Results from a Pilot Study of a Curriculum Unit Designed to Help Middle School 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
2012 NARST Annual International Conference
Indianapolis, IN
March 25 – 28, 2012

Abstract

Students often have trouble understanding key biology ideas because they lack an understanding of foundational chemistry ideas. AAAS Project 2061 is collaborating with BSCS in the development a curriculum unit that connects core chemistry and biochemistry ideas in order to help eighth grade students build the conceptual foundation needed for high school biology. The unit is designed to engage students in (a) observing phenomena that are explicitly aligned to the targeted ideas and selected to address common learning difficulties and (b) using models to help them interpret the phenomena in light of the targeted ideas. The unit was pilot tested with students from an urban school and a suburban school in the mid-Atlantic region of the U.S. at the end of or just after eighth grade. Multiple choice pretests and posttests were used to measure the change in students' understanding. The pre/posttest data were analyzed using Rasch modeling and the racking and stacking methods. The stacking method showed that the students at both schools made statistically significant gains in their performance on the items testing understanding of the targeted chemistry and biochemistry ideas. The racking method showed that the difficulty of most of the items decreased as a result of the intervention, suggesting that the unit successfully targeted most of the chemistry and biochemistry ideas. A distractor analysis showed that the students appeared to hold fewer misconceptions after participating in the unit. These results were used to inform revisions to the curriculum unit.

Introduction

Evidence from large-scale student assessments makes it clear that U.S. students are not being well prepared in science. In the 2009 National Assessment of Educational Progress (NAEP) science assessment, only 21% of 12th-graders reached the proficient level, and 40% performed at or below basic (National Center for Education Statistics, 2011). Although U.S. students are not performing well in any of the sciences, we are particularly concerned about students' low achievement on topics that are essential for further study of biology (see e.g. Bell, 1985; Andersson, 1986; Mohan et. al., 2009). Today's middle and high school students must be better prepared if they are going to succeed in college level biology courses, which demand a solid understanding of chemistry. The National Research Council has called attention to the increased dependency of biology on chemistry, noting that this "trend will continue, as more and more biological phenomena are explained in fundamental chemical terms" (2003, p. 136).

The AAAS Project 2061 research team is collaborating with BSCS in the development of a five-week curriculum unit, *Toward High School Biology*, 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. Guiding the development of the unit is 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 ideas and (b) having an opportunity to make sense of what they experience in terms of those ideas.

The unit differs from existing materials in several ways. First, the unit promotes students' sense making through a coherent presentation of the science ideas. A coherent content storyline is developed by (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 (Roth et al. 2009). Throughout the design and revisions process, the unit is being evaluated according to curriculum design specifications that enable us to systematically assess the coherence of the unit (Roseman, Stern, & Koppal, 2010).

Second, the unit addresses the most common and persistent misconceptions students have about chemical and biochemical changes and their molecular-level explanations. According to Anderson, Sheldon, and DuBay (1990), "students' difficulties in understanding the 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). Fewer than 20% of a national sample of about 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, Herrmann Abell, Wertheim, & Roseman, 2009). The unit is being designed to give students a solid grounding in chemical reactions and conservation and then explicitly relate experiences with physical science phenomena to experiences with life science phenomena. Students are given an atomic-molecular model for biological growth and use data and models to make sense of the chemical reactions that result in growth.

Third, the unit engages students with relevant real-world phenomena and helps them to develop scientific explanations. Students participating in the unit experience a range of phenomena related to chemical reactions in both non-living and living systems. The phenomena included in the Year 1 draft of the unit ranged from iron in steel wool rusting and burning butane to radiolabeling experiments tracing carbon atoms from carbon dioxide in the air to glucose in plants and crayfish building chitin to make a new shell. For each phenomenon presented, the unit related the macroscale observations to the underlying molecular explanation.

Finally, the unit takes advantage of physical models and other powerful representations to guide students' sense making. In order to help students visualize the discrete nature of matter and rearrangement of atoms during chemical reactions, the unit provides experiences with molecular models constructed out of LEGO® bricks, ball-and stick model kits, and images of space-filling and ball-and-stick models.

