

Engaging Undergraduate Biology Students in Scientific Modeling: Analysis of Group Interactions, Sense-Making, and Justification

Andrea M.-K. Bierema,^{1*} Christina V. Schwarz,² and Jon R. Stoltzfus³

¹Center for Integrative Studies in General Science and Department of Integrative Biology,

²Department of Teacher Education, and ³Department of Biochemistry & Molecular Biology, Michigan State University East Lansing, MI 48824

ABSTRACT

National calls for improving science education (e.g., *Vision and Change*) emphasize the need to learn disciplinary core ideas through scientific practices. To address this need, we engaged small groups of students in developing diagrammatic models within two (one large-enrollment and one medium-enrollment) undergraduate introductory biology courses. During these activities, students developed scientific models of biological phenomena such as enhanced growth in genetically modified fish. To investigate whether undergraduate students productively engaged in scientific practices during these modeling activities, we recorded groups of students as they developed models and examined three characteristics: how students 1) interacted with one another, 2) made sense of phenomena, and 3) justified their ideas. Our analysis indicates that students spent most of the time on task, developing and evaluating their models. Moreover, they worked cooperatively to make sense of core ideas and justified their ideas to one another throughout the activities. These results demonstrate that, when provided with the opportunity to develop models during class, students in large-enrollment lecture courses can productively engage in scientific practices. We discuss potential reasons for these outcomes and suggest areas of future research to continue advancing knowledge regarding engaging students in scientific practices in large-enrollment lecture courses.

INTRODUCTION

National calls for transformation of science, technology, engineering, and mathematics (STEM) education at both the K–12 and undergraduate levels emphasize the need for students to develop understanding of core disciplinary ideas by engaging in scientific practices (American Association for the Advancement of Science [AAAS], 2011; National Research Council [NRC], 2012; Cooper, 2015). Incorporating scientific practices into undergraduate STEM courses provides a promising route for active learning that directly connects to how scientists construct, evaluate, and apply knowledge (Cooper, 2016). These practices include arguing with evidence, analyzing and interpreting data, developing and using models, and constructing explanations. As undergraduate STEM instruction shifts to include scientific practices, there is a need to develop instructional approaches and materials that integrate scientific practices while emphasizing coherent understanding of core disciplinary ideas rather than memorization of fragmented details (Cooper *et al.*, 2015; Zagallo *et al.*, 2016). It is also essential to provide evidence that these instructional approaches promote scientific knowledge and competencies (Knight *et al.*, 2013).

The practice of scientific modeling allows scientists as well as others to make sense of phenomena and lies at the core of scientific endeavors (White and Frederiksen, 1998;

Nancy Pelaez, *Monitoring Editor*

Submitted February 9, 2017; Revised September 11, 2017; Accepted September 13, 2017

CBE Life Sci Educ December 1, 2017 16:ar68

DOI:10.1187/cbe.17-01-0023

*Address correspondence to: Andrea M.-K. Bierema (abierema@msu.edu).

© 2017 A. M.-K. Bierema *et al.* CBE—Life Sciences Education © 2017 The American Society for Cell Biology. This article is distributed by The American Society for Cell Biology under license from the author(s). It is available to the public under an Attribution–Noncommercial–Share Alike 3.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/3.0>).

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

Morgan and Morrison, 1999; Nersessian, 1999; Schwarz and White, 2005; Windschitl *et al.*, 2008; NRC, 2012; Passmore *et al.*, 2013). Scientists often use modeling to aid in constructing new knowledge by developing external representations that allow them to make sense of the structures and mechanisms underlying the phenomena (Windschitl *et al.*, 2008; Schwarz *et al.*, 2009; Passmore *et al.*, 2013). Mechanistic scientific models can be particularly useful in reasoning about phenomena, because they can help unpack what is occurring at the scalar level below the level of the observed phenomenon and use the behavior of relevant factors at that scale to explain the observed phenomenon (Wilensky and Resnick, 1999; Van Mil *et al.*, 2016).

Differences in model-based instructional approaches influence how students engage with this scientific practice (Svoboda and Passmore, 2010). Activities can focus primarily on engaging students in developing models to explain a phenomenon the students do not yet understand—similar to how scientists use modeling—or focus on providing students with developed canonical models so they learn about established disciplinary ideas (Schwarz and White, 2005; Passmore *et al.*, 2013). While the two approaches do not have to be at odds with one another, the tension between them should be considered in course design, because each has benefits and drawbacks (Passmore *et al.*, 2017). For example, developing models can enable participants to put together ideas and work to make sense of a phenomenon in ways that mirror scientists' efforts to understand novel phenomena. However, the practice of developing models does not necessarily lead students to reproduce scientifically accepted ideas about the phenomenon in question. At the same time, while providing students with canonical models gives students exposure to the established ideas, merely providing canonical models does not effectively enable students to understand them or make sense of the mechanisms underlying the model. Thus, a modified or combined approach may be particularly effective.

Actively engaging students in model development can provide students opportunities for collaborative learning, sense-making, and development of model-based reasoning skills. Because modeling is inherently a social endeavor (NRC, 2012), model-based instruction is well suited for engaging students in socially mediated construction, evaluation, and revision and the benefits that come from cooperative learning (Johnson *et al.*, 1998). Given the opportunity to work on teams with a scientific mentor, undergraduates can use modeling to make sense of phenomena (Passmore *et al.*, 2013). The construction of visual representations as part of classroom activities can also promote model-based reasoning in students (Quillin and Thomas, 2015).

Various studies on using modeling in undergraduate classrooms have revealed clear benefits of model-based instruction. Undergraduates in sections of physics and chemistry courses that focus on modeling outperform students from lecture-based sections of those courses on measures of conceptual understanding of core disciplinary ideas (Brewer *et al.*, 2010; Williams *et al.*, 2015). Biology instruction based on diagrammatic modeling tools like concept maps and structure/behavior/function type models reveals students' understanding of core disciplinary ideas and increases students' ability to create scientifically accurate models (Luckie *et al.*, 2011; Dauer *et al.*, 2013). Undergraduates in teaching real data interpretation with models (TRIM)-based courses use models as tools during data interpretation

and, in some cases, generate new ideas based on the model provided (Zagallo *et al.*, 2016).

Evidence supporting benefits of model-based instruction in large undergraduate classrooms has come primarily from measuring student learning following model-based instruction (Brewer *et al.*, 2010; Williams *et al.*, 2015) or by assessing the products of modeling activities (Luckie *et al.*, 2011; Dauer *et al.*, 2013; Williams *et al.*, 2015). This is problematic, because it does not identify potential mechanisms by which students' engagement in modeling improved their learning.

To determine whether and how activities engage students in scientific thinking and reasoning, it is essential to know what students are doing during the activities and not just how they perform on assessments after these activities. A case study of small groups of students working with a faculty mentor shows that modeling supports learning across the curriculum and can promote productive engagement in the science practice of modeling (Svoboda and Passmore, 2010, 2013). Analysis of recordings of small groups of students working in peer-led guided-inquiry teams shows that students in this setting successfully engage in scientific argumentation (Kulatunga *et al.*, 2013). Several recent studies recorded student discussions during active-learning exercises in large undergraduate STEM lecture courses and provided evidence that undergraduate students can successfully engage in scientific argumentation during clicker questions (Knight *et al.*, 2013) and can use existing scientific models to aid in data analysis (Zagallo *et al.*, 2016). To gain further insight into how developing and using models can improve students' engagement in scientific modeling and understanding of disciplinary ideas, we analyzed recordings from groups of undergraduate biology students as they developed models during classroom activities.

This study describes a novel model-based instructional approach and provides needed insight into student engagement in model development and revision by analyzing recordings of student discussions made while students were developing models to explain biological phenomena related to course content. We designed the activities in our study using a modified model development approach in which the model development process is supplemented with prompts to support developing causal mechanisms consistent with empirical evidence and theoretical knowledge. Using this approach, one can introduce established disciplinary ideas during model development to help students account for gaps in their explanations of the scientific phenomena and help move students toward canonical scientific understanding of the phenomenon under consideration while still exploring and attempting to make sense of the phenomenon. We incorporated these activities into two very different undergraduate biology courses, an intermediate-enrollment nonmajors course focusing on the use of biotechnology and a large-enrollment introductory cell and molecular biology course for life science majors with different instructors. Analyzing data from both of these contexts gives us a better understanding of how students engaged in modeling activities under different circumstances and provides information about the range of issues that can arise.

RESEARCH QUESTIONS

To determine how students engaged in modeling and whether they did so productively, we sought to address the following research questions:

TABLE 1. Comparison of all students in each course with students recorded in each course for mean incoming GPA and final grade in the course using independent *t* tests

Course	Measure	All: M (SD)	Recorded: M (SD)	<i>t</i> (df)	<i>p</i>
Biotechnology	Incoming GPA	2.96 (0.90)	3.04 (0.53)	0.553 (11)	0.592
	Final grade	2.7 (1.1)	2.4 (1.1)	0.86 (11)	0.41
Cell and Molecular	Incoming GPA	3.06 (0.82)	2.95 (1.1)	0.409 (17)	0.688
	Final grade	2.5 (1.0)	2.4 (0.53)	0.21 (17)	0.83

1. How did students use the class time provided to develop the model? This question is important for addressing whether students were off task and whether they were developing their models in a cooperative manner (e.g., clarifying ideas, agreeing, or disagreeing; hereafter referred to as “student interactions”).
2. Did students engage in sense-making while developing their models? If so, how did they engage in sense-making? In other words, did they generate questions regarding how and why a phenomenon occurs and seek to answer those questions? This question is important because making sense of phenomena is a major goal of constructing scientific models and evidence that students participated in sense-making would support the idea that they were engaging in scientific modeling.
3. When students collaborated to develop their models, how did they justify their decisions regarding the models? This question is important because scientists use evidence to support their claims during model development and student justification using evidence would support the idea that they were engaging in scientific modeling.

