Evaluating the Promise of an Intervention that Helps Students Understand Chemical Reactions in Living Systems

Cari F. Herrmann-Abell, Jean C. Flanagan, and Jo Ellen Roseman
AAAS Project 2061

Paper presented at the
2014 NARST Annual International Conference
Pittsburgh, PA
March 30 – April 2, 2014

Abstract

Modern biology has become increasingly molecular in nature, requiring students to understand basic chemical concepts. Yet studies show 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 collaborated in the development of a curriculum intervention that applies chemistry ideas to both living and non-living contexts. Seven eighth grade teachers participated in a field test of the unit during the spring of 2013. Three of the teachers had used an earlier version of the unit in the spring of 2012. The other four teachers were randomly assigned to either implement the unit or continue teaching the targeted ideas using the district curriculum. A pre- and post-test were administered and the data were analyzed using Rasch modeling and hierarchical linear modeling. The results showed that, when controlling for pre-test score, gender, and ethnicity, students who used the curriculum intervention performed better on the post-test compared to the students using the district curriculum covering the same content. Additionally, students who participated in the intervention held fewer misconceptions and were better able to use science ideas to explain real-world phenomena after experiencing the unit. These results demonstrate the unit's promise in improving students' understanding of the targeted ideas.

Introduction

According to the National Research Council, understanding modern biology depends more than ever on an understanding of chemistry:

> Much of modern biology has become increasingly chemical in character. This has always been true of biochemistry and medicinal chemistry, but molecular biology, genetics, cell biology, proteomics, physiology, microbiology, neurobiology, agriculture, and many other divisions of biology are now using chemistry as a major part of their language and their research. The trend will continue, as more and more biological phenomena are explained in fundamental chemical terms. (National Research Council [NRC], 2003, p. 136).

Past assessment research indicates that students do not have a coherent understanding of the molecular basis of biology, and misconceptions related to this topic are prevalent at both the middle and high school levels (AAAS Project 2061, n.d.). For example, students typically have difficulty predicting that mass will be conserved, especially for systems in which there appears to be an increase or decrease of visible "stuff" (Mitchell & Gunstone, 1984). A national assessment study conducted by AAAS Project 2061 found that 56% of middle school students and 50% of high school students thought that when mold grows in a closed system, the mass of the system must increase (DeBoer, Herrmann-Abell, Wertheim, & Roseman, 2009). Students also commonly assume that the mass of a plant comes from minerals in the soil (Vaz, Carola, & Neto, 1997), mostly because they think that gases have negligible mass (Mas, Perez, & Harris, 1987) and, therefore, cannot contribute significantly to the mass of the plant. This study found that 54% of middle school students and 58% of high school students held this misconception. Another common misconception is that food is either used for energy or eliminated as waste, ignoring the idea that some of the food is used to build or repair body parts (Smith & Anderson, 1986). This misconception was held by 60% of the middle school students and 69% of the high school students in the national sample.

A team of science education researchers and curriculum developers developed a six-week curriculum unit, called *Toward High School Biology* (THSB), that connects core chemistry and biochemistry ideas in order to help students build a strong conceptual foundation for their study of biology in high school and beyond. The curriculum intervention consists of instructional materials for both students and teachers and a set of hybrid (face-to-face and online) professional development materials. Guiding the development of the unit was a theory of change positing that students' science understanding develops from (a) having a wide range of experiences with the natural world that are explainable by a coherent set of fundamental ideas, (b) having an opportunity to make sense of what they experience in terms of those ideas, and (c) having those experiences sequenced into a coherent content storyline that is made explicit to students.

The THSB unit differs from existing materials in several ways. First, the unit incorporates the study of chemical reactions in both physical and life science contexts to ensure students appreciate the broad applicability of science ideas about atom rearrangement and conservation. Second, the unit addresses common and persistent misconceptions students have about chemical reactions at both substance and atomic/molecular levels by engaging students in activities that are designed to contradict these incorrect ideas. Third, the unit engages students with relevant phenomena that illustrate the science ideas. Finally, the unit takes advantage of physical models and other representations to guide student explanations of phenomena in terms of underlying molecular events.

This paper reports on the results of a student pre-test and post-test administered during the field-testing of the curriculum intervention. The results indicate that the intervention shows promise in its ability to increase students' understanding of the targeted science ideas.

