Subject/Problem

Science courses often focus on memorization and test preparation instead of deeper learning. Recent calls for reform in science education at both the K-12 and undergraduate level advocate moving from courses that focus largely on memorizing the breadth of disciplinary content to courses that focus on science practices and core disciplinary ideas (AAAS, 2011; Korber, 2014; NRC, 2012; Nelson Laird et al., 2008). One approach to shift courses from memorization of facts toward development of scientific competencies is to include the development and revision of scientific models as part of instruction. Scientific modeling encompasses a diverse set of practices at the heart of the scientific endeavor (Passmore et al., 2013). Models can be tested, are revised based on new evidence, have explanatory power because they are based on underlying structure and causation, are often based on processes that cannot be directly observed, and are springboards for generating new scientific knowledge, in part, by generating testable predictions (Passmore et al., 2013; Schwarz et al., 2009; Windschitl et al., 2008). Therefore, the process of constructing scientific models and reasoning with these scientific models serves as an ideal platform to help students learn both disciplinary content and disciplinary practices. The incorporation of modeling in the classroom also fulfills the call in *Vision and Change* for students to apply the process of science as part of instruction and to develop skills in modeling and simulation (AAAS, 2011). The practice of modeling in K-12 education has had positive impacts on student outcomes (Jackson et al., 2008; Schwarz, 2002; Seel, 2003; Svoboda and Passmore, 2013). At the university level, diagrammatic modeling tools like schematic representations, concept maps and Structure/Behavior/Function (SBF) models are being developed and used in undergraduate biology education (Dauer et al., 2013; Luckie et al., 2011).

Scientific practice also involves social construction of knowledge. Scientists must gather empirical evidence and use it to defend their scientific models or refute current scientific models. This is done in the context of the scientific community. In addition, the benefits of cooperative learning is well established (Johnson, 1991; Vygotsky, 1978), and it has been suggested that the use of cooperative learning enhances the benefits of model-based instruction (Coll et al., 2005). Based on best practices for productive cooperative learning, we developed instructional environments in which groups of students use evidence about phenomena to construct and evaluate scientific models of those phenomena. Then students use these scientific models to reason about phenomena with the overarching goal of engaging students in sense-making. During the modeling activities, we investigated group verbal interactions and sense-making using a ground theory approach. We also interviewed students to verify our findings.

Procedure

Participants and Activities

This study took place in a large mid-western research university. We introduced modeling activities to the curriculum of two large-enrollment undergraduate introductory biology courses: a non-majors course and a biology majors course (see Table 1 for additional variation between the two courses). One of the objectives for both courses was that students will be able to develop and revise scientific models to create explanations and make predictions. To meet this objective, students in groups of three completed periodic modeling activities. Before each activity, students read about a biological phenomenon from their textbook and science articles written for the general public (e.g., from *ScienceDaily*). In class, we give students a description of the phenomenon to model and key elements of the model to incorporate. Students work in groups of three to create the diagrammatic model and a cause-and-effect explanation (see Figure

1 for example). In the majors course, students created models on paper but were given the option to create them on electronic devices while in the non-majors course, students created models on provided Surface Pro tablets.

Data Collection

We collected data from two different sections of the non-majors course and one section of the majors course. Each course/section was taught by a different instructor with different teaching assistants; the second author taught one section of the non-majors course.

If all group members from a group signed individual consent forms, then the group was eligible to participate in the study. We provided an electronic tablet to every participating group whether or not it was the norm in the course to complete the activities on tablets. Then we instructed all participating groups to record their discussion and concurrent diagrammatic modeling during the entire duration of the activity via Camtasia Relay. We collected recordings from several modeling activities from both the biology majors course and the non-majors course.