We are currently in Year 2 of the three-year iterative-design study. This paper reports on the results of pretests and posttests administered during the piloting of the Year 1 draft of the unit. While the results are preliminary, they are promising and are helpful in informing revisions to the unit. The paper also discusses revisions to the unit that are currently underway.

Methodology

Curriculum unit. The Year 1 draft of the unit consisted of 10 chemistry lessons followed by 9 biochemistry lessons that build upon the chemistry lessons. The unit targeted ideas that are included in the 6-8 grade band in the science standards of nearly every state, including the states where we piloted the unit. The ideas are also found in the *2009 NAEP Science Framework*, *Benchmarks for Science Literacy* (National Assessment Governing Board, 2008), and the National Research Council's *Framework for K-12 Science Education* (NRC, 2012). The targeted chemistry and biochemistry ideas included:

- Chemical Reactions: Many substances react chemically in predictable ways with other substances to form new substances with different characteristic properties. When substances interact to form new substances, the atoms that make up the molecules of the original substances rearrange into new molecules.
- Conservation: Regardless of how substances within a closed system interact with one another, the total mass of the system remains the same. Whenever atoms interact with each other, regardless of how they are arranged or rearranged, the number of each kind of atom stays the same and, therefore, the total mass stays the same.
- Food: All organisms need food as a source of molecules that provide building materials and chemical energy.
- Photosynthesis: Plants make their own food in the form of sugar molecules from carbon dioxide molecules and water molecules. In the process of making sugar molecules, oxygen molecules are produced as well.
- Plant Growth: Plants use some of the sugar molecules to make a variety of larger carbon-containing molecules that become part of their body structures.
- Animal Growth: Animals use carbon-containing molecules from food to make a variety of other carbon-containing molecules that become part of their body structures.
- Obtaining Energy: Plants and animals obtain energy from a chemical reaction in which glucose and oxygen molecules react to produce carbon dioxide and water molecules.
- Storage: Plants and animals store molecules from food for later use.

The individual lessons within the unit involved (1) experiences with a range of phenomena to engage students in observing and raising questions and (2) a variety of molecular modeling activities including LEGO® bricks, ball-and-stick and space-filling models, chemical and structural formulas, and equations. Using a variety of models gave students different ways to represent and work with abstract ideas and to synthesize or connect seemingly disparate experiences and ideas.

Participants. Students from two schools in the mid-Atlantic region of the U.S. participated in the Year 1 pilot test. One school was located in a suburban area, and the other was located in an urban area. A total of 147 students participated in the lessons, but the data reported on here is from the 120 students who responded to both the pretest and the posttest. Of the 120 students, 91 were from the suburban school and 29 were from the urban school. The students at the suburban school were about half male and half female, and approximately 54% of the students were white, 22% were Asian, 13% were African American, and 7% were Hispanic. The students at the urban school were about half male and half female, and all were African American. The pilot test was conducted at the suburban school in May of 2011 during the students' eighth grade year and was conducted at the urban school in July of 2011 during a specially designed summer program for incoming ninth grade students. In both schools, the draft unit replaced the students' usual curriculum material, and the unit's lessons were taught by the classroom teacher with support from the research team.

Pretests and posttests. To determine whether students' understanding of the targeted learning goals changed as a result of the intervention, we administered a multiple choice test before and after the students participated in the unit. The tests were developed as part of an earlier effort to develop items aligned to national standards on the topics of Matter and Energy in Living Systems, and Substances, Chemical Reactions and Conservation of Matter. Item development 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 ideas, and item distractors were designed to probe for common student misconceptions. As part of this development project, the items were field-tested with a national sample of 573 ninth grade students in the spring of 2010. The data from this field test provided us with information on the current level of student understanding of the targeted ideas and was helpful in the development of the curriculum unit.

Due to the workflow of developing lessons for the new unit and the scheduling constraints of working with schools, the field test instruments from our earlier item development project were used as pretests and posttests with only minor modifications. As a result, while there was high overlap between the ideas targeted in the pre/posttests and the ideas targeted in the unit, one idea covered in the unit was not included on the test (Obtaining energy from food), other ideas were over-emphasized (Food), and still others were underemphasized (Plant growth).