Methodology

Learning Goals. As recommended in the National Research Council's *A Framework for K-12 Science Education* (NRC, 2012) and the *Next Generation Science Standards* (NGSS) (NGSS Lead States, 2013), the THSB unit is designed to address core disciplinary ideas, science practices, and crosscutting concepts. 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.

Disciplinary core ideas and crosscutting concepts. The science ideas to which the THSB unit and assessments were aligned are shown in Table 1. These statements were adapted from grade band endpoints for grade 8 articulated in sections PS1.A, PS1.B, and LS1.C from the National Research Council's *A Framework for K-12 Science Education* (NRC, 2012). The crosscutting concept *Energy and Matter: Flows, Cycles, and Conservation* is addressed by the unit as well, specifically the concept that "matter is conserved because atoms are conserved in physical and chemical processes" (NGSS Lead States, 2013, Appendix G).

Table 1: *Targeted science ideas*

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 carbon dioxide 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)

Science practices. The THSB unit addresses five of the eight science practices recommended by the NRC's *Framework:* (1) developing and using models, (2) analyzing and interpreting data, (3) constructing explanations, (4) engaging in Argument from Evidence, and (5) obtaining, evaluating, and communicating information (NRC, 2012). The pre- and post-test assessment includes the use of models and data in some of the items, but students are not assessed on their ability to develop models or analyze data. Some items require students to select the correct explanation of a phenomenon, and the open-ended items target students' ability to construct explanations.

Curriculum Intervention. The curriculum intervention was developed using a three-year iterative development process that is well aligned with the theory of research-based curriculum design described by Clements (2007) and current learning research. In the first year, a "backward design" strategy was used to develop an initial draft of the student materials (Wiggins & McTighe, 2005). The draft student materials were pilot tested by researchers in a small number of schools, and data from the pilot test was used to revise the student materials and 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 curriculum 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.

The unit tested in Year 3 consists of 20 lessons organized into four chapters. 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 1). 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 mas stays the same (Idea 3 in Table 1). Chapter 3 applies the concepts of atom rearrangement and conservation to animal growth and repair (Idea 4 in Table 1), and Chapter 4 applies these concepts to plant growth and repair (Idea 5 in Table 1).

The lessons within each chapter engage students with (1) a range of phenomena related to the learning goals and (2) a variety of models of underlying molecular events including using LEGO® bricks, ball-and-stick, and space-filling models to represent atom rearrangement and conservation before using chemical equations. Using the same models across chemical reactions in living and non-living systems highlights the common principles underlying both. Additionally, the unit provides scaffolds that familiarize students with using evidence and reasoning from science ideas and models to explain phenomena and justify their explanations.

Study Design. The field test was conducted in the spring of 2013 in two districts in the Mid-Atlantic U.S. This paper reports on the results from a study of one of the districts. Seven teachers from seven schools participated in the study. Three of the teachers had participated in the Year 2 pilot test of the curriculum intervention 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 using the unit for the first time and were matched based on school characteristics such as 8th grade state test scores in math and science and student demographic variables such as ethnicity to create pairs with similar characteristics. All of the classes in one school in each pair were randomly assigned to use the intervention (from this point forward referred to as the "novice group"), and all of the classes in the other school were assigned to the control group. Treatment assignment within each pair of schools was done randomly by Abt Associates. In both the experienced and novice groups, the THSB unit replaced the students' usual curriculum materials, and the unit's lessons were taught by the classroom teacher after the teacher participated in three days of face-to-face professional development. The students in the control group used the "business as usual" curriculum, which targets the same science ideas.

Participants. A total of 680 students participated in the study, but the data reported here is from the 674 students who completed both the pre-test and the post-test and responded to at least 25% of the items on both tests. Student demographic data indicated that 54% of the students were male and 46% were female; about 44% of the students were white, 15% were African American, 27% were Asian, 8% were Hispanic, and 7% were two or more ethnicities. A breakdown of the demographic data by group is presented in Table 2.