Exploratory Phase

Our exploratory research question asked "what are students doing while engaged in creating the model?" We created descriptive portraits of six recordings from the majors course and 12 recordings from the non-majors course. These descriptive portraits provided a timeline of what students were saying and what they were doing during the activity. We then qualitatively analyzed the summaries using a theory-based and inductive approach to identify key aspects of student modeling. Our original intention was to examine how often students were on task, but we found that students were on task most of the time. Instead, three, far richer, key questions emerged from the descriptive portraits:

1) How do students interact (e.g., agree, clarify, disagree) when deciding what to include in their models?
2) How do students justify their decisions?
3) How do students engage in sense-making during these modeling activities?

After developing the three research questions, we re-examined the descriptive portraits to create initial coding schemes for researching each question. The first two authors, with feedback from the third author, cycled several times through a process of coding two recordings together and then independently. Each cycle resulted in editing the coding schemes. Once all authors were satisfied with the coding schemes and the coding of the two recordings, the first two authors independently coded a third recording and compared results to establish inter-coder reliability (details for this particular recording are described in the "Data Analysis" section). We established inter-coder reliability for a recording to ensure that we developed clear definitions of each code (Cole et al., 2014). After developing a consensus for the third recording, the first author then coded the remaining recordings.

Coding Schemes and Data Analysis

How do students interact (e.g., agree, clarify, disagree) when deciding what to include in their models? We discovered four levels of interactions (Table 2; see Figure 2 for examples of each level). A low level of interaction occurs when a student works primarily independently while a higher level of interaction has students adding on to each other's ideas. Often discourse would take place and then the drawer would draw (the exception being "silent drawing" where only drawing took place). For our coding scheme, we considered discourse and the resulting drawing as an episode and labeled the episode with one of the four levels of interactions. We established inter-coder reliability of one recording: 74% (31/42) of episodes identified by the two coders overlapped and were coded with the same level of interaction. Five of the 31 (16%) episodes were identified as one episode by one coder and two episodes by the other coder.

In addition to the four levels of interaction, we also coded episodes of discussion that were not directly related to the creation of the model. These episodes included: 1) talking about off-task topics, 2) other course activities, 3) how to use the electronic tablet, 4) if they should ask for assistance, 5) that they are searching through resources, and 6) other students' models (not discussing how another model informs their own model). With these codes included, 71% of the episodes identified by the two coders overlapped and were coded with the same code.

How do students justify their decisions? Although students were never explicitly asked to do so, we discovered evidence of student justification. We identified two types of justification: source of information and interpretation of information (Berland et al., 2015). Students commonly reference three types of resources: 1) him or herself (e.g., "look, this is the best way"), 2) teacher or lecture (e.g., "yeah, she said the cascade goes after that"), and 3) course materials (e.g., "so this video says…that the construct works by the Chinook growth hormone producing at low levels"). When interpreting information, students most often provided an explanation. For instance, "but it won't be attached yet because when it's attached that's when it starts."

We coded each instance of justification as one type of justification. If a student provided the resource and an interpretation, we coded it as interpretation. For inter-coder reliability, both coders identified 10 instances of justification, which they coded as the same type of justification for nine of them. One coder found an additional instance that the other coder missed. This discrepancy led the coders to modify the methods slightly so that every recording would be coded again for instances of justification one week after completing the initial coding.

How do students engage in sense-making during these modeling activities? We recognize that sense-making can come in different forms and is likely often internal (i.e., not spoken out loud). Therefore, the verbal evidence we collected for sense-making likely under represents the actual number of instances. For our purposes, we defined sense-making as students are working to understand:

1. What a component in the model is or does
2. How components in the model work together to create a function
3. How the components interact within a system.

While making sense of these concepts, student may: 1) use logic and previous knowledge, 2) look for new information, 3) review material, and/or 4) ask for assistance.