There were three versions of the pre/posttests and each version, with the exception of one, covered all the targeted learning goals for which we had items. (For the pretest at the suburban school, the students responded to a version of the test that did not include any items targeting the ideas about chemical reactions.) Linking items were used so that the results across the different versions could be compared. The tests were administered online and students were given 30, 35, or 36 items depending on which version of the test they were assigned. Each student was assigned the same version for his/her pretest and posttest. A total of 52 items were included on the tests.

Description of Rasch modeling. The data from the pretests and posttests were analyzed using Rasch modeling. In the dichotomous Rasch model, the probability that a student will respond to an item correctly is determined by the difference in the student's ability and the difficulty of the item, according to the following equation:

$$\ln\left(\frac{P_{ni}}{1-P_{ni}}\right)=B_n-D_i$$

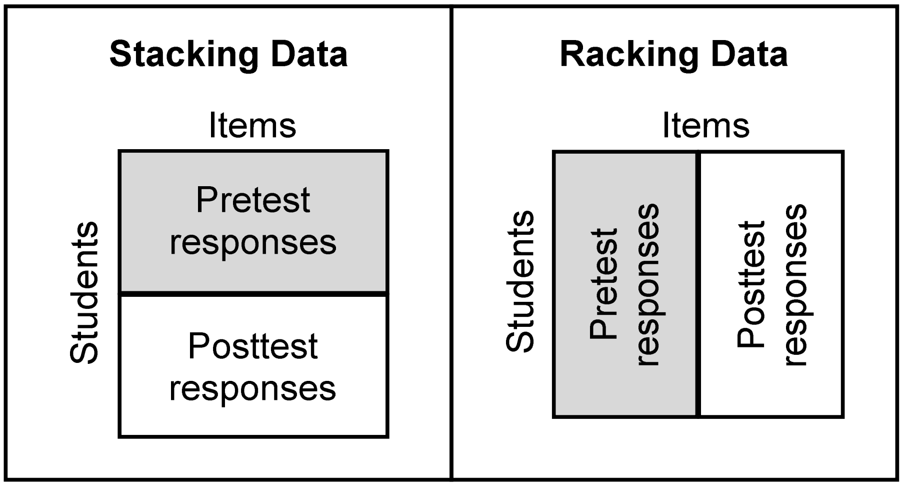
where P_{ni} is the probability that student n of ability B_n will respond correctly to item *i* with a difficulty of D_i (Bond & Fox, 2007; Liu & Boone, 2006). Student abilities and item difficulties, B_n and D_i, are measured in the unit of logarithm called log odds or logits, which can vary from $-\infty$ to $+\infty$. Student and item measures are expressed on the same interval scale and are mutually independent, which is not the case for percent correct statistics. (Note: Rasch modeling uses the term 'ability' to refer to the students' understanding of the science ideas being targeted. It should not be interpreted as an underlying, innate quality of the student, but more narrowly as the students' understanding of the topic.) In this study, the ability of students and the difficulty of items were estimated using Winsteps® Rasch measurement software (Lincare, 2011).

Measuring change using Rasch modeling. When using Rasch modeling to analyze change over time, Wright (2003) proposes two methods of structuring the data; stacking and racking. Stacking the data shows how the students have changed as a result of an intervention and racking the data shows how the items have changed. Recently, Cunningham and Bradley (2010) effectively applied these methods to study the impact of a teacher training program on student performance. In this paper, we apply the stacking and racking methods to the pretest and posttest data in order to investigate the change in student understanding and to determine on which ideas the unit had the greatest effect.

Stacking. Stacking data allowed us to study the effect of the unit on students' understanding of the targeted chemistry and biochemistry ideas. The stacked analysis was done by first preparing a data file that contained two rows of data per student (see Figure 1). One row contains their responses during the pretest and the second row contains their responses during the posttest. This analysis results in two ability measures per student: a pretest ability and a posttest ability. The difference between these ability measures represents the change in the students' understanding as a result of participating in the unit. If the unit was effective in improving students' understanding of the targeted chemistry and biochemistry ideas, we would expect the students' ability measures to increase from pretest to posttest.