Table 2: *Summary of class and student level variables*

	Control	Novice	Experienced
Number of Classes	9	10	14
Gifted & Talented Classes	67%	50%	50%
Number of Students	196	194	284
Average Pre-test score	-0.14	-0.44	-0.61
Gender			
Male	56%	55%	53%
Female	44%	45%	47%
Ethnicity			
White	45%	41%	44%
Asian	27%	29%	26%
Black	14%	11%	19%
Hispanic	9%	10%	6%
Two or more ethnicities	6%	9%	5%

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, a test was administered before and after instruction. The test contained 36 items, which were a mix of distractor-driven multiple-choice items and two-tiered items. The development of the multiple-choice items used 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, Herrmann-Abell, & Gogos, 2007; DeBoer, Herrmann-Abell, et al., 2008; DeBoer, Lee, & Husic, 2008). Each item was aligned to one or two of the targeted science ideas listed in Table 1, and item distractors were designed to probe for relevant student misconceptions (Sadler, 1998). 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 choice they selected was 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. The ideal response, which included both substance level and atomic/molecular level ideas, was broken down into a series of essential elements. There were four or five essential elements for each item. Table 3 presents the essential elements for the plant growth two-tiered item. Students' scores were based on the number of essential elements their responses included. One point was deducted for each misconception included in the response. Each response was rated by two researchers, and any disagreements were resolved by consulting a third researcher. The students received one score for these two-tiered items that summed their scores on the multiple-choice item (zero or one) and the written explanation (from zero to four or five depending on the number of essential elements for the item). None of the students who selected an incorrect answer choice received points for their written explanations because their explanations included misconceptions which reduced any points they did earn to zero.

Table 3: *The essential elements of an ideal response to the plant growth two-tiered item*

1. *Photosynthesis*: Carbon dioxide molecules react [with water molecules] to form glucose/sugar/food molecules [and oxygen molecules].
2. *Polymer formation*: Glucose/sugar/food molecules react to form carbohydrates/polymers/cellulose/starch molecules.
3. *Body structures made of polymers:* Carbohydrates/polymers/cellulose/starch molecules make up/used to make [most] of the wood.
4. *Therefore, atom rearrangement & incorporation*: Atoms from carbon dioxide molecules are rearranged into other molecules that become part of the wood.

Description of Rasch Modeling. Student-level scale score were created using Rasch modeling (Bond & Fox, 2007; Liu & Boone, 2006; Boone, Staver, & Yale, 2014). The "Partial Credit" model was used because the test included both dichotomous and polytomous items (Masters, 1982). 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. In Rasch modeling, 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 $+\infty$. The average item difficulty was set at zero.

Measuring change using Rasch modeling. In this paper, we apply the stacking method to the pre-test and post-test data (Wright, 2003). Stacking data allowed us to create two scale scores per person: a pre-test score and a post-test 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 pre-test and the second row contains their responses during the post-test.

Hierarchical Linear Modeling. Once the pre-test and post-test scale scores were created, the post-test scale scores were modeled as outcome measures in two-level hierarchical linear models (HLM) 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 post-test scores varied between classes of the same teacher. Student-level variables included pre-test scores, gender, and ethnicity. Class-level variables included whether or not the class was part of the novice group of a match 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.

A fully unconditional model containing only the post-test outcome variable and no independent variables except an intercept was estimated first. This was followed by a conditional model in which pre-test score, gender, and ethnicity were included as controls and modeled as fixed-effects. HLM 7 software was used in this study (Raudenbush, Bryk, Cheong, Congdon, & du Toit, 2011). The method of estimation was restricted maximum likelihood. Effect sizes were calculated by dividing the coefficient by the square root of the total standard deviation.

Results and Discussion

Rasch Fit. The stacked data had a good fit to the Rasch model as shown in Table 4. The separation indices and corresponding reliabilities were 17.74 and 1.00 for the items and 2.88 and .89 for the students. Both of the separation indices are considered acceptable—i.e., greater than 2, according to Wright and Stone (2004). Additionally, the standard errors for the items and students were small (see Table 4). The infit and outfit mean-square values for the majority of the items and students were within the acceptable range of 0.7 to 1.3 for multiple-choice tests (Bond & Fox, 2007).

Table 4: *Rasch fit statistics for the stacked data*

	Item			Person		
	Min	Max	Median	Min	Max	Median
Standard error	0.03	0.10	0.06	0.32	1.25	0.35
Infit mean-square	0.82	1.25	1.00	0.08	3.60	0.98
Outfit mean-square	0.68	1.39	1.00	0.02	8.92	0.92
Point-measure correlation coefficients	0.29	0.70	0.46	-0.27	0.76	0.46
Separation index (reliability)	17.74 (1.00)			2.88 (0.89)		

Fully Unconditional HLM. A fully unconditional hierarchical linear model with no independent variables at either level was run in order to calculate the intraclass correlation coefficient. The results of the model are shown in Table 5. Over half (55%) of the variance in post-test score could be the function of class characteristics. The proportion of the variance in post-test scores that exists at the individual level is 45%. A chi-square test indicated that post-test scores varied significantly between classes ($\chi^2 = 746.391$, $p < .001$).