METHODS

Research Design

To determine what students were doing during modeling activities, we used tablet computers equipped with audio and screen capture software to record groups of consenting students as they developed their models. The tablet computers were equipped with a stylus to aid in drawing. Most groups contained three students. We interviewed a subset of these students to provide information to guide analysis of the recordings and supplement findings from these analyses. We analyzed recordings using a combination of coding and descriptive statistics, interviews using coding, and student demographic data using independent *t* tests and χ^2 tests for goodness of fit. We performed all statistical analyses using SPSS, version 20 (IBM, Armonk, NY).

Participants

This study took place in two introductory biology courses at a large, public, land-grant, R1 university located in the Midwest:

a biotechnology course for non-science majors and an introductory cell and molecular biology course for life science majors. The institution's review board approved all research investigations (IRB #x14-308e).

In the biotechnology course, we analyzed recordings of the same four groups completing two different modeling activities. Because the groups were constant throughout the semester, a total of 12 students from this course participated in the study. Three of the groups were in J.R.S.'s section, and one group was in another section of the course. In the cell and molecular biology course, five groups participated in the first modeling activity, and three groups participated in the second activity. In this course, students formed new groups for each activity, but several groups from the second activity included students who also participated in the first activity, resulting in a total of 18 students from this course who participated in the study.

We contacted all students who participated in the study with a request to be interviewed in exchange for a gift card. Six students from the biotechnology course and five students from the cell and molecular biology course agreed and were interviewed.

We obtained student university grade point average (GPA) before the start of the course, gender, and race/ethnicity from the university registrar through standard university data requests. In both courses, students who participated in the study were representative of the class as a whole with respect to incoming university GPA and final grade in the course (Table 1). The genders of students who participated in the study in the biotechnology course were representative of the course as a whole, while in the cell and molecular biology course there was an overrepresentation of females in the study (Table 2). Students in the study were predominantly white (83% and 89% for the biotechnology and cell and molecular biology courses, respectively). We note that it is important to keep in mind the preponderance of female and white students in our samples when interpreting the results and applying them to other contexts. It also highlights the need for larger studies that specifically focus on recruiting males and non-white students.

TABLE 2. Comparison of observed number of female and male students who participated in the study in each course with the expected number based on the total frequencies of females and males in each course using χ^2 tests for goodness of fit

Course	Gender	Observed	Expected	χ^2 (df)	<i>p</i>
Biotechnology	Female	8	6.9	0.437 (1)	0.509
	Male	4	5.1		
Cell and Molecular	Female	16	10.4	7.3 (1)	0.007
	Male	2	7.6		

Instructional Context

The biotechnology course was designed for non-science majors and fulfilled part of these students' general university requirements. Two different sections of the biotechnology course participated in this study. J.R.S. designed the modeling activities and taught one section. He had a decade of experience teaching undergraduates, including upper-level biochemistry for biochemistry majors and introductory cell and molecular biology for life science majors. This was the 11th semester J.R.S. taught this biotechnology course. An instructor who was familiar with the research study but not involved in the study taught the second section. This instructor, who had not previously taught the course but had several years of undergraduate biology teaching experience, observed J.R.S.'s section before teaching his own section and used similar activities and assessments. Each section had one undergraduate learning assistant. J.R.S.'s section had 50 students and took place in a SCALE-UP (Student-Centered Active Learning Environment with Upside-down Pedagogies) classroom with round tables, movable chairs, and computer monitors for each group (Beichner *et al.*, 2007; Stoltzfus and Libarkin, 2016). The other section had 67 students and took place in a similar classroom with round tables and movable chairs but no computer monitors for the groups. Both sections used a flipped-classroom format in which the instructor expected students to read material and view informational videos before class time and then participate in activities during class.

Modeling activities in this course typically took students 30–50 minutes to complete within the 80-minute class sessions. There were 10 modeling activities throughout the semester. Students worked within the same groups during the entire semester, with the exception that students could submit requests for reassignment halfway through the semester. The instructor assigned formal groups during the second week of class based on responses to a survey students completed. The instructor aimed to create group sizes of three students, but group size varied from two to five students as students dropped the course and groups combined. The instructor grouped students primarily based on similar amounts of time they indicated they intended to spend working on the course outside class time.

The cell and molecular biology course was designed for life science majors. It is the first semester in a two-semester introductory biology series that is a prerequisite for many advanced biology courses. One section of the cell and molecular biology course participated in the study, and the instructor knew of the research study but was not involved in the study. This instructor had several years of experience teaching both upper-level genetics courses and biotechnology for non-science majors science courses, but it was the first time this instructor taught this cell and molecular biology course. She taught this section in collaboration with J.R.S., who was concurrently teaching a parallel section of the course that incorporated the same modeling activities but was not included in this study. Two-hundred eighty students took the course in a lecture hall setting. Most of the course instruction on days in which students did not construct models consisted of lectures with integrated active-learning activities such as clicker questions. Five of the class days in this course contained modeling activities in which students worked in groups of three during most of the 80-minute class period to develop models. Students formed their own groups at

the beginning of each modeling activity. In addition to the instructor of the course, J.R.S., three graduate student teaching assistants, and two undergraduate learning assistants aided students during the modeling activities.

In both courses, students were expected to complete homework before the start of class. The homework included a reading from the popular press about a biological phenomenon or process and questions to answer regarding basic biological concepts needed to explain that phenomenon or process. In class, the instructor provided students with an instruction sheet to guide model development. Groups used either tablets with a stylus or paper and pencil/pen to construct a diagrammatic, concept-process model (Harrison and Treagust, 2010) and a written explanation incorporating a causal mechanism. The instructor then selected one or two groups, typically at random, to present their models to the entire class. Following the model presentation, groups discussed the presented models and compared them with their own model. When time allowed, the instructor randomly chose individuals from several groups to provide feedback about the models. Finally, groups revised their models before final submission. When using a paper and pencil/pen, groups captured their models digitally using the camera on their mobile devices. Groups submitted their models and received feedback on their models from instructors using the university's learning management system.

Modeling activities composed 3.7% of the total grade in the biotechnology course and 4% of the total grade in the cell and molecular biology course. The instructional teams from each course graded the models based on completion of key aspects of the model required in the basic instructions. Although instructors did not grade models based on biological accuracy, students did receive additional feedback on the scientific accuracy of biological concepts found in their models.

Modeling Activity Development

We initially developed the modeling activities described in this paper during the process of revising the biotechnology course for delivery in a SCALE-UP instructional space (Stoltzfus and Libarkin, 2016) with fewer than 70 students per section. We designed the modeling activities using ideas from model-based inquiry (Windschitl *et al.*, 2008) and a learning progression for scientific modeling (Schwarz *et al.*, 2009). The course used a flipped instructional approach in which students used readings and videos to acquire basic knowledge of core disciplinary ideas before class and applied these ideas during class to develop models, explanations, and evaluations of claims in the popular press. Because the SCALE-UP instructional space is ideal for students to work in groups and cooperative learning has many benefits (Johnson *et al.*, 1998), students developed models in groups. Choosing groups to present during the activity and specifically connecting the modeling activities to individual assessments on exams provided some of the accountability needed for productive cooperative learning (Johnson *et al.*, 1991). Because previous analysis of recordings of students suggested that group activities involving comparisons promote sense-making (Young and Talanquer, 2013), we designed most of these activities so that groups constructed two or more diagrams depicting different versions of the phenomenon or process in which some aspect had changed (hereafter referred to as “scenarios”). Based on successful implementation of modeling activities in the

TABLE 3. Characteristics of activities analyzed for this study

Course	Activity	No. of scenarios	Timing in semester
Biotechnology	Self-renewal	1	Second out of 10 models
	Modification	2	Ninth out of 10 models
Cell and Molecular	Ras in cancer	4	Third out of five models
	Specialization	3	Fifth out of five models

biotechnology course for nonmajors, instructors incorporated modeling activities into several sections of an introductory cell and molecular biology course for life science majors.

Description of Modeling Activities Analyzed and Example Modeling Activity

Because of limited resources and the significant effort required to analyze each recording, we chose to carry out in-depth analysis on a subset of the available recordings rather than attempt superficial analysis of all available recordings. We analyzed two modeling activities in each course for a total of four modeling activities. We selected these activities to represent modeling from both courses, from various time points in the courses, from activities that targeted both cellular and organismal level phenomena, and from activities with differences in complexity regarding the number of scenarios to model (Table 3). We chose diverse activities to increase the chance that observations from these activities might generalize and to provide an opportunity to observe how differences between activities might influence model development. Instructional materials for these four activities are available in Supplemental Material 1.

All four modeling activities in this study relate to the core concept for biological literacy of information flow, transfer, and storage (AAAS, 2011). A key scientific practice objective for all

four activities is developing student's modeling abilities (NRC, 2012). However, the models draw on four different biological contexts and have four different content-focused objectives. The first modeling activity analyzed from the biotechnology course asked students to develop a model using ideas from mitosis to explain why the two cells produced during stem cell self-renewal have identical genetic information. The second activity analyzed from the biotechnology course asked students to develop a model using ideas from gene expression to explain why the gene added to AquAdvantage salmon (i.e., genetically modified salmon) caused these salmon to grow more quickly than unmodified Atlantic salmon. The first activity analyzed from the cell and molecular biology course involved the role of Ras signaling in cancer (Figure 1). The activity asked students to develop a model using ideas from cell signaling to explain why a mutation in Ras causes uncontrolled cell division and how a new cancer treatment drug can prevent this. The second modeling activity analyzed from this course asked students to develop a model using ideas from differential gene expression to explain the molecular similarities and differences between an unspecialized stem cell and specialized eye and heart cells.