The evidence that we found for sense-making began with a student asking a question or providing a tentative explanation. Then the student was provided with an explanation (from another student, instructor, textbook, etc). The original student then agreed with the provided explanation, restated the explanation in his or her own words, and/or added to the explanation. If the student simply agreed with the explanation, we considered it potential sense-making since it is unclear if the student was in agreement because he or she understood the concept or did not wish to further pursue the question he or she originally had. For instance:

> Aaron: So we just divide the Ras, GDP a bunch, is that right? *(Question)*
> Teaching Assistant: Well, you'll have Ras is bound to GTP because it's active, so you add that part. Think about whether or not it's supposed to be active. *(Explanation)*
> Aaron: Oh, okay. *(Agreement)*

In the above example, Aaron asks a question about a concept, the teaching assistant provides an explanation, and Aaron agrees with the teaching assistant's explanation. Other forms of sense-making were considered evident sense-making. For instance:

> Anna: So the new drug it, like, attaches to the pocket thingy.
> Bobby: At the very beginning, right? *(Tentative explanation)*
> Anna: Yeah, and then it stops it. *(Explanation)*
> Bobby: So it stops the activation. (*Restating explanation)*

In the above example, Bobby asks a question about the concept that Anna first brought up and Anna provides an explanation. Bobby then restates Anna's explanation, which we established as an indicator of evident sense-making.

We coded every instance of sense-making in the recordings. Some instances were brief, like the examples above, and others lasted a few minutes. For inter-coder reliability of one recording, both coders identified 12 instances of sense-making and coded them with the same form of sense-making for all but one of the instances. Five additional instances were found by one coder but not the other and nearly all of these instances were confirmed by the other coder while reaching consensus. Therefore, similar to coding for justification, the coders agreed that each recording would be coded again for sense-making after one week of the initial coding.

For this study, we analyzed recordings from two activities of the non-majors course (four recordings) and one activity from the majors course (five recordings). We intended to utilize our coding scheme for a variety of scenarios (e.g., various times within a semester), so from the majors course we used the third activity from five total activities (see Figure 1 for example from this activity). Then we selected the second and ninth activities (out of 10 activities total) in the non-majors course. Students stay in the same groups within the non-majors course, and we selected a recording from the ninth activity that was done by a group that we analyzed from the second activity.

Interviews

After developing the research questions, we interviewed a subset of the participating students (five biology majors and six non-biology majors). We interviewed the biology majors a few weeks after they completed the course, and we interviewed the non-biology majors around the same time that students completed the last modeling activity. The first author conducted semi-structured interviews one-on-one with students outside of class time. Interviews lasted about 30 minutes. We used the interviews to validate our findings from the recordings.

Findings

Below, each research question is addressed by providing the results for the recordings and interviews. Since we obtained the data from volunteers, they may not be representative of all of the students in the classes. Nonetheless, we believe that this study provides insightful information about the process that students are undergoing while developing models. Additionally, the act of recording can cause students to act differently, but we observed that students often seem to forget that they are being recorded (e.g., swearing, critiquing the instructor, etc.).

How do students interact when deciding what to include in their models? We discovered that students rarely disagree and split their time fairly evenly between the other three levels (solitary, agreement, and clarification), although variation existed across the groups (Figure 3). During interviews, when asked how the group functioned or contributed to the model, all students either described how the work was dispersed (*n* = 7) or how they worked together (*n* = 4). After that question, we asked if there were any disagreements, and 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 and another student experienced "some tension between [her] and [her] female teammate."

How do students justify their decisions? In order to justify their decisions, students either referred to sources of information or interpreted information. Each group used both types of justification but varied in how often they used them (Figure 4). Three of the groups interpreted information more often than referred to a resource for justification. During the interviews, when asked about which parts of the model were most important, all 11 interviewed students provided an interpretation as their justification. Seven of the interviewed students were in a group that revised their model at least once during a recorded activity. We asked these students why the revision took place. Similar to the activity recordings, students varied in how they justified the revision. Four students referred to a source of information, one student provided an interpretation and two students used both types of justification.