Table 5: *Fully unconditional HLM*

Within-classroom variance (σ^2)	0.75
Between-classroom variance (τ)	0.90
Between-classroom standard deviation	0.95
Reliability (λ)	0.96
Intraclass correlation (ρ)	0.55

Conditional HLM. The mixed-model for the conditional hierarchical linear model is

$$POSTTEST_{ij} = \gamma_{00} + \gamma_{01}*NOVICE_j + \gamma_{02}*EXPERIENCED_j + \gamma_{03}*GT_j + \gamma_{10}*PRETEST_{ij} + \gamma_{20}*FEMALE_{ij} + \gamma_{30}*BLACK_{ij} + \gamma_{40}*HISPANIC_{ij} + \gamma_{50}*ASIAN_{ij} + \gamma_{60}*2ORMORE_{ij} + u_{0j} + r_{ij}$$

where $POSTTEST_{ij}$ and $PRETEST_{ij}$ are the post- and pre-test 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 experience implementer of the THSB unit. The control group, which was using the district curriculum, was used as a reference group. FEMALE is a dummy variable indicating the gender of student *i* in class *j* (Female = 1 and Male = 0). Four dummy variables were created for ethnicity (BLACK, HISPANIC, ASIAN, and 2ORMORE) and white was used as a reference group. 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.

According to the results of the conditional HLM shown in Table 6, the average post-test score for a non-GT control group class is -0.37 logits (controlling for pre-test scores, gender, and ethnicities). On average, non-GT classes in the novice group score 0.91 logits higher, non-GT classes in the experienced group score 1.49 logits higher, and GT classes in the control group score 0.50 logits higher on the post-test than the non-GT control classes. Compared to the district curriculum, the effect size for the THSB unit being implemented for the first time (novice group) is 0.88 and the effect size for the THSB unit being implemented by teachers with prior experience with THSB (experienced group) is 1.88. Both of these effect sizes are considered to be large (i.e. greater than 0.80) (Cohen, 1977).

The larger effect size for the experienced group may be due to teachers' increased familiarity and comfort with the THSB unit and the fact that more of their classes reached the end of the unit. During conversations with the teachers in the experienced group, most of them indicated that during the Year 3 implementation of the THSB unit, they better understood the storyline of the unit and the science content and were more comfortable with the molecular models used in the unit. Regarding completion of the unit, ten out of the fourteen classes in the experienced group completed all of the lessons, compared to none of the classes in the novice group.

Table 6: *Results from the conditional HLM*

Fixed Effects	Coefficient	Standard error	*t*-ratio	Approx. *d.f.*	*p*-value
Class level variables					
Intercept, γ_{00}	-0.37	0.17	-2.18	29	0.04
Novice, γ_{01}	0.91	0.19	4.77	29	<0.001
Experienced, γ_{02}	1.49	0.18	8.38	29	<0.001
GT, γ_{03}	0.50	0.15	3.33	29	0.002
Individual level variables					
Pre-test, γ_{10}	0.82	0.04	19.36	635	<0.001
Female, γ_{20}	0.13	0.05	2.38	635	0.02
Black, γ_{30}	-0.36	0.09	-4.07	635	<0.001
Hispanic, γ_{40}	-0.29	0.11	-2.57	635	0.01
Asian, γ_{50}	0.03	0.07	0.38	635	0.71
2 or more, γ_{60}	-0.16	0.11	-1.41	635	0.16
Random Effects	**Standard Deviation**	**Variance**	***d.f.***	**χ^2**	***p*-value**
Intercept, u_0	0.38	0.14	29	216.02	<0.001
level-1, *r*	0.68	0.46			

Change in Student Understanding of Learning Goals. Table 7 presents the percentage of students in each group who selected the correct answers on the multiple-choice items by learning goal. For the novice and experienced groups, the increase in percent correct from pre- to post-test was statistically significant at the .001 level for all learning goals. For the control group, only the increases for the chemical reactions, conservation, and animal growth learning goals were statistically significant at the .01 level.