Racking. Racking data permitted us to examine the effect of the unit on the items' difficulty level. The racked data set includes one row per student and two columns per item as illustrated in Figure 1. One column contains the students' pretest responses and the second column contains the students' posttest responses. The assumption here is that the items change in difficulty from pretest to posttest but the students remain unchanged. Racking the data in Rasch results in two difficulty measures per item: a pretest difficulty and a posttest difficulty. The difference in the difficulty measures indicates the degree to which the unit successfully targeted the ideas tested by the items. If the unit was effective in improving students' understanding of the targeted chemistry and biochemistry ideas, we would expect that the items would be easier for students to respond to after participating in the unit and, therefore, the difficulty measure for each item would decrease from pretest to posttest.

Figure 1: *Illustration of stacking and racking data for Rasch modeling*



Results and Discussion

Fit. Summaries of the Rasch fit statistics for the stacked and racked data in Tables 1 and 2 showed an adequate fit to the Rasch model. The separation indices and corresponding reliabilities were acceptable—i.e., greater than 2, according to Wright and Stone (2004). The separation index represents the spread of the abilities or difficulties and indicates the approximate number of different levels of difficulty or ability. Additionally, the standard errors for the items and students were small. 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). In the stacked data, the infit mean-square values for 8 students were not in this range. In the racked data, the infit mean-square values for 5 items were not in this range. We focused on the infit statistics because they gave more weight to the responses of students with abilities closer to the item difficulty, whereas outfit statistics are unweighted and, therefore, are more sensitive to outlying scores. The fit statistics for the racked data flagged 6 items that had negative point-measure correlations, which indicated that the pattern of student responses did not follow expectations (i.e. students with high ability measures were responding incorrectly to items with difficulties lower than their ability measure).

Table 1: *Fit statistics for the stacked data*

	Item			Person		
	Min	Max	Median	Min	Max	Median
Standard error	0.15	0.52	0.19	0.35	1.84	0.40
Infit mean-square	0.73	1.42	0.97	0.68	1.50	0.99
Outfit mean-square	0.67	1.63	0.94	0.27	2.07	0.95
Point-measure correlation coefficients	-0.05	0.66	0.45	-0.17	0.72	0.35
Separation index (reliability)	2.84 (0.89)			2.44 (0.86)		

Table 2: *Fit statistics for the racked data*

	Item			Person		
	Min	Max	Median	Min	Max	Median
Standard error	0.20	1.06	0.25	0.26	1.84	0.29
Infit mean-square	0.70	1.41	0.97	0.71	1.45	0.98
Outfit mean-square	0.46	1.83	0.98	0.57	2.20	0.95
Point-measure correlation coefficients	-0.30	0.74	0.36	-0.09	0.68	0.43
Separation index (reliability)	2.13 (0.82)			2.88 (0.89)		

Stacked method: Changes in student understanding. The data were stacked to investigate the changes in students' understanding of chemistry and biochemistry. Figure 2 shows the Wright map for the stacked data. The map shows the range of item difficulties on the left side and student abilities on the right side. Low ability/difficulty is represented at the bottom of the map and high ability/difficulty is represented at the top of the map. The mean of the item difficulties was set at zero. When item difficulty and student ability level match, the student has a 50% chance of answering the item correctly. The student abilities are divided into two groups; one for the students' performance on the items during the pretest and one for the students' performance on the items during the posttest. The students at the urban school are represented by "U"s and the students at the suburban school are represented by the "S"s. The map reveals that the ability levels of the students during the posttest are higher than the ability levels of the students during the pretest. This indicates that the students' understanding of the chemistry and biochemistry ideas increased between the pretest and the posttest. The map also shows that the difficulty range for the items on the pre-/posttests matches the urban students' ability range well. On the other hand, there are little to no items that match the suburban students on the upper end of the ability range.

The ability measures for 104 of the 120 students increased from pretest to posttest. A paired samples t-test was used to investigate the significance of the change in students' abilities. Table 3 presents a summary of the pre- and posttest abilities for the students at the two schools. Overall, the posttest abilities were significantly higher than the pretest abilities for both schools (Suburban: t = -10.93, p <.001; Urban: t = -5.44; p < .001). Additionally, the effect sizes were large for both schools.