The models from the biotechnology course required less complex mechanistic explanations than those from the cell and molecular biology course. For example, the underlying mechanism in both the genetic modification model from the biotechnology course and the cellular specialization model from the cell and molecular biology course is regulated gene expression. However, the model from the biotechnology course only incorporates the idea that promoters help control gene expression, while the model from the cell and molecular biology course incorporates ideas related to enhancers, specific transcription factors, and the role of DNA sequence in the control of gene expression.

To provide a more complete picture of modeling activity implementation, the following describes Ras signaling in cancer

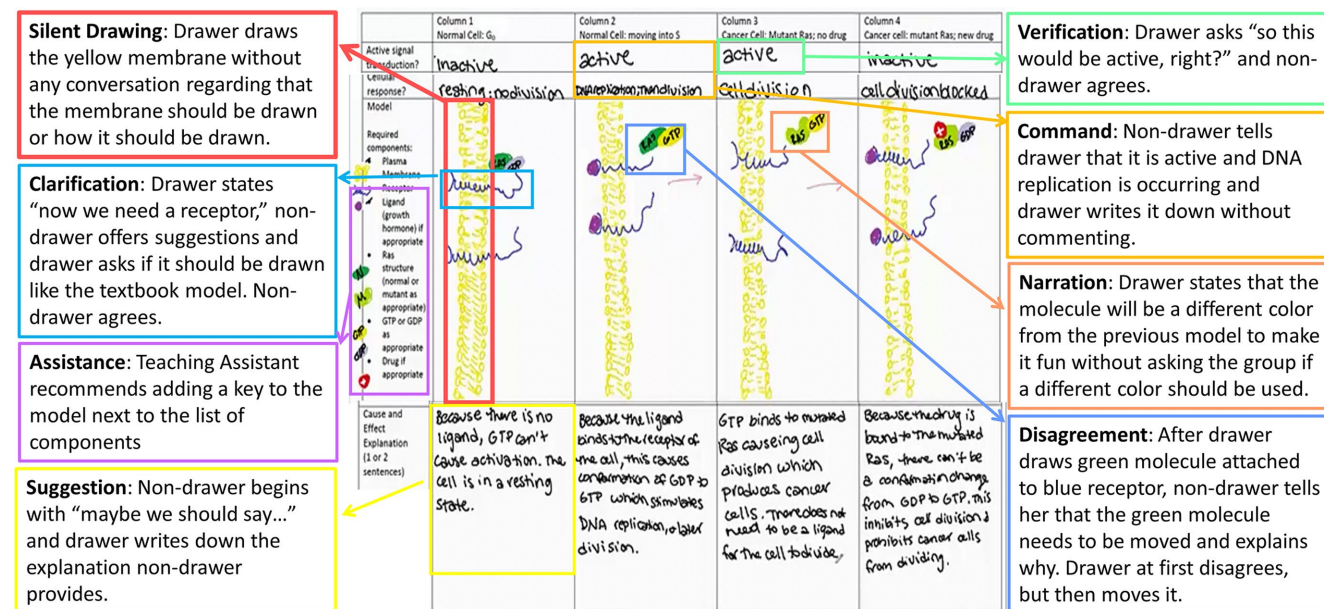


FIGURE 1. Example of an annotated student model from the cell and molecular biology course activity on Ras in cancer. The typed portions across the top and on the left side are part of the original template, and the handwritten portions represent the group's work. Each annotation describes the discussion that resulted in the addition of the model component indicated by the box on the model.

activity. This activity is one of the two activities analyzed from the cell and molecular biology course and requires students to model four scenarios related to the phenomenon. Before the activity, students engaged with core disciplinary ideas related to cell signaling in readings from the textbook and during previous lectures. Before the class meeting during which they developed the model, students read a *ScienceDaily* article describing a new drug that targets a specific type of cancer (University of California, 2013). During class, the modeling activity instructed students to create a model that illustrated the structure–function relationship of Ras, a protein involved in a signaling pathway that causes cells to divide when growth factors are present. The activity asked students to model the structure–function relationship of Ras in four related scenarios: 1) a normal cell not undergoing cell division, 2) a normal cell preparing for cell division, 3) a cell in which one allele of Ras has been mutated from a proto-oncogene to an oncogene, and 4) the cell from scenario 3 with a new drug that impacts mutant Ras. The instructor supplied students with a template on which students drew a diagrammatic model and provided a cause-and-effect statement for each of the four scenarios (Figure 1). The textbook depicted the scenario involving normal cell division. Students had to apply that information, information previously learned in class, and information from the preclass homework to develop remaining scenarios in the model.

Data Collection

We collected data during the in-class modeling activities and during interviews that subsequently took place outside class. During the in-class modeling activities, participating groups recorded their discussions and concurrent diagrammatic modeling for the entire duration of the activity using Microsoft Surface Pro 2 tablet computers running audio-recording and screen capture software (Camtasia Relay, version 4.3.1; Techsmith, Okemos, MI).

To provide additional insights into what students were doing during the modeling activities, A.M.-K.B. conducted interviews with students who participated in the in-class portion of the study. We asked students questions regarding a particular model

and the modeling activities in general. Interviews consisted of four main parts. First, we asked students to describe what their models explained and which components of the models were most important for the explanation. Second, we asked about any revisions they made to the models, why they made those revisions, and whether they would have made any further changes to the models. The third part of the interview asked students to describe how their groups worked as a team and how they resolved any disagreements. Finally, we asked students what they thought was the purpose of the modeling activities and about the purpose of doing the activities in groups instead of individually.

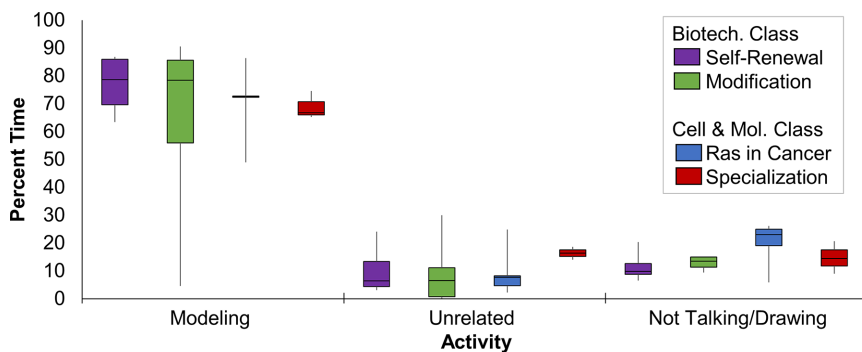
The interviews were semistructured, such that we tailored each one to the respective modeling recording (see Supplemental Material 2 for the general interview protocol). For instance, we adjusted the protocol to account for the types of revisions across the groups, taking into account that some groups did not make any revisions. Also, we allowed students to discuss additional ideas about the activities outside the protocol questions in case there were important aspects to the modeling activities that we had not initially considered.

A.M.-K.B. interviewed students from the biotechnology course toward the end of the semester during the week of the last modeling activity and students from the cell and molecular biology course a few weeks after the course ended (i.e., the beginning of the following semester). All interviews were audio-recorded and transcribed for analysis.

ANALYSIS OF GROUP MODELING RECORDINGS

Our initial goal for analyzing the modeling activities was to determine how much time students spent off task or on task, as well as whether they engaged in fact-based discussion or sense-making (e.g., Young and Talanquer, 2013). However, we found this coding scheme did not adequately capture the rich interactions present in our recordings. Therefore, we used iterative cycles of inductive coding (Berg, 2009) to develop novel coding schemes better suited for the complexity of what students were doing during the modeling activities and to determine whether those actions were productively associated with engaging in modeling to make sense of the biological phenomenon. We analyzed the recordings using Studiocode, version 4 (Vosaic, Lincoln, NE), video analysis software.

To address our first research question about how students use the class time during modeling activities, we developed a coding scheme describing what students were doing and how they interacted during the modeling activities and how this contributed to model development. As an initial step in the coding process, we grouped student activity during modeling into three broad categories: 1) interactions that resulted in adding to, clarifying, or revising the model; 2) talk that was not directly related to model development; and 3) periods of inactivity in which students were neither talking nor drawing (Figure 2). Because students spent the vast



Note: Lower end of each box represents the 25% percentile, upper end of each box represents the 75% percentile, the whiskers represent the minimum and maximum, and the line within each box represents the median.

FIGURE 2. Box plot of percent of time that each group spent on adding to or revising the model ("modeling"), discussing content not directly related to model development ("unrelated"), and neither talking nor drawing ("not talking/drawing") for each activity analyzed.

TABLE 4. Coding scheme for analysis of group verbal interactions

Level	Type	Explanation
Solitary	Silent drawing	The drawer draws something without anyone talking about it.
	Narration	One of the group members is talking out loud without receiving feedback from the group.
Agreement	Suggestion	A nondrawer suggests to the drawer what to do and the drawer does it.
	Verification	The drawer asks the nondrawers if he or she should do something, and at least one nondrawer agrees.
	Confirmation	The drawer confirms with the group that something that he or she did is correct, legible, etc.
Clarification		Two or more group members discuss what or how to do something without disagreement, such as refining an original idea.
Disagreement		Group members discuss what or how to do something without coming to an instant agreement.

majority of the time during the activity interacting in ways that added to the model (category 1), subsequent analysis focused on these interactions. Typically, for interactions resulting in adding to or revising the model, a discussion took place, and then the person doing the drawing added to or revised the model. Iterative cycles of coding revealed four levels of interaction: 1) solitary, 2) agreement, 3) clarification, and 4) disagreement (see Table 4 and Box 1 for explanations and examples of levels). For our coding scheme, we considered an “instance” as each time students 1) added new drawing or writing to the model, 2) revised the drawing or writing, or 3) discussed ideas in the model, potential additions to the model, or revisions to the model but did not follow through by changing the model. We coded each instance with a level of interaction. The interactions in Box 1 illustrate typical group dialogue observed on other recordings as well as a range of common interactions and codes.