Table 7: *Percentage of students selecting the correct answer by learning goal*

Learning Goal	# of items	Control		Novice		Experienced	
		pre-test	post-test	pre-test	post-test	pre-test	post-test
Substances	4	75%	77%	68%	82%	67%	80%
Chemical reactions	6	57%	63%	51%	74%	49%	74%
Conservation	6	63%	75%	50%	76%	42%	73%
Chemical reactions & conservation	3	53%	56%	46%	68%	42%	66%
Animal growth	6	42%	52%	44%	80%	38%	82%
Plant growth	8	32%	36%	33%	44%	28%	66%

Substances. An analysis of the results of the Year 2 version of THSB found that the unit did not significantly improve students' understanding of the substances learning goal (Herrmann-Abell, Roseman, & Flanagan, 2013). As a result, the initial lessons of the unit were modified in order to attempt to better build students' understanding of substances and characteristic properties. Additionally, new activities in the later lessons about animal plant growth reinforced this idea by having students compare the properties and molecular structures of different proteins and carbohydrates. Table 7 shows that in Year 3 both the experienced and novice groups saw significant improvement on items aligned to this learning goal.

Chemical reactions. On the post-test, the novice and experienced groups performed significantly better on the chemical reaction items than the control group (χ^2= 44.39, p < .001).

Conservation. The control group's performance on the conservation items on the post-test did not significantly differ from the novice and experienced groups performance on these items on the post-test (χ^2= 3.94, n.s.). However, the control group performed significantly higher on these items on the pre-test than the other groups, suggesting they had a better understanding of conservation at the start of the study.

Animal growth. On the post-test, the novice and experienced groups performed significantly better on the animal growth items than the control group (χ^2= 357.46, p < .001).

Plant growth. On the plant growth items, the experienced group outperformed the novice group on the post-test and the novice group outperformed the control group (χ^2= 368.67, p < .001). The difference in performance on the plant growth items between the students of experienced and novice users may have been caused by differences in the number of lessons teachers completed: whereas most of the experienced users completed all of the lessons targeting plant growth ideas, none of the novice users did.

Distractor Analysis. An analysis of the students' selection of distractors was performed to gain insight into the effects the THSB unit had on students' ideas and misconceptions.

Transmutation of atoms. 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 during a reaction (Andersson, 1986). Five items on the pre- and post-test included distractors that probed this misconception. Overall, these distractors were selected on the pre-test by 31% of the time by students in control classrooms and 33% of the time by students who used the THSB unit (in classrooms of both novice and experienced users) (χ^2= 0.99, n.s.). On the post-test, the percentages decreased to 23% for the control group students and 14% for students who used the unit. The post-test percentage for THSB users is significantly lower than the percentage for the control group (χ^2= 34.39, p < .001). This indicates that the THSB was more successful in reducing the prevalence of this misconception than the district curriculum.

The THSB unit targets this misconception by having students model a variety of chemical reactions in both non-living and living systems. Students build physical models of the reactant molecules using either LEGO bricks or ball-and-stick models and then rearrange the "atoms" to form models of the product molecules. They observe that the product molecules are always made from the same number of each type of atoms as the reactant molecules and that each LEGO brick or ball maintains its identity during each reaction.

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 organism takes in additional matter that becomes part of their bodies (Krüger, Fleige, & Riemeier, 2006). Six items on the pre- and post-test included distractors aligned to this misconception. On the pre-test, these distractors were selected 34% of the time by students in the control group and 29% of the time by students who used the THSB unit (in classrooms of both novice and experienced users) (χ^2= 10.52, p < .01). On the post-test, the percentage decreased to 26% for the control group and 6% for the novice and experienced groups (χ^2= 305.19, p < .001).

The THSB unit targets this misconception by providing experiences with data and models that contradict it. For example, students examined evidence from ^{14}C labeling experiments showing that some of the atoms from a fish's food were incorporated into the fish's body structures. Using ball-and-stick models, the students then followed a "labeled" atom through protein digestion and synthesis. Similar activities were included for photosynthesis and cellulose production in plants.

Explaining Real-World Phenomena Using Science Ideas. A major goal of the THSB unit is for students to be able to explain observable events (e.g., the production of new substances) in terms of underlying atomic/molecular events, specifically atom rearrangement and conservation during chemical reactions in non-living systems and in the bodies of living organisms as they grow. Three of the two-tiered items on the pre- and post-test presented the students with real-world scenarios in which new materials appear: (1) mold grows on a slice of bread in a sealed bag, (2) a tree grows wood, and (3) a sea star grows back two arms. For the moldy bread item, the students were asked to determine how the mass of the sealed bag containing the bread would change after the mold grew, and then they were asked to explain their answer. For the tree and sea star items, the students were asked where the mass to build the body parts came from and then asked to explain their answer. Table 8 provides the percentage of students in each group who included atomic/molecular ideas in their explanations.

