Keynes, UK: Open University Press.
- Driver R, Squires A, Rushworth P, Wood-Robinson V (1994). *Making Sense of Secondary Science: Research into Children's Ideas*, New York: Routledge.
- Eaton JF, Anderson CW, Smith EL (1984). Student preconceptions interfere with learning: case studies of fifth-grade students. *Elem School J* 64, 365–379.
- Feltovich PJ, Spiro RJ, Coulson RL, Anderson DK (1989). Multiple analogies for complex concepts: antidotes for analogy-induced misconception in advanced knowledge acquisition. In: *Similarity and Analogical Reasoning*, ed. S Vosniadou and A Ortony, Cambridge, UK: Cambridge University Press, 498–531.
- Flavell JH (1979). Metacognition and cognitive monitoring: a new area of cognitive-developmental inquiry. *Am Psychol* 34, 906–911.
- Flick LB (1995). Navigating a sea of ideas: teacher and students negotiate a course toward mutual relevance. *J Res Sci Teach* 32, 1065–1082.
- Glaser R (1994). *Application and Theory: Learning Theory and the Design of Teaching Environments*, Pittsburgh, PA: Learning Research and Development Center, University of Pittsburgh.
- Grosslight L, Unger C, Jay E, Smith CL (1991). Understanding models and their use in science: conceptions of middle and high school students and experts. *J Res Sci Teach* 23, 799–822.
- Harp SF, Mayer RE (1998). How seductive details do their damage: a theory of cognitive interest in science learning. *J Educ Psychol* 90, 414–434.
- Hartley LM, Wilke BJ, Schramm JW, D'Avanzo C, Anderson CW (2011). College students' understanding of the carbon cycle: contrasting principle-based and informal reasoning. *BioScience* 61, 65–75.
- Heller P (2001). Lessons learned in the CIPS (Constructing Ideas in Physical Science) curriculum project. Paper presented at the AAAS Project 2061 Science Textbook Conference, held 27 February to 2 March 2001, in Washington, DC. [www.project2061.org/events/meetings/textbook/literacy/heller.htm](http://www.project2061.org/events/meetings/textbook/literacy/heller.htm) (accessed 22 November 2016).
- Herrmann-Abell CF, Flanagan JC, Roseman JE (2013). Developing and evaluating an eighth grade curriculum unit that links foundational chemistry to biological growth: using student measures to evaluate the promise of the intervention. Paper presented at the Annual Meeting of the National Association of Research in Science Teaching, held 6–9 April 2013, in Rio Grande, PR.
- Hesse JJ, Anderson CW (1992). Students' conceptions of chemical change. *J Res Sci Teach* 29, 277–299.
- Karataş FÖ, Ünal S, Durland G, Bodner G (2013). What do we know about students' beliefs? Changes in students' conceptions of the particulate nature of matter from pre-instruction to college. In: *Concepts of Matter in Science Education*, ed. G Tsapalis and H Sevan, Dordrecht, Netherlands: Springer, 231–247.

- Kesidou S, Roseman JE (2002). How well do middle school science programs measure up? Findings from Project 2061's curriculum review. *J Res Sci Teach* 39, 522–549.
- Krajcik J, McNeill KL, Reiser B (2008). Learning-goals-driven design model: developing curriculum materials that align with national standards and incorporate project-based pedagogy. *Sci Educ* 92, 1–32.
- Krippendorff K (2004). Reliability in content analysis: some common misconceptions and recommendations. *Hum Comm Res* 30, 411–433.
- Krnel D, Watson R, Glazar SA (1998). Survey of research related to the development of the concept of "matter." *Int J Sci Educ* 20, 257–289.
- Krüger D, Fleige J, Riemer T (2006). How to foster an understanding of growth and cell division. *J Biol Educ* 40, 135–140.
- Kruse R, Howes EV, Carlson J, Roth K, Bourdélát-Parks B, Roseman JE, Herrmann-Abell CF, Flanagan JC (2013). Developing and evaluating an eighth grade curriculum unit that links foundational chemistry to biological growth: changing the research-based curriculum. Paper presented at the Annual Meeting of the National Association of Research in Science Teaching, held 6–9 April 2013, in Rio Grande, PR.
- Kyle WC, Bonnstetter RJ, Gadsden T, Shymansky JA (1988). What research says about hands-on science. *Sci Child* 25(7), 39–40.