It was common for instances to intertwine and overlap as the drawer finished adding a component to the model while the other students began discussing a new idea or as students jumped back and forth between ideas. For example, in Box 1, starting at 17:40, students are trying to figure out how to draw the Ras protein in the second scenario in the model (a normal cell preparing for cell division). At 18:30, before they add Ras to the diagram, the students change topics to clarify the connection between growth factor and signal transduction activity. At

19:00, one group member expresses confusion about the receptor and Ras, two distinct components in the model, and the students again change topics to clarify this confusion. At 19:30, the student making the drawing begins to draw Ras, and the way it is drawn causes disagreement. Because the discussion of Ras (before the clarification instances) is linked with the disagreement, we coded this as one instance of disagreement with two instances of clarification of unrelated ideas occurring in the middle (see Figure 3, starting at ~17:40).

To address our second research question about whether students engage in sense-making during these modeling activities, we developed a coding scheme that identified instances when students verbally worked to make sense of some aspect of the model (Figure 4). Iterative analysis revealed cases when a student asked a question or provided a tentative explanation related to important ideas in the model followed by a second student providing an explanation. We coded these instances in which a question was followed by an explanation as sense-making. In some cases, it was clear that sense-making had occurred, because the original student either restated the explanation in his or her own words or added to the explanation. We coded these instances as evident sense-making. In other instances, the original student simply agreed with the explanation without restating or adding to the explanation. We categorized these instances as potential sense-making, because it was unclear whether the student agreed because he or she understood the

BOX 1. Recording Excerpt with Commentary and Coding

This box contains coded excerpts from a roughly 7-minute portion of a recording of the majors' Ras in cancer activity, with commentary added to describe what students are drawing. Interaction instances are indicated to the left of the excerpt with approximate start times. Sense-making instances are indicated by blue, italic text. Justification instances are in purple, bold text. This portion of the recording is from 17:30 to 25:00 on the timeline in Figure 3.

In this recording, students are developing a model to explain why a mutation in Ras causes uncontrolled cell division and how a new cancer treatment drug can prevent this. Each column in the diagram represents a different scenario in order to model the structure–function relationship of Ras in normal cells; how the mutation that converted Ras from a proto-oncogene to an oncogene influences the structure and function of Ras; and how the new drug influences both normal Ras and mutant Ras:

- Column 1: Normal cell in G_0 phase
- Column 2: Normal cell moving into S phase
- Column 3: Cancer cell with mutant Ras (no drug)
- Column 4: Cancer cell with mutant Ras and new drug

In this recording, students spent roughly 15 minutes determining whether signal transduction is active or inactive and how the cell responds in each of the four scenarios. Student A has drawn the plasma membrane (yellow), receptors (blue squiggly line), and ligands (purple circles) in every column, as shown below. Then the students go back to the first column (normal cell in G_0) and add Ras (green) and GDP (purple). The transcript below begins as students move from drawing a normal cell in G_0 to a normal cell moving into S phase (column 2).

	Column 1 Normal Cell: G ₀	Column 2 Normal Cell: moving into S	Column 3 Cancer Cell: Mutant Ras; no drug	Column 4 Cancer cell: mutant Ras; new drug
Active signal transduction?	inactive	active	active	inactive
Cellular response?	resting; no division	replication of DNA	cell division	cell division blocked
Model				
Required components:	<ul style="list-style-type: none"> Plasma Membrane Receptor Ligand (growth hormone) if appropriate Ras structure (normal or mutant) 			

17:30

17:40 Beginning of disagreement instance #1 (disagreement occurs below)

[The group begins working on the second column (normal cell moving into S phase).]

Student B: *I think that we should, like try this way because like when it's active, it has GDP bonded in the S phase and then converted to GTP.*

[It is unclear what Student B is referring to when she states "this way."]

Student A: *Yeah...yeah, that works, so in this one we would draw it with GTP.*

Student B: *Yeah.*

Student A: *Because that would mean it's active right?*

Student B: *Yeah, there is replication of the DNA.*

18:30 Clarification instance #9

Student C: *Wait.*

Student B: *S phase is not completed here, we just have replication of the DNA.*

Student A: *What were you going to say?*

Student C: *Um, because like, so this one doesn't have like a growth factor since it's inactive, is this one going to have it still even though it's inactive?*

Student A: *Well, the mutant Ras won't be activated but like the normal Ras can still be activated.*

Student C: *Okay.*

Student A: *You know?*

Student C: *Yeah.*

19:00 Clarification instance #10

Student A: *Ummmm, Oh, no, we're doing fine. Plasma membrane, receptor...*

Student B: *Where is receptor? Receptor is Ras?*

Student A: *No, no, no, receptor is this blue thing.*

Student B: *Oh, yeah*

Student A: *Ligand.*

Student B: *Ras structure?*

Student A: *We don't need in this stage. Ras, yeah. And then GDP, and then drug, so okay, we're fine. So this one will be...Ras.*

19:20 End of disagreement instance #1

[Student A draws Ras (green) and GTP (yellow) in column 2, the normal cell preparing for cell division (S phase) as Students A and B discussed at the beginning of this recording clip. She begins labeling the Ras protein.]

Student A: *That looks bad.*

[Student A Removes the labels and begins to re-do the labels.]

Student B: *It's not the same place. You have to move it.*

Student A: *Why?*

Student B: *Why? You have to move it from the cell...we have to move it from the receptor and like go into the cell.*

Look, this is the best way.

Student A suggests: *We're not gonna get...like, the only thing we need to focus on is like this part of it.*

Student B: *Like I was thinking that it should not be connected to the receptor. It must move away from it a little bit.*

[As Student B is talking, Student A erases the part of the receptor (the blue squiggly line) that is touching Ras.]

Student B: *Yeah, this is fine.*

Student A: *"Yeah?"*

Student B: *"Yeah."*

21:00 Start of clarification instance #11

[Student A finishes labeling Ras and GTP in the second column.]

Student B: *Should we draw like this? Like, like how it...*

Student A: *No, I don't think so because it says GTP or GDP.*

Student B: *Oh, okay.*

[Student A pauses drawing.]

Student B: *Which one is the normal cell? This one. In the normal cell, it will have cancer because we have no drug yet. And we have mutation..."*

- 21:30 Solitary instance #8 — Student A: *Alright, I'm just gonna cross these off; we've done them.*
[Student A crosses off the first few components listed in the first column of the worksheet (normal cell not preparing for cell division).]
- 22:30 End of clarification instance #11 — [Student A selects light green to begin the third column (cancer cell with mutant Ras).]
Student A: *So this one would be mutant Ras so I'm gonna do it a different color and make it look fun.*
[Student A begins drawing.]
Student B: *It's active and cell division what happen. How we can ... this is the cancer so how we can address that there's cancer here?*
[No answer is caught on the recording.]
- 22:40 Agreement instance #4 — [Student A finishes drawing and labeling Ras and GTP in the cancer cell (column 3).]
Student A: *Is this looking okay so far?*
Students B: *Yeah.*
Student C: *Yeah.*
- 23:30 Clarification instance #12 — [As Student A draws the first protein in column four, which represents a cancer cell with a new drug that targets mutant Ras.]
Student B: *So if it is getting inactive, so, we have the Ras and GDP"*
Student A: *Yeah.*
Student B: *And how would we draw the drug?*
Student A: *Yeah, I don't know how we would draw a drug. It would be like connected to here, right?*
[As Student A refers to "here," she moves the pointer around the pocket in the Ras protein.]
Student B: *Is that the Ras or GDP?*
Student A: *That's the Ras; this is the GDP.*
[Student A finishes drawing the Ras and starts drawing the GDP.]
Student A: *And then...*
[Student A draws a red circle where her cursor was before.]
- 24:10 Agreement instance #5 — Student A: *Like that?*
Student C: *For the drug?*
Student A: *For the drug.*
- End of clarification instance #12 — [Student A puts a white cross on the red circle.]
Student A: *Drug!*
- 24:40 Agreement instance #6 — Student B: *Looking good.*
Student A: *Is it?*
Student B: *I think.*

25:00

	Column 1 Normal Cell: G ₀	Column 2 Normal Cell: moving into S	Column 3 Cancer Cell: Mutant Ras; no drug	Column 4 Cancer cell: mutant Ras; new drug
Active signal transduction?	inactive	active	active	inactive
Cellular response?	resting; no division	replication of DNA	cell division	cell division blocked
Model				
Required components:				
Plasma Membrane				
Receptor				
Ligand (growth hormone) if appropriate				
Ras structure (normal or				

After the recording clip, the instructor then begins to talk to the class and the students stop working on the model for the moment. They then have an assistant look over the model and receive feedback on changes that they make before moving on to their explanations of the diagrams and submitting the model. See Figure 2 for the completed model.

idea or just did not wish to further pursue the question he or she originally posed. In some cases, a student posed a question aloud and then found an explanation while looking through course materials related to the modeling activity. We also coded these instances as sense-making, even though they did not involve a second student. We coded every instance of sense-making related to important ideas in the model whether or not the sense-making resulted in accurate scientific descriptions of the

phenomenon. We recognize that some sense-making may take other forms and may have occurred silently; therefore, our number of identified instances is likely smaller than the actual number of instances when students were making sense of some aspect of the model. To gain additional insight into student sense-making, we analyzed the sense-making instances from the Ras in cancer activity to determine which aspects of the activity prompted sense-making.