Table 8: *Percentage of students using atoms and molecules in their written explanations*

		Moldy Bread	Tree Growth	Sea Star Growth
Control	Pre-test	1%	0%	40%
	Post-test	1%	1%	30%
Novice	Pre-test	0%	1%	39%
	Post-test	36%	13%	53%
Experienced	Pre-test	0%	1%	29%
	Post-test	37%	27%	45%

Moldy bread item. An analysis of the students' pre-test explanations for the moldy bread item showed that students provided only substance-level explanations (with the exception of two students in the control group) (see Table 8). Similar results were seen on the post-test for the control group with only one student using the words "atoms" and "molecules" in their explanation. The results for the students in the novice and experienced groups, however, were different on the post-test. As seen in Table 8, the percentage of these students providing atomic/molecular explanations increased from pre- to post-test to over one third of the students.

Because the moldy bread item did not include any prompts for atomic/molecular ideas, students might not have felt the need to go beyond a substance-level explanation (e.g., the bag was sealed so nothing can get in or out). However, at least some students in the novice and experienced groups offered explanations on the post-test that included what happened to the atoms in the sealed bag without prompting. This suggests that the THSB unit was more successful than the district curriculum in supporting students in explaining conservation phenomena in terms of underlying atomic/molecular events.

Sample pre- and post-test explanations from two students who used the THSB unit are shown in Table 9. The students who wrote these explanations selected the correct answer choice on both the pre- and post-tests indicating that they knew 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 pre-test and an atomic-level on the post-test. The examples also indicate that the students who experienced the THSB unit gained a better understanding of the concept of conservation as articulated in the NRC's *Framework* and NGSS, which explicitly includes atom conservation.

Table 9: *Sample explanations for the moldy bread item from students in the experienced group*

Pre-test Explanation	Post-test 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."

Tree growth item. As with the moldy bread item, students failed to use atomic/molecular ideas in their pre-test explanations (see Table 8). On the post-test, only one control group student included atomic/molecular ideas, whereas 13% of the novice group and 27% of the experienced group did. These results mirror the results found for the multiple choice plant growth items discussed earlier. The difference in performance between the novice and experienced group may be attributed to the fact that the novice group did not complete all of the plant growth lessons and, therefore, did not receive guidance in how to explain plant growth in terms of the underlying atomic/molecular events.

Sample pre- and post-test explanations from two students who used the THSB unit are shown in Table 10. The first student was in a novice group class, and the second student was in an experienced group class. Both students selected the correct answer on the pre- and post-tests, indicating that they knew the mass of the tree came from carbon dioxide in the air. In each example, the student wrote a substance-level explanation on the pre-test and an atomic-level explanation on the post-test. The explanations provide evidence that the students' understanding of plant growth improved from pre- to post-test. Both students' pre-test explanations rely heavily on photosynthesis which is only part of the plant growth idea. On the post-test, the first student refers explicitly to the incorporation of the atoms from carbon dioxide into the tree but does not detail the process by which this occurs. The second student references data provided during the unit and works backwards from trees being made up of carbohydrates, to carbohydrates being made up of glucose monomers, to glucose monomers being made up of atoms from carbon dioxide.

Table 10: *Sample explanations for the tree growth item from two students in the novice and experienced groups*

Pre-test Explanation	Post-test Explanation
"Trees use carbon dioxide for photosynthesis, so it seems reasonable that most of their mass is made of it."	"Unlike humans who breathe in oxygen, plants like trees need carbon dioxide; they take in carbon dioxide from the air. These additional carbon dioxide atoms are incorporated into their existing mass so the tree's measured mass increases."
"Trees need carbon dioxide in order to photosynthesize. In this process, trees take the carbon, water and light to create a chemical reaction to grow."	"In the tables that we saw earlier showed that trees are made of 96% carbohydrates. Later, I learned that the carbohydrates were made of glucose monomers. During atom modeling, carbon dioxide and water molecules reacted to form glucose monomers. In another experiment, radio-labeled atoms showed that most of the atoms in glucose monomers came from carbon dioxide."

Sea star growth item. Roughly one-third of the students in each group included ideas about atoms and/or molecules in their pre-test response to the sea star item. This differs from the other two items. This difference could be due partially to the fact that all of the answer choices for the sea star item included the word "atoms," which cued the students to use this word in their response. The other items did not use the word "atom" or "molecule."