How do students engage in sense-making during these modeling activities? Similar to justification, groups varied in how often they exhibited sense-making but most groups periodically showed potential and evident signs of sense-making (Figure 5; note that although the majors groups experienced more sense-making, on average, the timespan of the activities was twice as long as the non-majors' activities). Two groups were the exception: we did not find any instances of sense-making in one group, and another group had a fair number of potential sense-making episodes but no evident sense-making. During the interviews, all students described that the models were helpful in understanding concepts, and three of the students also stated that the activities helped in seeing connections among concepts. We asked seven of the students if it would have been just as helpful to be given a completed model and although one student could not decide, the rest of the students preferred to create the model. A few of these students had earlier in the interview described that they preferred lecture over the modeling activities. Therefore, it is possible that some students may have felt under pressure during the interview to state that the models were helpful.

Contribution

Recent national calls for improving science education (AAAS, 2011; Korber, 2014; NRC, 2012; Nelson Laird et al., 2008) emphasize the need to focus on core disciplinary concepts and incorporate scientific practices during instruction. This can present a challenge at research universities where introductory science courses typically have large enrollments (Handelsman et al., 2004), and research is needed to evaluate whether activities used in these courses meet the goals of deeper learning. We implemented a series of modeling activities in large enrollment introductory biology courses and found evidence that students were engaged in the activity and worked to make sense of the phenomena, sometimes as often as every 10 minutes. However, because students tend to agree with one another without argument, the observed patterns may suggest engagement in everyday discussion norms that do not necessarily foster explicit scientific thinking. Therefore, these group modeling activities may foster understanding of core concepts, but need further modification, such as asking students to develop written explanations of the phenomenon before working on the model in a group. This might lead to additional evaluation and comparison of ideas and to improve scientific thinking. In conclusion, we found that these modeling activities in large enrollment classrooms helped foster sense-making of biological phenomena.

Continuing Research Plans

We have developed coding themes to analyze group verbal interactions, justification, and sense-making by using a ground theory approach and used the themes to conduct a preliminary analysis on the descriptive portraits. In addition to coding more recordings, we are also interested in examining the relationship between the three concepts (interaction, justification, and sense-making). Moreover, we will be investigating how these three concepts relate to the content of these activities.

References

AAAS. (2011). *Vision and change in undergraduate biology education: A call to action*. C. A. Brewer and D. Smith. Washington DC: American Association for the Advancement of Science.

Cole RS, Becker N, Stanford C. (2014). Discourse analysis as a tool to examine teaching and learning in the classroom. In Tools of Chemistry Education Research (Eds. Bunce DM, Cole RS), Washington, DC: American Chemical Society.

Coll, R. K., France, B., et al. (2005). The role of models and analogies in science education: Implications from research. *International Journal of Science Education 27*(2), 183-198.

Cooper, M. M., Underwood, S. M., et al. (2012). Development and assessment of a molecular structure and properties learning progression. *Journal of Chemical Education 89*(11), 1351-1357.

Dauer, J. T., Momsen, J. L., et al. (2013). Analyzing change in students' gene-to-evolution models in college-level introductory biology. *Journal of Research in Science Teaching 50*(6), 639-659.

Ebert-May, D., Brewer, C., et al. (1997). Innovation in large lectures-teaching for active learning. *Bioscience* 47(9), 601-607.

Handelsman, J., Ebert-May, D., et al. (2004). Scientific teaching. *Science 304*(5670), 521-522.

Harrison, A. G., & Treagust, D. F. (2000). A typology of school science models. *International Journal of Science Education 22*(9), 1011-1026.

Jackson, J., Dukerich, L., et al. (2008). Modeling instruction: An effective model for science education. *Science Educator 17*(1), 10-17.

Johnson, D. W., Johnson, R. T., & Smith, K. A. (1991). Cooperative learning: Increasing college faculty instructional productivity. ASHE-ERIC Higher Education Report No. 4. *ASHE-ERIC Higher Education Reports*, Washington, DC: George Washington University.

Kober, L. (2014). *Reaching students: What research says about effective instruction in undergraduate science and engineering*. Washington, DC: National Academies Press.

Luckie, D., Harrison, S. H., et al. (2011). Model-based reasoning: Using visual tools to reveal student learning. *Advances in Physiology Education 35*(1), 59-67.