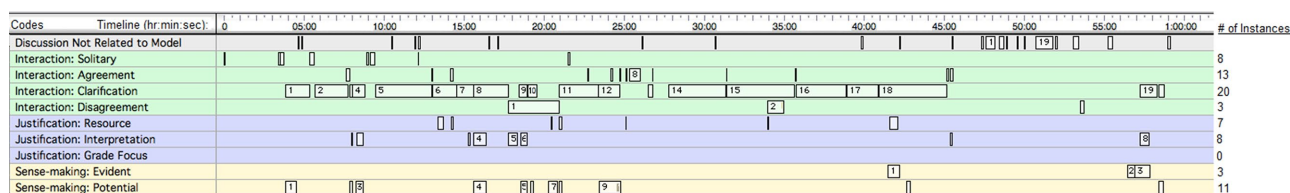


FIGURE 3. Timeline of a cell and molecular biology group creating a model from start to finish for the Ras in cancer activity. Each row is one code. White blocks (sometimes short enough to appear as a line) within each row are instances. The length of each block corresponds to the time it took to complete the instance.

To address our third research question, regarding how students justified their modeling decisions, we developed a coding scheme to identify instances of justification. During the analysis of the videos, we noticed instances of students explaining (i.e., justifying) why something needed to be added, changed, or removed from the model. We identified three types of justification used in the recordings: 1) justification that included the interpretation of information (Berland *et al.*, 2016); 2) justification in which students cited the source of information but did not explain why that information supported the claim; and 3) justification in which students indicated that modeling should either be done in a certain way to improve their grade or not be done because it would not influence the grade (Table 5). When students justified a decision by referring to a source of information, they commonly referenced three types of resources: 1) themselves, 2) the instructor, or 3) course materials. Unlike the student interaction codes, justification was often brief and may or may not have resulted in drawing. We coded an instance of justification every time a student used at least one of the types of justification described above. Furthermore, we coded each instance of justification as only the highest level of justification present, using interpretation as the highest level and grade-based justification as the lowest level.

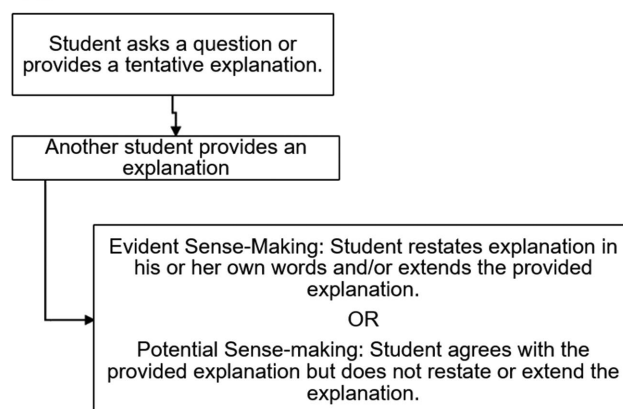
To determine intercoder reliability around each of these coding schemes, A.M.-K.B. and J.R.S. developed a consensus on the coding of two recordings and used a third recording to establish intercoder reliability (Cole *et al.*, 2014). We divided the recording into 30-second intervals and identified intervals in which a new instance of interaction, sense-making, or justification was initiated, based on the type of interaction, sense-making, or justification. We identified intervals with no new instance as “none.” Comparison of coding resulted in Cohen’s kappa (Cohen, 1960) values of 0.75 for interaction codes, 0.76 for justification codes, and 0.69 for sense-making. According to Landis and Kock’s (1977) observer scale, kappa values between 0.61 and 0.80 indicate substantial agreement between coders. After coming to a consensus on the codes for the recording used for reliability testing, A.M.-K.B. then coded the remaining recordings.

ANALYSIS OF INTERVIEWS

Our goal in analyzing the interviews was to provide complementary data to triangulate the findings from our analysis of the recordings. Similar to the analysis of the modeling recordings, we used inductive coding (Berg, 2009) to develop an interview coding scheme. We focused the coding scheme on aspects that informed the research questions. For example,

interview questions regarding how the groups worked together and how they resolved disagreements informed how students contributed to the model. Additionally, interview questions regarding the general purpose of the modeling activities revealed students’ ideas regarding how modeling helped with their sense-making. Finally, during the interviews, students often justified their reasoning regarding which aspects of the model were most important. As a result, we applied our justification coding scheme (from research question 3) to analyze the interview statements. A.M.-K.B. coded the interviews and provided examples of representative quotes that C.V.S. and J.R.S. verified as fitting the coding scheme.

A.



B.

<i>Tentative Explanation</i>	Anna: So the new drug it, like, attaches to the pocket thingy.
<i>Provided Explanation</i>	Bobby: At the very beginning, right?
<i>Restating Explanation</i>	Anna: Yeah, and then it stops it. Bobby: So it stops the activation.

C.

<i>Question</i>	Michaela: ... so this one doesn't have like a growth factor since it's inactive, is this one going to have it still even though it's inactive?
<i>Provided Explanation</i>	Duane: Well, the mutant Ras won't be activated but like the normal Ras can still be activated.
<i>Agreement</i>	Michaela: Okay.

FIGURE 4. Flowchart of the identifying features of sense-making (A) with examples of evident (B) and potential (C) sense-making.

TABLE 5. Coding scheme for analysis of justification

Justification type	Description	Example
Interpretation	A student provides an explanation of the phenomenon.	"But it won't be attached yet because when it's attached that's when it starts."
Resource	A student refers to a source of information such as him- or herself, the instructor, or course materials.	"Look, this is the best way." "Yeah, she said the cascade goes after that." "So this video says...that the construct works by the Chinook growth hormone producing at low levels."
Grade focus	A student suggests something is needed because it will result in a better grade or is unnecessary because it will not impact the grade.	"I don't think that they are really necessarily looking for order, I think they are just looking for shape."

RESULTS

Student Interactions

Our analysis of student interactions as they worked in small groups to develop models found that, while some variation occurred, students spent the majority of time allotted for the modeling activity engaged in discussions that contributed to model development (Figure 2). In typical recordings, instances of the various types of interactions, sense-making, and justification occurred throughout the length of each activity (see Figure 3 for an example of this distribution). In addition to the 75% spent on such work, students spent around 15% of the recording neither talking nor drawing. This time included when students silently searched for information and waited for assistance. Finally, students spent ~10% of the recording discussing something not directly related to the development of the model, although periodically a student silently worked on the model on the tablet during this discussion. An extreme version of this was encountered with one group working on the genetic modification activity. In this recording, the drawer diligently worked on the model and occasionally asked the group brief questions regarding model development, while the entire group (including the drawer) participated in off-task discussion 83% of the activity time.

Because students spent the majority of the time working together to develop their models, we analyzed how they interacted when doing this work. Students in both courses developed their models primarily by clarifying information, agreeing on aspects of the models, or working independently

(Figure 5). The segment in Box 1 and timeline in Figure 3 illustrate typical ways students interacted with one another. This transcript represents most recordings analyzed, in that the types of interactions varied across the recording and occurred in no particular order.

Box 1 is an excerpt from a group working on the Ras in cancer activity. As described in *Description of Modeling Activities Analyzed and Example Modeling Activity*, this modeling activity entailed students developing a model that illustrates the structure–function relationship of Ras and its role in four different scenarios, with each scenario modeled in a separate column.

In this excerpt, students A, B, and C are working with one another to decide how to represent key aspects of the scenarios in the second, third, and fourth columns. The transcript begins when then they are working to determine how Ras functions in a normal cell that is preparing for cell division (second scenario). Students touch on key aspects needed to understand the phenomenon. They discuss physical interactions of signal transduction components, the relationship between growth factors and signal transduction, the impact of normal and mutant Ras on signal transduction, and how the drug fits into the process. They do so using a variety of interaction types that overlap, which was typical of most group work and interactions (see Figure 3).

We note that the transcript illustrates that A, B, and C productively interacted while still missing some opportunities for deeper exploration. For example, the transcript indicates that students could have explored why mutant Ras is always active but did not do so. However, they did focus on important ideas needed to understand the phenomenon—for example, the connection between GDP/GTP binding to RAS and signal transduction activity. Overall, students largely worked together to coconstruct the model.

Interviews confirmed that students commonly interacted by coconstructing models through clarification and that students rarely disagreed ($n = 11$). When asked how the group functioned or contributed to the model, all students either described how they dispersed the work ($n = 7$) or how they worked together ($n = 4$). For instance, “usually we had someone that was designated the writer and then we had two other people that would like

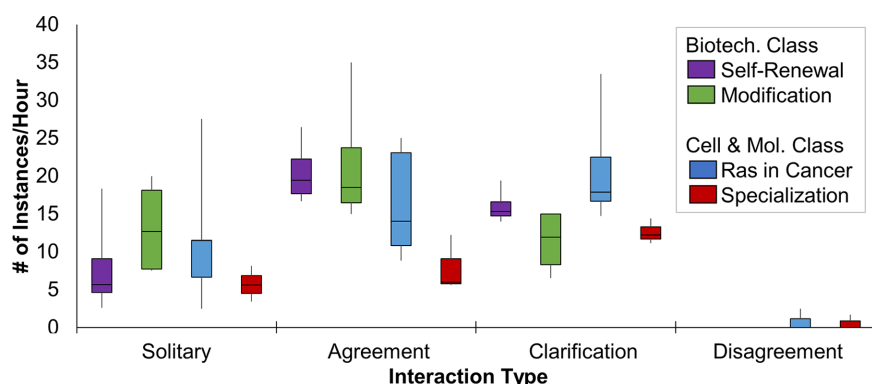


FIGURE 5. Box plot of interaction instances per hour for each activity in each interaction level. An instance includes discussion on the resultant drawing. Lower end of each box represents the 25th percentile; upper end represents the 75th percentile; whiskers represent minimum and maximum; and the horizontal line within each box represents the median.

bounce ideas off as well as the writer.” When asked whether there were any disagreements, most students described that they rarely disagreed ($n = 5$) or if they did disagree, they talked it out ($n = 4$). One student mentioned that her group “just couldn’t agree what [they] needed to do” on one activity, but they did fine on the rest of the activities.