As shown in Table 8, the percentage of students using atomic/molecular ideas increased from pre- to post-test for the novice and experienced groups [14% gain for the novice group (χ^2= 7.76, $p < .01$) and 16% gain for the experienced group (χ^2= 15.27, $p < .001$)]. The control group percentage decreased 10%, which is not significant at the 0.01 level (χ^2= 4.06, n.s.). The difference in the number of students using atomic/molecular ideas on the post-test among the three groups was statistically significant (χ^2= 21.86, $p < .001$). These results indicate that the THSB unit was more successful that the district curriculum in supporting students' ability to explain animal growth in terms of underlying atomic/molecular events.

From the written responses to this item on the pre-test, it was clear that students associated food with growth, but they did not understand the role atoms from food play in the growth process. Samples of students' explanations for the sea star growth item are provided in Table 11. The first student, who was a member of the novice group, simply says that the food the sea star eats "produces" enough atoms to grow body parts. The second student, who was a member of the experienced group, seems to think that food remains in the stomach until it is needed as energy for growth. The post-test explanations indicate that these students have learned that chemical reactions play a prominent role in growth.

Table 11: *Sample explanations for the sea star growth item from two students in the novice and experienced groups*

Pre-test Explanation	Post-test Explanation
"Answer A is correct because, as stated before, sea stars eat mussels and snails. These creatures that were consumed by the sea star would produce enough atoms to grow back the two arms which the sea star needed."	"Option A is correct because the sea star ate food that contained atoms and proteins, both contain mass. These nutrients and atoms were used in the body through chemical reactions to produce the two new arms grown by the sea star."
"There are more atoms in the sea stars body now that is has eaten, so it can use them to grow its arms back. The food is kind of just floating around in its stomach, so it can use that energy to grow new arms."	"The sea star's diet has protein in it. (from the mussels and snails) The sea star's body can take the protein polymers from these animals and use them in a chemical reaction to form a different substance that will be used to help the sea star grow and repair its legs. The additional mass comes from the atoms in the mussels and snails because they were not originally in the measured mass. (they were outside the open system)."

Even though there was an overall increase in the percentage of students providing atomic/molecular explanations from pre- to post-test on these three items, the increase is modest. Only 13% to 37% of the students who experienced the THSB unit used atomic/molecular ideas in their post-test responses when not prompted by the use of the word "atom" in the item. Furthermore, the average score (expressed as percentage of total possible points) for students in the novice and experienced groups who selected the correct answer choice on the post-test ranged from 34% to 50% across the three items indicating that the majority of the students did not include all of the essential elements that make up an ideal response.

A further analysis of the instructional support the unit provided identified some deficiencies. For example, the unit did not make explicit that the use of atoms and molecules was included as a criterion for a good explanation. The activities within the unit intermittently prompted students to use atoms and molecules, which could have lead students to conclude that they should only mention atoms and molecules when prompted. In response to this finding, the unit has been revised to better support students in constructing explanations of phenomena and the effects of these revisions are being examined this spring. Revisions include (a) making explicit to students the expectation that changes in substances should be explained in terms of underlying changes in molecules, (b) initially prompting students to do so and then fading prompts, (c) reminding teachers to monitor students' use of atom rearrangement and conservation in their explanations after prompts have been removed, and (d) giving teachers rubrics for scoring students' final explanations.

Conclusions

This paper reports on a subset of data from the Year 3 field test of a new curriculum 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. The unit aims to improve on currently available materials by engaging students with phenomena that occur in both non-living and living systems and scaffolding students' sense making. This scaffolding includes questions and modeling tasks that help students connect activities 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 experiences.

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 learning goals. Rasch modeling was used to create scale scores for both the pre- and post-test. These scale scores were then modeled as outcomes in a two-level hierarchical linear model to investigate effects of the intervention controlling for pre-test score, gender, and ethnicity. The results of the model showed a significantly positive correlation between using the THSB unit and post-test score. Large effect sizes were found for both the novice group and for the experienced group. 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 non-living systems and for the unit's feasibility, which improves in the hands of experienced teachers.

Acknowledgements

The authors would like to acknowledge the members of the BSCS project team: Janet Carlson, Brook Bourdelat-Parks, Elaine Howes, Rebecca Kruse, Kathy Roth, Aleigh Raffelson, Kerry Skaradznski, Rhiannon Baxter, Stacey Luce, and Chris Moraine. The authors would like to thank Mary Koppal of Project 2061 for her help in editing the curriculum unit, Martin Fernandez of Project 2061 for his help in scoring the written explanations, Bernard Koch of Project 2061 for his help in analyzing the written explanations, and Caitlin Klein for her help in scoring the written explanations.