Table 3: *Summary of pretest and posttest student measures*

		Min	Max	Median	Mean	SD	Effect size[a]
Suburban School (N=91)	Pretest	-1.78	5.08	-0.08	-0.02	0.97	1.11
	Posttest	-1.98	5.20	0.96	1.06	1.28	
Urban School (N=29)	Pretest	-2.26	0.31	-0.97	-0.96	0.68	1.72
	Posttest	-1.50	1.95	-0.36	-0.21	0.91	

[a] Effect size calculated by dividing the difference of the means by the pretest standard deviation (SD)

No significant differences were observed between the gains of males and females at either school (Suburban: t = -0.514, p >.05; Urban: t = 1.945; p > .05). Furthermore, there was no significant difference in the gains of the students at the urban school compared to the gains of the students at the suburban school (t = 1.730, p > .05).

Figure 2: *Wright map from stacked analysis*

```
    ITEM - MAP - PERSON
      <rare>|<more>
 4          +          S           |      SSSSS
            |                      |
            |      Pretest         |      Posttest
            |                      |
            |                      |
            |                      |
            |                      |      SS
            |                      |
 3          +                      |
            |                      |
            |                      |      SS
            |T                     |
            |                      |
            |                      |      SS
         X  |          S           |
            |                      |      SSSSSSS
 2          +                      |U
            |                      |U     SSSSSS
         X T|                      |
            |          S           |      SSSSS
            |          SS          |      SSSSS
         X  |S                     |      S
         X  |          SSS         |      SS
         X  |          SS          |      SSSSSS
 1       X  +                      |U     SSSSSS
           S|          SS          |U     SSSSSSSS
       XXX  |          SSSS        |U
        XX  |                      |UU    SS
     XXXXX  |          SSSSSSSS    |      S
      XXXX  |          SSSSSS      |      SSSS
     XXXXX  |M UU      SSSSS       |UU    SSSS
         X  |          SSSSSSSS    |U     SSS
 0      XX M+  U                   |U     SSSSSS
     XXXXX  |          SSSSSS      |UU    S
         X  |  UU      SSSSSSS     |U     SSS
       XXX  |          SSSS        |UUU   SSS
        XX  |  UUUUU   S           |U
       XXX  |          SSS         |
         X  |  UU      SSSSSSSSSS  |UU
           S|  UU      SSSS        |UU    SSS
-1       X  +S UU      S           |U     SS
       XXX  |  UUUUU   SSSSSSS     |UU
         X  |  UU      SSSSS       |UU
         X  |                      |U
        XX  |  U                   |U     S
            |  U                   |
           T|          S           |
            |                      |
-2          +  U                   |      S
            |T                     |
         X  |  UUU                 |          M = mean performance/difficulty
            |                      |          S = 1 standard deviation away from mean
            |                      |          T = 2 standard deviations away from mean
            |                      |          X = 1 item
            |                      |          U = 1 Urban student
            |                      |          S = 1 Suburban student
-3          +                      |
     <frequ>|<less>
```

Racked method: Changes in item difficulties. The data were racked to investigate the changes in item difficulties as a result of the intervention. Because the pretest at the suburban school did not include items targeting ideas about chemical reactions, the racked data set did not include the suburban students' posttest responses to these items. The Wright map for the racked data is shown in Figure 3. On this map, the range of item difficulties on the right side and student abilities on the left side. The item difficulties are divided into two groups; one for the pretest measures and one for the posttest measures. The map reveals that the item difficulties decreased from pretest to posttest. More specifically, the difficulties of 48 items decreased from pretest to posttest, the difficulty of one item remained the same, and the difficulties of three items increased. This suggests that overall the knowledge targeted by the items was learned by the students who participated in the unit.

Table 4 shows a summary of the item difficulties broken down by idea. Paired samples t-tests were used to determine the significance of the changes in mean item difficulty. The mean item difficulty for items aligned to the ideas about Chemical Reactions, Conservation, Food, Photosynthesis, and Animal Growth decreased significantly from pretest to posttest. We also saw a decrease in item difficulty for the item aligned to ideas about Food & Storage and for the item aligned to ideas about Photosynthesis & Plant Growth. These results indicate that the unit was successful in teaching students these ideas.