Nelson Laird, T. F., Shoup, R., et al. (2008). The effects of discipline on deep approaches to student learning and college outcomes. *Research in Higher Education 49*(6), 469-494.

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

Passmore, C., Svoboda, J., et al. (2013). Models in science and in learning science: Focusing scientific practice on sensemaking. In M. R. Matthews (Ed.). *International handbook of* *research in history, philosophy and science teaching* (pp. 1171-1202). New York, NY: Springer.

Russ, R. S., Scherr, R. E., et al. (2008). Recognizing mechanistic reasoning in student scientific inquiry: A framework for discourse analysis developed from philosophy of science. *Science Education 92,* 499-525.

Schwarz, C. V. (2002). *Using model-centered science instruction to foster students' epistemologies in learning with models.* Paper presented at the annual meeting for the American Educational Research Association. New Orleans, LA.

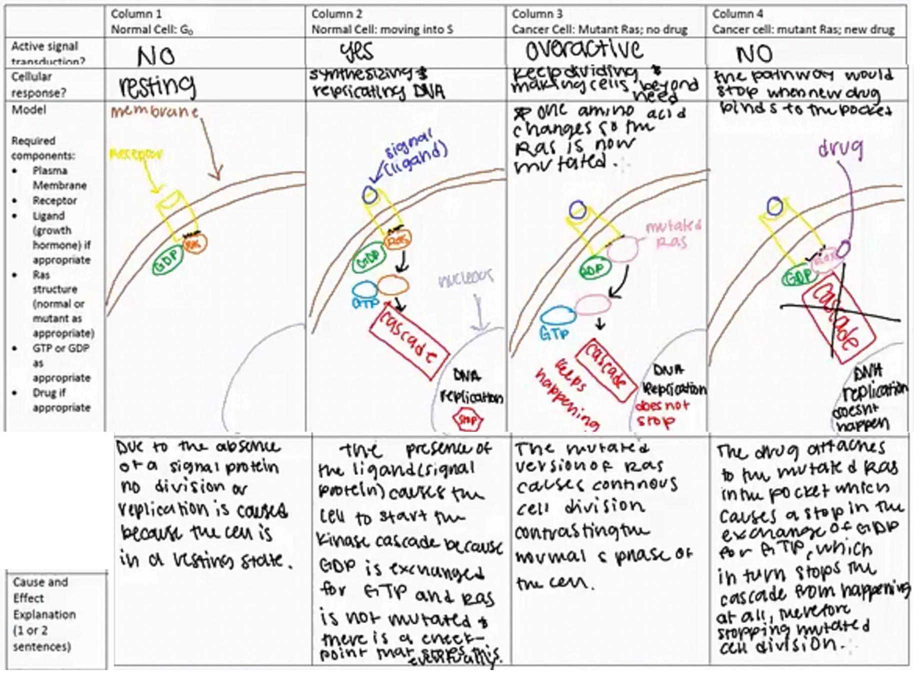
Schwarz, C. V., Reiser, B. J., et al. (2009). Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners. *Journal of Research in Science Teaching 46*(6), 632-654.

Seel, N. M. (2003). Model-centered learning and instruction. *Technology, Instruction, Cognition and Learning 1*(1), 59-85.

Svoboda, J. & Passmore, C. (2013). The strategies of modeling in biology education. *Science & Education 22*(1), 119-142.