- Leach J, Driver R, Scott P, Wood-Robinson C (1992). Progression in Understanding of Ecological Concepts by Pupils Aged 5 to 16, Leeds, UK: Centre for Studies in Science and Mathematics Education, University of Leeds.
- Lee O, Eichinger DC, Anderson CW, Berkheimer GD, Blakeslee TD (1993). Changing middle school students' conceptions of matter and molecules. *J Res Sci Teach* 30, 249–270.
- Lehrer R, Chazan D (1998). Designing Learning Environments for Developing Understanding of Geometry and Space, Mahwah, NJ: Erlbaum.
- Linacre JM (2013). Winsteps® Rasch Measurement Computer Program, Beaverton, OR: Winsteps.com.
- Linn M, Burbules N (1993). Construction of knowledge and group learning. In: *The Practice of Constructivism in Science Education*, ed. K Tobin, Hillsdale, NJ: Erlbaum, 91–119.
- Liu X, Boone WJ (2006). Introduction to Rasch measurement in science education. In: *Applications of Rasch Measurement in Science Education*, ed. X Liu and WJ Boone, Maple Grove, MN: JAM Press, 1–22.
- Marmaroti P, Galanopoulou D (2006). Pupils' understanding of photosynthesis: a questionnaire for the simultaneous assessment of all aspects. *Int J Sci Educ* 28, 383–403.
- Mas CJ, Perez JH, Harris H (1987). Parallels between adolescents' conception of gases and the history of chemistry. *J Chem Educ* 64, 616–618.
- Masters GN (1982). A Rasch model for partial credit scoring. *Psychometrika* 47, 149–174.
- McDermott L (1991). Millican Lecture 1990: what we teach and what is learned—closing the gap. *Am J Phys* 59, 301–315.
- McNeill KL, Krajcik J (2012). Supporting Grade 5–8 Students in Constructing Explanations in Science: The Claim, Evidence and Reasoning Framework for Talk and Writing, New York: Pearson/Allyn & Bacon.
- Minstrell J (1984). Teaching for the understanding of ideas: forces on moving objects. In: *Observing Science Classrooms: Perspectives from Research and Practice*, 1984 Yearbook of the Association for the Education of Teachers in Science, ed. CW Anderson, Columbus, OH: ERIC Center for Science, Mathematics and Environmental Education, 55–73.
- Mitchell I, Gunstone R (1984). Some student conceptions brought to the study of stoichiometry. *Res Sci Educ* 14, 78–88.
- Mohan L, Chen J, Anderson CW (2009). Developing a multi-year learning progression for carbon cycling in socio-ecological systems. *J Res Sci Teach* 46, 675–698.
- Morse MP (2001). A Review of Biological Instructional Materials for Secondary Schools, Washington, DC: American Institute of Biological Sciences.
- National Research Council (NRC) (2000). *How People Learn: Brain, Mind, Experience, and School*, Washington, DC: National Academies Press.
- NRC (2009). *A New Biology for the 21st Century*, Washington, DC: National Academies Press.
- NRC (2012). *A Framework for K–12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*, Washington, DC: National Academies Press.
- Needham R (1987). *Teaching Strategies for Developing Understanding in Science*, Leeds, UK: Centre for Studies in Science and Mathematics Education, Children's Learning in Science Project.
- Next Generation Science Standards Lead States (2013). *Next Generation Science Standards: For States, By States*, Washington, DC: National Academies Press.
- Osborne R, Freyberg P (1985). *Learning in Science: The Implications of Children's Science*, Auckland, New Zealand: Heinemann.
- 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), Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education, <http://ies.ed.gov/ncee/wwc/Docs/PracticeGuide/20072004.pdf> (accessed 22 November 2016).
- Perkins DN, Salomon G (1988). Teaching for transfer. *Educ Leadership* 46, 22–32.
- Piaget J (1954). *The Construction of Reality in the Child*, New York: Basic.
- Posner GJ, Strike KA, Hewson PW, Gertzog W (1982). Accommodation of a scientific conception: toward a theory of conceptual change. *Sci Educ* 66, 211–227.
- Raudenbush SW, Bryk AS, Cheong YF, Congdon RT, du Toit M (2011). *HLM 7: Hierarchical Linear and Nonlinear Modeling*, Chicago, IL: Scientific Software International.
- Riemer T, Gropengießer H (2008). On the roots of difficulties in learning about cell division: process-based analysis of students' conceptual development in teaching experiments. *Int J Sci Educ* 30, 923–939.