Sense-Making

Our analysis of whether students engaged in sense-making in the modeling activities indicated that students worked to make sense of why certain components were needed in the model and how those components worked together to contribute to the overall function of the system. This is critical because sense-making is a hallmark of scientific modeling practice, and if students are working to make sense of key aspects of a phenomenon, it supports the idea that students are productively participating in this practice.

We found instances of sense-making in all recordings analyzed (Figure 6). These instances were typically distributed throughout the recording (see Figure 3 for a typical distribution). Note that some instances were brief, like the examples in Figure 4, while other instances lasted several minutes. In some cases, as illustrated in the final instance of sense-making in Box 1, students moved between interaction types as they worked to make sense of an idea over an extended period of time.

We further examined the 65 instances of sense-making in the cell and molecular biology modeling activity on Ras in cancer to determine whether the students worked to make sense of key mechanistic aspects of the model (e.g., how the G-protein Ras functions) or whether they focused on trivial aspects of the model (e.g., the activity directions). Analysis indicates that the majority (93%) of sense-making instances focused on four areas that are key to the mechanistic aspects of the model: 1) differences across the four scenarios (63%); 2) how the G-protein Ras functions (12%); 3) how a drug designed to treat cancer works (12%); and 4) how, in general, signal transduction works (6%). These different aspects were discussed throughout the recordings. For instance, the transcript in Box 1 illustrates examples of sense-making around all four main mechanistic aspects. The first sense-making instance discusses differences between the scenarios. The next two

examples focus on general signal transduction function. The last two sense-making instances address G-protein function and drug function, respectively.

Students’ self-reflections during the interviews confirmed the utility of the modeling activities in helping students make sense of phenomena. All students interviewed described that they found developing the models helpful in understanding ideas, and three of the students also stated that the activities aided in seeing connections among ideas. For instance, one student stated, “When I’m drawing this, I have to consciously think about ... how everything is connected.” One of these students described throughout the interview that he preferred “to go down notes and memorize” but still agreed that the models were helpful in understanding the ideas.

Justification

Our analysis of how students justified their modeling decisions indicated that students did so without any specific prompting from instructors or the modeling instructions. During modeling activities, students referred to sources of information, interpreted information, and focused on grade improvement when explaining why something needed to be added, deleted, or revised in the model (Table 5). The majority of justification instances either referred to sources of information or interpreted that information (Figure 7). Students justified their decisions throughout the activities and during all interaction levels (Box 1 and Figure 3).

Analysis of interviews also revealed spontaneous justification. During the interviews, we asked students questions, such as which parts of the model were most important. All 11 students interviewed provided a justification within their explanations without being explicitly prompted to do so. Additionally, when asked whether their groups revised their model during the modeling activity, seven students spontaneously provided justification for their revisions.

DISCUSSION

Engaging students in scientific practices in STEM courses is an important part of improving STEM education (AAAS, 2011; NRC, 2012). Determining whether and how instructional approaches engage students in scientific practices in large-enrollment STEM lecture courses presents significant challenge but needs to be undertaken if we are going to design evidence-based instruction that incorporates science practices. Studies have examined how students participate in the scientific practices of argumentation during clicker questions (Knight *et al.* 2013), use of models during data interpretation in large-enrollment courses (Zagallo *et al.*, 2016), and development of mathematical models in small courses in conjunction with faculty mentors (Svoboda and Passmore, 2010, 2013). Less is known about how small groups of students in large-enrollment courses participate in the scientific practice of model development and which aspects of instructional prompts support productive engagement in this practice.

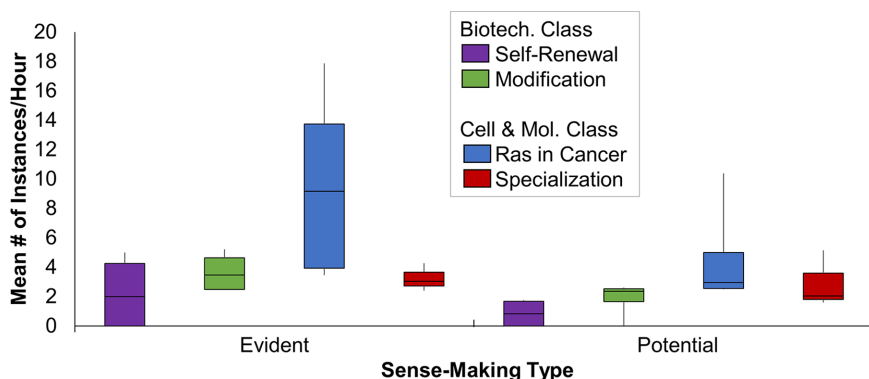


FIGURE 6. Box plot of potential and evident sense-making instances per hour for each activity. Lower end of each box represents the 25th percentile; upper end represents the 75th percentile; whiskers represent minimum and maximum; and the horizontal line within each box represents the median.

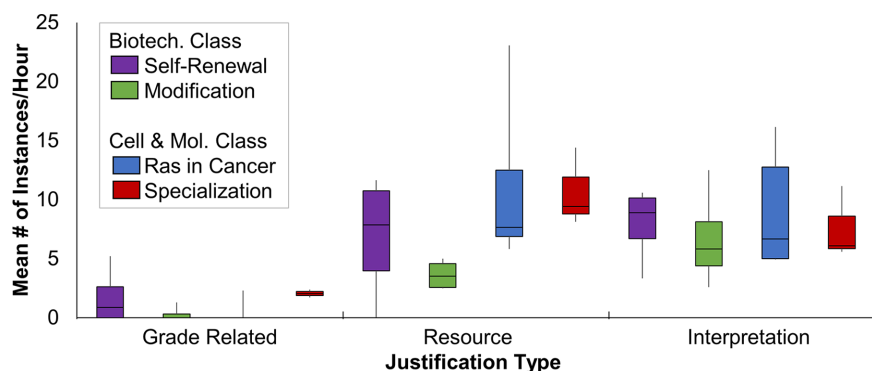


FIGURE 7. Box plot of justification instances per hour for each activity in each justification type. Lower end of each box represents the 25th percentile; upper end represents the 75th percentile; whiskers represent minimum and maximum; and the horizontal line within each box represents the median.

Our work provides evidence that students can engage in the scientific practice of model development in meaningful ways in large-enrollment lecture courses and suggests aspects of activity design that support engagement in scientific modeling.

For our first research question, we initially investigated whether students made productive use of class time during the activities. Our analysis indicates that students spent most of the time during the 30- to 80-minute modeling activities on task developing and evaluating their models (Figure 2). Moreover, they collaborated with one another, worked to make sense of key ideas, and justified their ideas to one another throughout the activities, not just at the start or end of the time during which they were given to complete the activity (Figure 3). These findings concur with those of Young and Talanquer (2013). In their study, groups of students in a chemistry course for nonmajors spent only 6% of the time during the activity in off-task discussions. We found that groups of students typically spent less than 10% of the activity in discussions that did not relate directly to their model (Figure 2). Our study adds to the Young and Talanquer work, as their activities averaged 9 minutes in length, while our activities ranged between 30 and 80 minutes, showing that students can remain engaged in scientific practices for extended periods during class.

As part of this work related to our first research question, we also investigated how group members interacted. We wanted to determine whether one member of the group completed the activity while other members of the group remained passive or off task. We found that most groups spent the majority of time during the activity building on one another's ideas by either confirming ideas put forth by one group member before it was added to the model (agreement) or working together to make sure group members understood an idea before it was added to the model (clarification; Figure 5). This trend suggests that it is possible to design and implement activities in large-enrollment lecture courses during which group members work to coconstruct models and benefit from cooperative learning (Johnson *et al.*, 1998).

While interacting to build on one another's work, the students seldom directly disagreed with one another (Figure 5). A study that recorded groups of students as they worked to interpret data using scientific models in TRIM activities also found that, rather than disagreeing and trying to persuade one another who was

correct, the students typically built on one another's ideas (Zagallo *et al.*, 2016). As pointed out by the authors of that study, this trend differs from the pattern seen during recordings of clicker questions, in which students frequently disagreed on the correct answer and tried to persuade one another who was correct (Knight *et al.*, 2013). A study on peer-led guided-inquiry groups in chemistry found disagreement as part of coconstructed arguments (Kulatunga *et al.*, 2013). These differences are important, as they may suggest that the more open-ended activities in which students develop their own models or interpret data in the context of a current scientific model might offer students opportunities to coconstruct their own understanding rather than choosing from a list of

pre-existing ideas and defending their choices. Alternatively, it could also mean that the activities lend themselves to collaboration—working together to get a job done—perhaps facilitated by collective decision-making on novel representations rather than comparing previously developed representations and ideas and being asked to choose the best one.