For the items aligned to the ideas about Food & Photosynthesis and for the items aligned to idea about Storage, we did not see a significant change in difficulty. These results are not surprising because the Food & Photosynthesis items tested students' understanding of what is (sugars that plants make) or is not (water, carbon dioxide, sunlight, and minerals) considered food for plants, and the unit did not explicitly focus on defining food for plants. Half of the items aligned to the idea about Storage tested students' understanding of the storage structures in plants and animals. While the unit did include a lesson on the storage of carbohydrates in both plants and animals, it did not emphasize the structures used to store molecules from food. For this reason, we would not expect to see a significant decrease in difficulty for these items.

Figure 3: *Wright map from racked analysis*

```
      PERSON - MAP - ITEM
          <more>|<rare>
 4           X  +                                            |
                |             Pretest                        |       Posttest
                |                                            |
                |                                            |
                |                                            |
                |                                            |
                |  ME325                                     |
                |                                            |
 3           X  +                                            |
                |                                            |
                |                                            |
                |                                            |
                |                                            |
             X  |  ME954  ME956                              |
                |                                            |
                |                                            |
 2          XX T+T                                           |
             X  |                                            |
                |                                            |
            XX  |  ME017  ME096  ME169                       |
            XX  |  ME309  SC753                              |
             X  |  ME955  SC505  SC883                       |
         XXXXX  |                                            |ME325
   XXXXXXXXXXX S|  ME367  ME1591                             |ME096  ME956
 1           X  +S ME1123                                    |
           XXX  |  ME1093 SC673                              |
           XXX  |  ME109  ME1042 ME1533                      |
    XXXXXXXXXX  |  ME1012 ME1103 ME296  SC354  SC913         |ME017  SC673
           XXX  |  ME1062 ME1473 SC933                       |ME1123 ME1472
       XXXXXXX  |  ME1125 ME1501 ME0412                      |ME1012 ME1103 ME1125
         XXXXX  |  ME1541                                    |ME1042 ME147  ME1533 ME169  ME0412
        XXXXXX M|  ME026  ME0411                             |ME109  ME1062
 0       XXXXX  +M ME1472 ME1532 ME1611 ME0410 SC765  SC944  |ME1473 ME1541 ME367
             X  |  ME1063 SC665  SC903                       |ME1093 SC753  SC883
                |                                            |
          XXXX  |                                            |ME1532 ME026  ME954
     XXXXXXXXX  |                                            |
       XXXXXXX  |  ME1491                                    |ME0410 ME0411 ME955
        XXXXXX  |  ME982                                     |ME1591 ME309  SC933
          XXXX S|                                            |SC354
         XXXXX  |  ME147  ME1443 ME844  ME853                |ME1063 ME296  SC765
-1              +S ME1002 ME138                              |ME1601
        XXXXXX  |  ME1601                                    |SC505  SC913  SC944
            XX  |                                            |ME1611 ME844  SC903
            XX  |                                            |SC665
           XXX  |                                            |ME1491 ME1501 ME853  ME982
                |                                            |
               T|                                            |
             X  |                                            |ME1443
-2              +T ME992                                     |ME1002 ME138
                |                                            |
                |                                            |
                |                                            |
                |                                            |ME992
                |                                            |
                |                             M = mean performance/difficulty
                |                             S = 1 standard deviation away from mean
-3              +                             T = 2 standard deviations away from mean
          <less>|<frequ>                      X = 1 student
```