- Roseman JE, Fortus D, Krajcik J, Reiser B (2015). Curriculum materials for next generation science standards: what the science education research community can do. Paper presented at the Annual Meeting of the National Association of Research in Science Teaching, held 11–14 April 2015, in Chicago, IL.
- Roseman JE, Herrmann-Abell CF, Flanagan J, Kruse R (2016). Integrating NGSS core ideas and practices: supporting and studying teachers' implementation. Paper presented at the Annual Meeting of the National Association of Research in Science Teaching, held 14–17 April 2016, in Washington, DC.
- Roseman JE, Herrmann-Abell CF, Flanagan J, Kruse R, Howes E, Carlson J, Roth K, Bourdélát-Parks B (2013). Developing and evaluating an eighth grade curriculum unit that links foundational chemistry to biological growth: selecting core ideas and practices—an iterative process. Paper presented at the Annual Meeting of the National Association of Research in Science Teaching, held 6–9 April 2013, in Rio Grande, PR.
- Roseman JE, 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. *J Res Sci Teach* 47, 47–70.
- Roth K (1991). Reading science texts for conceptual change. In: *Science Learning Processes and Applications*, ed. CM Santa and DE Alverman, Newark, DE: International Reading Association, 48–63.
- Roth KJ, Lemmens M, Garnier HE, Chen C, Wickler NI, Roseman JE, Barton AC, Zembal-Saul C (2009). Coherence and science content storylines in science teaching: Evidence of neglect? Evidence of effect? Symposium presented at the Annual Meeting of the National Association of Research in Science Teaching, held 17–21 April 2009, in Garden Grove, CA.
- Roth WM (1996). Teacher questioning in an open-inquiry learning environment: interactions of context, content, and student responses. *J Res Sci Teach* 33, 709–736.
- Sadler PM (1998). Psychometric models of student conceptions in science: reconciling qualitative studies and distractor-driven assessment instruments. *J Res Sci Teach* 35, 265–296.
- Sanchez CA, Wiley J (2006). An examination of the seductive details effect in terms of working memory capacity. *Mem Cognit* 34, 344–355.
- Sandoval WA (2003). Conceptual and epistemic aspects of students' scientific explanations. *J Learn Sci* 12, 5–51.
- Sandoval WA, Reiser BJ (2004). Explanation-driven inquiry: integrating conceptual and epistemic scaffolds for scientific inquiry. *Sci Educ* 88, 345–372.



- Smith EL, Anderson CW (1986). Alternative student conceptions of matter cycling in ecosystems. Paper presented at the Annual Meeting of the National Association of Research in Science Teaching, held April 28–31, 1986, in San Francisco, CA.
- StataCorp (2009). Stata: Release 11, Statistical Software, College Station, TX: StataCorp.
- Stern L, Roseman JE (2004). Can middle school science textbooks help students learn important ideas? Findings from Project 2061's curriculum evaluation study: life science. *J Res Sci Teach* 41, 538–568.
- Strike K, Posner GS (1985). A conceptual change view of learning and understanding. In: *Cognitive Structure and Conceptual Change*, ed. L West and AL Pines, Orlando, FL: Academic, 211–231.
- Thagard P (1992). *Conceptual Revolutions*, Princeton, NJ: Princeton University Press.
- Vaz AN, Carola MH, Neto AJ (1997). Some contributions for a pedagogical treatment of alternative conceptions in biology: an example from plant nutrition. Paper presented at the Annual Meeting of the National Association of Research in Science Teaching, held March 21–24, 1997, in Oak Brook, IL.
- Vygotsky L (1978). *Mind in Society: The Development of Higher Psychological Processes*, Cambridge, MA: Harvard University Press.
- White BY, Fredrickson JR (1998). Inquiry, modeling, and metacognition: making science accessible to all students. *Cogn Sci* 16, 90–91.
- Wiggins G, McTighe J (2005). *Understanding by Design*, 2nd ed., Alexandria, VA: Association for Supervision and Curriculum Development.
- Wise KC, Okey JR (1983). A meta-analysis of the effects of various science teaching strategies on achievement. *J Res Sci Teach* 20, 419–435.
- Wright BD (2003). Rack and stack: time 1 vs. time 2. *Rasch Meas T* 17, 905–906.
- Wright BD, Stone MH (2004). *Making Measures*, Chicago, IL: Phareon.