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.

References

AAAS Project 2061 (n.d.) *AAAS Science assessment website*. Retrieved from http://assessment.aaas.org on March 26, 2014.

Andersson, B. (1986). Pupils' explanations of some aspects of chemical reactions. *Science Education*, *70*(5), 549-563.

Bond, T., & Fox, C.(2007). *Applying the Rasch model*. Mahwah, NJ: Lawrence Erlbaum Associates.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. NY: Academic Press.

Clements, D. (2007). Curriculum Research: Toward a Framework for "Research Based Curricula" *Journal of Research in Mathematics Education*, *38*(1) 35-70.

DeBoer, G. E., Herrmann-Abell, C. F., & Gogos, A. (2007, April). *Assessment linked to science learning goals: Probing student thinking during item development.* Paper presented at the National Association of Research in Science Teaching Annual Conference, New Orleans, LA.

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

DeBoer, G. E., Herrmann-Abell, C. F., Wertheim, J., & Roseman, J. E. (2009, April). *Assessment linked to middle school science learning goals: A report on field test results for four middle school science topics*. Paper presented at the National Association of Research in Science Teaching Annual Conference, Garden Grove, CA.

DeBoer, G. E., Lee, H. S., & Husic, F. (2008). Assessing integrated understanding of science. In Y. Kali, M. C. Linn, & J. E. Roseman (Eds.), *Coherent science education: Implications for curriculum, instruction, and policy* (pp. 153-182). New York, NY: Columbia University Teachers College Press.

Herrmann-Abell, C. F., Flanagan, J.C., and Roseman, J.E. (2013, April). *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 National Association of Research in Science Teaching Annual Conference, Rio Grande, PR.

Krüger, D., Fleige, J., and Riemeier, T. (2006). How to foster an understanding of growth and cell division, *Journal of Biological Education*, *40*(3), 135-140.

Linacre, J. M. (2013). Winsteps® Rasch measurement computer program. Beaverton, Oregon: Winsteps.com.

Mas, C. J., Perez, J. H., & Harris, H. (1987). Parallels between adolescents' conception of gases and the history of chemistry. *Journal of Chemical Education*, 64(7), 616-618.

Mitchel, I., & Gunstone, R. (1984). Some student conceptions brought to the study of stoichiometry. *Research in Science Education, 14*, 78-88.

Mohan, L., Chen, J., & Anderson, C. (2009). Developing a multi-year learning progression for carbon cycling in socio-ecological systems. *Journal of Research in Science Teaching*, *46*(6), 675-698.

National Center for Education Statistics. (2011). *The Nation's report card: Science 2009* (NCES 2011–451). Washington, DC: Institute of Education Sciences, U.S. Department of Education.

National Research Council. (2003). *BIO 2010: Transforming undergraduate education for future research biologists* [Electronic version]. Washington, DC: National Academy Press.

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

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

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., Congdon, R. T., & du Toit, M. (2011). HLM 7: Hierarchical linear and nonlinear modeling. Chicago, IL: Scientific Software International.

Sadler, P.M. (1998). Psychometric models of student conceptions in science: Reconciling qualitative studies and distractor-driven assessment instruments. *Journal of Research in Science Teaching*, *35*(3), 265-296.

Smith, E. L., & Anderson, C. W. (1986, April). *Alternative student conceptions of matter cycling in ecosystems*. Paper presented at the the National Association of Research in Science Teaching Annual Conference, San Francisco, CA.

Vaz, A. N., Carola, M. H., & Neto, A. J. (1997, April). *Some contributions for a pedagogical treatment of alternative conceptions in biology: An example from plant nutrition*. Paper presented at the National Association of Research in Science Teaching Annual Conference, Oak Brook, IL.

Wiggins, G., & McTighe, J. (2005). *Understanding by design* (Expanded 2nd ed.). Alexandria, VA: Association for Supervision and Curriculum Development (ASCD).

Wright, B. D. (2003) Rack and stack: Time 1 vs. time 2. *Rasch Measurement Transactions, 17*(1), 905-906.

Wright, B. D., & Stone, M. H. (2004). *Making Measures*. Chicago, IL: The Phaneron Press.