Table 4: *Summary of item difficulties by idea*

Idea		Min	Max	Median	Mean	t	Sig.
Chemical Reactions	Pretest	-0.13	1.36	0.66	0.66	4.93	<.01
(N=5)	Posttest	-0.84	-0.38	-0.95	-0.64		
Conservation	Pretest	-0.14	1.54	0.34	0.56	8.26	<.001
(N=6)	Posttest	-1.21	-0.08	-0.95	-0.74		
Food	Pretest	-1.95	1.64	0.00	-0.09	4.93	<.001
(N=18)	Posttest	-2.45	0.65	-0.94	-0.76		
Photosynthesis	Pretest	0.67	2.40	1.10	1.39	4.52	<.01
(N=6)	Posttest	-0.88	1.09	-0.23	-0.07		
Animal Growth	Pretest	0.78	3.22	1.37	1.69	3.31	<.05
(N=4)	Posttest	0.01	1.31	0.62	0.64		
Food & Photosynthesis	Pretest	-0.04	0.99	0.32	0.35	2.66	n.s.
(N=5)	Posttest	-0.47	0.47	0.23	0.02		
Storage	Pretest	-1.00	0.75	-0.42	-0.28	1.05	n.s.
(N=6)	Posttest	-2.01	0.23	-0.17	-0.62		
Food & Storage	Pretest				1.48		
(N=1)	Posttest				-0.61		
Photosynthesis & Plant Growth	Pretest				1.65		
(N=1)	Posttest				0.31		

Distractor analysis. An analysis of the students' selection of distractors was performed to gain insight into the effects the curriculum unit had on students' ideas and misconceptions. Table 5 summarizes the changes in students' responses to an item that was aligned to both the Photosynthesis and Plant Growth ideas. The item asked students where most of the material that makes up a wood table originally comes from. During the pretest (and for the national ninth grade sample), the most popular answer choice was that the material came from minerals in the soil, which is based on the misconception that minerals are a source of food for plants (Vaz et al, 1997). After participating in the unit, fewer students chose this answer, and more students chose the correct answer that the material came from carbon dioxide in the air. This trend was noticed in both of the pilot test schools but was only statistically significant for the suburban school. The unit included several activities that targeted the ideas that carbon dioxide in the air provides the carbon atoms for glucose, which is then used to make plant structures. For example, the students engaged in a discussion of radiolabeling experiments that traced carbon atoms from carbon dioxide molecules to glucose molecules in plants and used LEGO® bricks to model the photosynthesis reaction and the polymerization reaction for building cellulose from glucose.

Table 5: *Responses to an item about where the material that makes up a wood table comes from.*

	Suburban School				Urban School			
Answer choice	Pre	Post	χ^2	Sig.	Pre	Post	χ^2	Sig.
Minerals in the soil	51.6%	25.6%	12.99	<.001	65.5%	44.8%	2.51	n.s.
CO_2 in the air	23.1%	48.9%	13.10	<.001	20.7%	41.4%	2.90	n.s.

Table 6 provides an example of the change in students' performance on an item about where trees get energy to form new leaves in the spring. The item was aligned to both the Food and Storage ideas. During the pretest at the suburban school (and for the national ninth grade sample), the most popular answer choice was that the energy came from minerals from the soil. During the pretest at the urban school, the most popular answer choice was that the energy came from water from the soil. After participating in the unit, significantly more students at both schools chose the correct answer that the energy came from the stored starch molecules. Activities within the unit that may have contributed to students' improvement on this item included class discussions focused on questions such as "Why do radish seeds that are planted too deeply die before they can grow to the surface?", images of storage phenomena, and experiences linking structural models of the starch polymer and glucose monomer to a LEGO® model of glucose.

Table 6: *Responses to an item about where trees get the energy to make new leaves.*

	Suburban School				Urban School			
Answer choice	Pre	Post	χ^2	Sig.	Pre	Post	χ^2	Sig.
H_2O molecules in the soil	24.2%	8.9%	7.65	<.01	35.7%	17.9%	2.28	n.s.
Minerals from the soil	35.2%	12.2%	13.15	<.001	28.6%	14.3%	1.70	n.s.
Stored starch molecules	27.5%	71.1%	34.48	<.001	17.9%	50.0%	6.45	<.02

Research has shown that a particularly resilient 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/repair body parts (Smith & Anderson, 1986). Approximately 58% of the ninth grade students in our national field test sample held this misconception. Because we were aware of the prevalence of this misconception during the curriculum development stage, several activities in the unit were designed to give students evidence that contradicted this misconception. For example, a chemistry lesson focused on how the formation or "growth" of a polymer chain occurs by adding atoms, which account for the added mass. In the biochemistry lessons, this idea of growth by forming polymers from monomers and forming body structures from polymers was illustrated with protein polymers (e.g. keratin in human hair, collagen in tendons, and actin and myosin in muscle) and with carbohydrate polymers (e.g. chitin in crayfish exoskeleton and cellulose in celery stalks). After participating in the unit, the percentage of students choosing distractors based on this misconception dropped from 73% to 7% at the suburban school and from 48% to 6% at the urban school.