For the second research question, we investigated the occurrence and nature of students' sense-making. We found that students engaging in the modeling activities raised questions regarding how scenarios in the model differ, what certain components in the model do, and why a particular aspect of the phenomenon or process works in a particular way. Students then worked together to answer these questions—in other words, to make sense of some aspect of the model they did not understand. This pattern is important, as successfully engaging students in a scientific practice requires that students employ strategies that are hallmarks of that practice. Scientists develop models to aid in making sense of phenomena or processes that they do not yet fully understand. Therefore, successfully engaging students in the scientific practice of model development requires that students work to make sense of a phenomenon or process that the students do not yet fully understand, even if the phenomenon or process is well understood by scientists (Windschitl *et al.*, 2008; Schwarz *et al.*, 2009; Passmore *et al.*, 2013). Our work shows that it is possible to design activities during which students in large-enrollment lecture courses realize they do not understand some aspect of a phenomenon and then work to make sense of the things they do not understand (Figure 6). We recognize that sense-making can come in different forms and is sometimes internal (i.e., not spoken out loud). Therefore, the verbal evidence we collected for sense-making in this study likely underrepresents the actual number of instances of sense-making that took place during each activity.

While analyzing sense-making in the recordings, we observed a connection between the scenarios in the model and the instances of sense-making. The Ras in cancer modeling activity involved four scenarios and resulted in the most instances of sense-making per hour, while the self-renewal modeling activity involved only one scenario and resulted in the least sense-making per hour (Figure 6). Clarification instance #9 in Box 1 illustrates

this, as the student tries to make sense of the role of growth factor in two different scenarios. This connection between scenarios and sense-making instances is supported by post hoc analysis of the sense-making discussions from the Ras in cancer activity recordings, which revealed that 63% of the sense-making instances related to differences between the scenarios. Previous work on peer-led team learning showed that activities requiring teams to contrast different systems or ideas increased the instances of sense-making compared with activities that did not incorporate such contrasts (Young and Talanquer, 2013). Comparison of contrasting cases followed by expert explanations also results in improvements in student learning (Bransford and Schwartz, 1999; Schwartz *et al.*, 2011; Roelle and Berthold, 2015). Taken together, this evidence indicates that drawing attention to important contrasts helps students start making sense of these contrasts in valuable ways. This merits further investigation into the role of contrasting scenarios in developing meaningful modeling activities for use in instruction.

The post hoc analysis of sense-making instances in the Ras in cancer modeling activity indicates an additional area in which sense-making occurred. Twelve percent of the sense-making instances related to G-protein (Ras) function and another 12% of the instances related to how a drug designed to treat cancer works. The phenomenon being modeled, cancer, is a cellular phenomenon, and the components of the model, G-proteins and the drug, are both molecules. Molecules occupy a scalar level below the cellular phenomenon. Differences in scale are important, as mechanistic scientific modeling requires thinking a scalar level below the phenomenon (Wilensky and Resnick, 1999; Van Mil *et al.*, 2016). In their models, students could have simply drawn images of cells reproducing or not reproducing and ignored molecular explanations. Instructing students to include specific components in the model, such as the Ras G-protein and the drug, may have caused the students to struggle with underlying mechanisms of the phenomenon, resulting in instances of sense-making. Further work is needed to understand the relationship between the specific scaffolding provided in the instructions and what students do during model development related to thinking across scales and sense-making. For example, it would be useful to explore how requiring students to include specific components in their models or allowing students to decide which components should be present in their models impacts thinking across scales and sense-making during model development.

Finally, for the third research question, we investigated whether and how students justify their decisions while modeling. An important characteristic of scientific models is that the models are supported by evidence, and students' use of evidence to justify their modeling decisions supports the idea that students are participating in the scientific practice of modeling. During the modeling activities, although not prompted to do so by the instructions, students justified their ideas to one another (Figure 7). These instances of justification occurred without disagreement; in other words, students justified their ideas, even though the instructor did not explicitly direct the students to defend their ideas against other ideas. During interviews in our study, students also spontaneously justified their statements. Other studies have also noted spontaneous justification (Knight *et al.*, 2013; Kulatunga *et al.*, 2013; Zagallo *et al.*, 2016). The observation that students spontaneously justify their decisions

in a variety of active-learning situations, even when not prompted to do so, indicates an opportunity to strengthen scientific practice in these activities. Scientists use experimentation and observation as evidence to justify their models (Nersessian, 1999; Windschitl *et al.*, 2008). Further research is needed to optimize instructions that leverage students' natural tendencies toward justification and guide them to use scientific evidence as part of model development and revision.

The patterns of justification found in our study differ from those found in the other studies that identified justifications based on claims, evidence, and warrants (Toulmin, 1964). This pattern did not emerge in our coding scheme and was not evident in post hoc analysis of student justification. Rather than citing specific evidence and then reasoning about how this evidence justified their claims, students typically cited the textbook or instructor without specific details or they constructed explanations that did not clearly cite any specific evidence. Other studies on introductory STEM students also found lack of the claim, evidence, and reasoning pattern of argumentation in student justifications during clicker questions (James and Willoughby, 2011; Schen, 2012). It is also possible that differences in activity design or instructional cues account for the different patterns of justification observed. For example, the TRIM activities (Zagallo *et al.*, 2016) provided the students with an obvious source of evidence for use in justification: the data from a scientific paper. In addition, instruction that explicitly asks students to provide reasoning when discussing clicker questions increased the use of evidence and reasoning during the discussions (Knight *et al.*, 2013). This suggests that including data as part of modeling activities or explicitly asking students to provide evidence and reasoning to justify their decisions may increase student use of scientific argumentation during model development. More work on the nature of the activity design is needed to understand the impact of these variables on student argumentation.

We analyzed recordings from courses designed for both non-science majors and life science majors and found similar results. As described in the *Methods* section, the general parameters used to design and implement the activities were the same for both courses, but we adjusted the level of mechanistic detail to match the course goals. For example, the genetic modification model from the non-science majors' course and the specialization model from the life science majors' course both focus on the same general phenomenon that changing gene expression alters the proteins produced in a cell, which alters the cell's characteristics. The non-science majors' models only required students to understand that genes are turned on and off, while the life science majors' model required students to understand the role of transcription factors in the process of increasing and decreasing gene expression. Our findings suggest that students from a variety of backgrounds can participate in modeling in large-enrollment lecture courses if the activities are carefully adjusted so the expected mechanistic details are appropriate for the audience.

While enabling students to develop and apply models provides opportunity for sense-making and developing understanding of the underlying mechanisms and meanings of the representations in established models, developing models does not guarantee that students will fully understand important aspects of canonical models and related disciplinary ideas. For example, the excerpt in Box 1 illustrates that, although students

disagreed about how to draw the physical relationship between the receptor and Ras, presumably while referencing a figure from the textbook, they did not pursue the opportunity to further explore any underlying importance of this relationship. It appears that the student who raised the concern was focused on developing a model that matched the illustration from the text rather than exploring the underlying mechanism of why the physical connection between Ras and the receptor might be important. In addition, while the group did work to make sense of the roles of Ras, GTP, and GDP, aspects of the group's final model (Figure 1) do not match canonical explanation of this process. More work is needed to understand productive strategies that both support students in exploring and developing models in ways that mirror scientists' efforts to understand novel phenomena and enable students learn about fundamental established models and ideas.

Several limitations of this research bound our findings. Due to the complexities of collecting recordings of group work that occurs during class and the significant time investment required to analyze the audio and video of an activity that lasts more than an hour, we analyzed a subset of four activities that represented key scientific tasks at different times in the semester. While these do not include every task under every condition, they represent purposeful sampling that is likely to give some indication of the outcomes. In addition, we analyzed these data for a small percentage of students in the class (8%). While the sample groups shared some of the same characteristics as those from the entire class (e.g., GPAs and final course grades) there was an overrepresentation of white females. The impact of these demographic variables on student interactions, sense-making, and justification need to be addressed by further research. Additionally, it is also possible that the act of recording could have caused students in the recorded groups to act differently. However, we observed that students sometimes engaged in behaviors we predicted would be inhibited if they were conscious of being recorded. Examples of these behaviors included using profanity and critiquing the instructor of the course. On the basis of these observations, we suggest that students in the groups we analyzed interacted fairly typically and did not behave in ways aimed at pleasing the instructor simply because they were being recorded.

CONCLUSIONS

National calls for improving science education (e.g., *Vision and Change* [AAAS, 2011]) emphasize the need to focus on disciplinary core ideas using scientific practices during instruction. We engaged students in two large undergraduate biology courses in scientific modeling activities and analyzed what they did during the activities. From the recordings, we found that students worked collaboratively, made sense of key ideas, and justified their ideas while making explanatory, mechanistic models. It is important to understand how students are engaged in the scientific practice of modeling in order to assess whether the activities are meeting the intended scientific practice objectives of developing students' modeling abilities. These results indicate that the activities met our scientific practices learning objective and add to our knowledge about how to productively engage students in scientific practices within large undergraduate courses. For future work, we recommend building on this analysis to understand the role of the activity design (e.g., nature of scenarios, scaffolding, inclusion of data) and

instruction on students' sense-making. Further, we recommend additional analysis regarding how particular groups of students (students from underrepresented groups or those from varied linguistic backgrounds) interact and respond to engagement in practices to determine productive ways to engage all students. Understanding productive scientific engagement in practices will enable the community to advance science learning for all.

ACKNOWLEDGMENTS

We thank the CREATE for STEM Institute at Michigan State University for their financial support (LPF-CMP 2 Innovation Grant). For revision recommendations on earlier versions of this article, we thank members of the Michigan State University Geocognition Lab (Dr. Julie Libarkin, Patricia Jaimes, Caitlin Kirby, and Amanda Lorenz), the AACR (Automated Analysis of Constructed Response) Research Group (Jennifer Kaplan and Alex Lyford from University of Georgia and Rosa Moscarella from Michigan State University), and the reviewers.