Conclusions

This paper reports on the pilot test of a new curriculum unit that targets foundational chemistry and biochemistry ideas. Designed to emphasize the underlying molecular explanations for observable biological events in the real world, the unit aims to improve on currently available materials by engaging students with phenomena that occur in 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.

Rasch modeling was used to investigate the change in student understanding from pretest to posttest and the impact of the unit on the difficulty of the items on the pre/posttests. The stacked data set showed that, overall, the students' understanding of chemistry and biochemistry improved significantly. The racked data set showed that most of the items got easier from pretest to posttest. An analysis of the students' answer choice selections also revealed a decrease from pretest to posttest in the popularity of several misconceptions.

Next steps. Following the Year 1 pilot of the curriculum unit, the AAAS and BSCS teams met with the pilot teachers and planned for extensive revisions that included reducing the number of learning goals to allow a more focused and coherent treatment of the following overarching goal:

> Students will be able to use the idea that all matter is made out of atoms to explain growth and repair in living organisms (plants and animals). In order to grow and repair body structures, plants and animals build polymers through chemical reactions from subunits (monomers) that plants make through other chemical reactions. Through all this, atoms are rearranged and conserved.

These revisions have been implemented and the revised unit, including teacher materials and professional development, is being tested this Spring with nine teachers in four states.

Given the extent of these revisions, the pre/posttest of student content knowledge is being revised accordingly to create an instrument that more closely aligns to the ideas targeted in the unit. For example, many new items have been developed to assess the idea that plants grow by synthesizing polymers from the glucose monomers they make, for which there was only one item in the Year 1 pre/posttest. In addition, whereas the Year 1 pre/posttest used exclusively multiple choice items, the revised pre/posttest requires students to write explanations for their responses. This will give us additional information about the ideas and misconceptions students are using to answer the items.

Acknowledgements

The authors would like to acknowledge the members of the BSCS curriculum development team: Janet Carlson, Kathleen Roth, Rebecca Kruse, Elaine Howes, and Aleigh Raffelson. The authors would like to thank Mary Koppal of Project 2061 for her help in editing the curriculum unit and Abigail Burrows of Project 2061 for her help in coordinating the pilot test of the unit.

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

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

Anderson, C. W., Sheldon, T., & Dubay, J. (1990). The effects of instruction on college nonmajors' conceptions of respiration and photosynthesis. *Journal of Research in Science Teaching*, *27*(8), 761–776.

Bell, B. (1985). Students' ideas about plant nutrition: what are they? *Journal of Biological Education*, *19*(3), 213-218.

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (Second ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Cunningham, J. D., & Bradley, K. D. (2010, April) *Applying the Rasch model to measure change in student performance over time*. Paper presented at the American Educational Research Association Annual Meeting, Denver, CO.

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., 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.

Linacre, J.M. (2011). Winsteps® (Version 3.73.0) [Computer Software]. Beaverton, Oregon: Winsteps.com. Retrieved January 1, 2011. Available from http://www.winsteps.com/.

Liu, X., & Boone, W. J. (2006). Introduction to Rasch measurement in science education. In X. Liu & W. J. Boone (Eds.), *Applications of Rasch Measurement in Science Education* (pp. 1-22). Maple Grove, Minnesota: JAM Press.

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 Assessment Governing Board. (2008). Science Framework for the 2009 National Assessement of Educational Progress, Washington, DC: Author.

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: Author.

Roseman, J. E., Stern, L., & Koppal, M. (2010). A method for analyzing the coherence of biology textbooks: A study of its application to the topic of matter and energy transformations in four textbooks. *Journal of Research in Science Teaching*, *47*(1), 47-70.

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.

Wright, B. D. (2003) Rack and stack: Time 1 vs. Time 2. *Rasch measurement transactions*. *17*(1), 905-906. Retrieved March 12, 2012, from http://www.rasch.org/rmt/rmt171a.htm.