REFERENCES

- American Association for the Advancement of Science (AAAS). (2011). *Vision and change in undergraduate biology education: A call to action*. Washington, DC.
- Beichner, R. J., Sual, J. M., Abbot, D. S., Morse, J. J., Deardorff, D. L., Allain, R. J., ... Risley, J. S. (2007). The student-centered activities for large enrollment undergraduate programs (SCALE-UP) project. In Redish, E., & Cooney, P. (Eds.), *Reviews in PER. Volume 1: Research-based reform of university physics*. College Park, MD: American Association of Physics Teachers. Retrieved September 7, 2014, from www.percentral.com/PER/per_reviews/media/volume1/SCALE-UP-2007.pdf
- Berg, B. L. (2009). *Qualitative research methods for the social sciences* (7th ed.). Boston, MA: Allyn & Bacon, Pearson Education.
- Berland, L. K., Schwarz, C. V., Krist, C., Kenyon, L., Lo, A. B., & Reiser, B. J. (2016). Epistemologies in practice: Making scientific practices meaningful for students. *Journal of Research in Science Teaching*, 53(7), 1082–1112. doi: 10.1002/tea.21257
- Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple implications. *Review of Research in Education*, 24, 61–100. doi.org/10.2307/1167267
- Brewe, E., Sawtelle, V., Kramer, L. H., O'Brien, G. E., Rodriguez, I., & Pamelá, P. (2010). Toward equity through participation in modeling instruction in introductory university physics. *Physical Review Special Topics—Physics Education Research*, 6(1). doi: 10.1103/PhysRevSTPER.6.010106
- Cohen, J. A. (1960). Coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37–46.
- Cole, R. S., Becker, N., & Stanford, C. (2014). Discourse analysis as a tool to examine teaching and learning in the classroom. In Bunce, D. M., & Cole, R. S. (Eds.), *Tools of chemistry education research* (pp. 61–81). Washington, DC: American Chemical Society.
- Cooper, M. M. (2015). Why ask why? *Journal of Chemical Education*, 92(8), 1273–1279. doi: 10.1021/acs.jchemed.5b00203
- Cooper, M. M. (2016). It is time to say what we mean. *Journal of Chemical Education*, 93(5), 799–800. doi: 10.1021/acs.jchemed.6b00227
- Cooper, M. M., Caballero, M. D., Ebert-May, D., Fata-Hartley, C. L., Jardeleza, S. E., Krajcik, J. S., ... Underwood, S. M. (2015). Challenge faculty to transform STEM learning. *Science*, 350(6258), 281–282. doi: 10.1126/science.aab0933
- Dauer, J. T., Momsen, J. L., Speth, E. B., Makohon-Moore, S. C., & Long, T. M. (2013). Analyzing change in students' gene-to-evolution models in college-level introductory biology. *Journal of Research in Science Teaching*, 50(6), 639–659. doi: 10.1002/tea.21094
- Harrison, A. G., & Treagust, D. F. (2010). A typology of school science models. *International Journal of Science Education*, 22(9), 1011–1026. doi: 10.1080/095006900416884

- James, M. C., & Willoughby, S. (2011). Listening to student conversations during clicker questions: What you have not heard might surprise you! *American Journal of Physics*, 79(1), 123–132. doi: 10.1119/1.3488097
- Johnson, D. W., Johnson, R. T., & Smith, K. A. (1991). *Cooperative Learning: Increasing College Faculty Instructional Productivity* (ASHE–ERIC Higher Education Report No. 4). Washington, DC: George Washington University.
- Johnson, D. W., Johnson, R. T., & Smith, K. A. (1998). Cooperative learning returns to college: What evidence is there that it works? *Change: The Magazine of Higher Learning*, 30, 26–35.
- Knight, J. K., Wise, S. B., & Southard, K. M. (2013). Understanding clicker discussions: Student reasoning and the impact of instructional cues. *CBE—Life Sciences Education*, 12, 645–654. doi: 10.1187/cbe.13-05-0090
- Kulatunga, U., Moog, R. S., & Lewis, J. E. (2013). Argumentation and participation patterns in general chemistry peer-led sessions. *Journal of Research in Science Teaching*, 50, 1207–1231. doi: 10.1002/tea.21107
- Landis, J. R., & Kock, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–174.
- Luckie, D., Harrison, S. H., & Ebert-May, D. (2011). Model-based reasoning: Using visual tools to reveal student learning. *Advances in Physiology Education*, 35(1), 59–67. doi: 10.1152/advan.00016.2010
- Morgan, M. S., & Morrison, M. (1999). *Models as mediators*. Cambridge, UK: Cambridge University Press.
- National Research Council (NRC). (2012). *A framework for K–12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: National Academies Press.
- Nersessian, N. (1999). Model-based reasoning in conceptual change. In Magnani, L., Nersessian, N., & Thagard, P. (Eds.), *Model-based reasoning in scientific discovery* (pp. 5–22). New York: Kluwer Academic/Plenum.
- Passmore, C., Gouvea, J. S., & Giere, R. (2013). Models in science and in learning science: Focusing scientific practice on sensemaking. In Matthews, M. R. (Ed.), *International handbook of research in history, philosophy and science teaching* (pp. 1171–1202). New York: Springer.
- Passmore, C., Schwarz, C., & Mankowski, J. (2017). Developing and using models. In Schwarz, C., Passmore, C. & Reiser, B. (Eds.). *Helping students make sense of the world using next generation science and engineering practices*. Arlington, VA: NSTA Press.
- Quillin, K., & Thomas, S. (2015). Drawing-to-learn: A framework for using drawings to promote model-based reasoning in biology. *CBE—Life Sciences Education*, 14, es2. doi: 10.1187/cbe.14-08-0128
- Roelle, J., & Berthold, K. (2015). Effects of comparing contrasting cases on learning from subsequent explanations. *Cognition and Instruction*, 33(3), 199–225. doi: 10.1080/07370008.2015.1063636
- Schen, M. (2012). Assessment of argumentation skills through individual written instruments and lab reports in introductory biology. Paper presented at the Annual Meeting of the National Association for Research in Science Teaching (Indianapolis, IN).
- Schwartz, D. L., Chase, C. C., Oppezzo, M. A., & Chin, D. B. (2011). Practicing versus inventing with contrasting cases: The effects of telling first on learning and transfer. *Journal of Educational Psychology*, 103(4), 759–775. doi: 10.1037/a0025140
- Schwarz, C. V., Reiser, B. J., Davis, E. A., Kenyon, L., Achér, A., Fortus, D., ... Krajcik, J. (2009). Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners. *Journal of Research in Science Teaching*, 46(6), 632–654. doi: 10.1002/tea.20311
- Schwarz, C. V., & White, B. Y. (2005). Metamodeling knowledge: Developing students' understanding of scientific modeling. *Cognition and Instruction*, 23(2), 165–205. doi: 10.1207/s1532690xci2302_1
- Stoltzfus, J. R., & Libarkin, J. (2016). Does the room matter? Active learning in traditional and enhanced lecture spaces. *CBE—Life Sciences Education*, 15, ar68.
- Svoboda, J., & Passmore, C. (2010). Evaluating a modeling curriculum by using heuristics for productive disciplinary engagement. *Cell Biology Education*, 9(3), 266–276. doi: 10.1187/cbe.10-03-0037
- Svoboda, J., & Passmore, C. (2013). The strategies of modeling in biology education. *Science & Education*, 22(1), 119–142. doi: 10.1007/s11191-011-9425-5
- Toulmin, S. E. (1964). *The uses of argumentation*. Cambridge, UK: Cambridge University Press.
- University of California–San Francisco. (2013, November 20). Drug strategy blocks leading driver of cancer. *ScienceDaily*. Retrieved October 14, 2014, from www.sciencedaily.com/releases/2013/11/131120133735.htm
- Van Mil, M. H. W., Postma, P. A., Boerwinkel, D. J., Klaassen, K., & Waarlo, A. J. (2016). Molecular mechanistic reasoning: Toward bridging the gap between the molecular and cellular levels in life science education. *Science Education*, 100(3), 517–585.
- White, B. Y., & Frederiksen, J. R. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and Instruction*, 16(1), 3–118.
- Wilensky, U., & Resnick, M. (1999). Thinking in levels: A dynamic systems approach to making sense of the world. *Journal of Science Education and Technology*, 8(1), 3–19. doi: 10.1023/A:1009421303064
- Williams, L. C., Underwood, S. M., Klymkowsky, M. W., & Cooper, M. M. (2015). Are noncovalent interactions an Achilles heel in chemistry education? A comparison of instructional approaches. *Journal of Chemical Education*, doi: 10.1021/acs.jchemed.5b00619
- Windschitl, M., Thompson, J., & Braaten, M. (2008). Beyond the scientific method: Model-based inquiry as a new paradigm of preference for school science investigations. *Science Education*, 92(5), 941–967. doi: 10.1002/sce.20259
- Young, K. K., & Talanquer, V. (2013). Effect of different types of small-group activities on students' conversations. *Journal of Chemical Education*, 90, 1123–1129. doi: 10.1021/ed400049a
- Zagallo, P., Meddleton, S., & Bolger, M. S. (2016). Teaching real data interpretation with models (TRIM): Analysis of student dialogue in a large-enrollment cell and developmental biology course. *CBE—Life Sciences Education*, 15, ar17. doi: 10.1187/cbe.15-11-0239