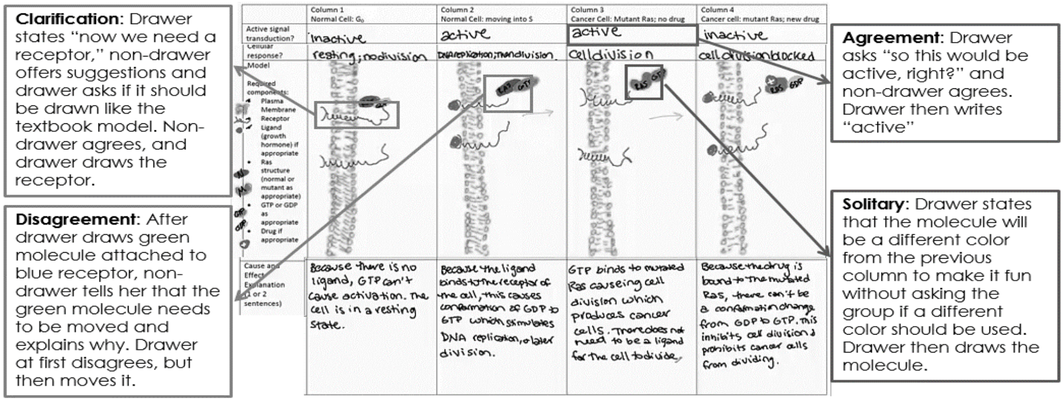
Vygotsky, L. S. (1978). *Mind in society* (M. Cole, V. John-Steiner, S. Scriberner, & E. Souberman, E., Eds. & Trans.). Cambridge, MA: Harvard University Press.

Windschitl, M., Thompson, J., et al. (2008). Beyond the scientific method: Model-based inquiry as a new paradigm of preference for school science investigations. *Science Education 92*(5): 941-967.



	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?	NO	yes	overactive	NO
Cellular response?	resting	synthesizing & replicating DNA	keep dividing & making cells beyond need	the pathway would stop when new drug binds to the pocket
Model Required components: • Plasma Membrane • Receptor • Ligand (growth hormone) if appropriate • Ras structure (normal or mutant as appropriate) • GTP or GDP as appropriate • Drug if appropriate			* one amino acid changes so the Ras is now mutated.	
Cause and Effect Explanation (1 or 2 sentences)	Due to the absence of a signal protein no division or replication is caused because the cell is in a resting state.	The presence of the ligand(signal protein) causes the cell to start the kinase cascade because GDP is exchanged for GTP and Ras is not mutated & there is a checkpoint that stops eventually this.	The mutated version of Ras causes continous cell division contrasting the normal S phase of the cell.	The drug attaches to the mutated Ras in the pocket which causes a stop in the exchange of GDP for GTP, which in turn stops the cascade from happening at all, therefore stopping mutated cell division.

Figure 1. Example of a student's model from the majors course. For this modeling activity, students were instructed to create a model that illustrates the structure-function relationship of Ras (a protein that is involved in a signaling pathway that causes cells to divide when growth factors are present) in normal cells, how the mutation that converted Ras from a proto-oncogene to an oncogene impacts the structure and function of Ras, and how a new drug impacts both normal Ras and mutant Ras.



Clarification: Drawer states "now we need a receptor," non-drawer offers suggestions and drawer asks if it should be drawn like the textbook model. Non-drawer agrees, and drawer draws the receptor.

Disagreement: After drawer draws green molecule attached to blue receptor, non-drawer tells her that the green molecule needs to be moved and explains why. Drawer at first disagrees, but then moves it.

	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	DNA replication; then division	Cell division	cell division blocked
Model Required components: • Plasma Membrane Receptor Ligand (growth hormone) if appropriate • Ras structure (normal or mutant as appropriate) • GTP or GDP as appropriate • Drug if appropriate				
Cause and Effect Explanation (1 or 2 sentences)	Because there is no ligand, GTP can't cause activation. The cell is in a resting state.	Because the ligand binds to the receptor of the cell, this causes conformation of GDP to GTP which stimulates DNA replication, & later division.	GTP binds to mutated Ras causeing cell division which produces cancer cells. There does not need to be a ligand for the cell to divide.	Because the drug is bound to the mutated Ras, there can't be a conformation change from GDP to GTP. This inhibits cell division & prohibits cancer cells from dividing.

Agreement: Drawer asks "so this would be active, right?" and non-drawer agrees. Drawer then writes "active"

Solitary: Drawer states that the molecule will be a different color from the previous column to make it fun without asking the group if a different color should be used. Drawer then draws the molecule.

Figure 2. A student group's final "Ras in Cancer" model with an example episode of each interaction level. An episode is discourse and resulting drawing.

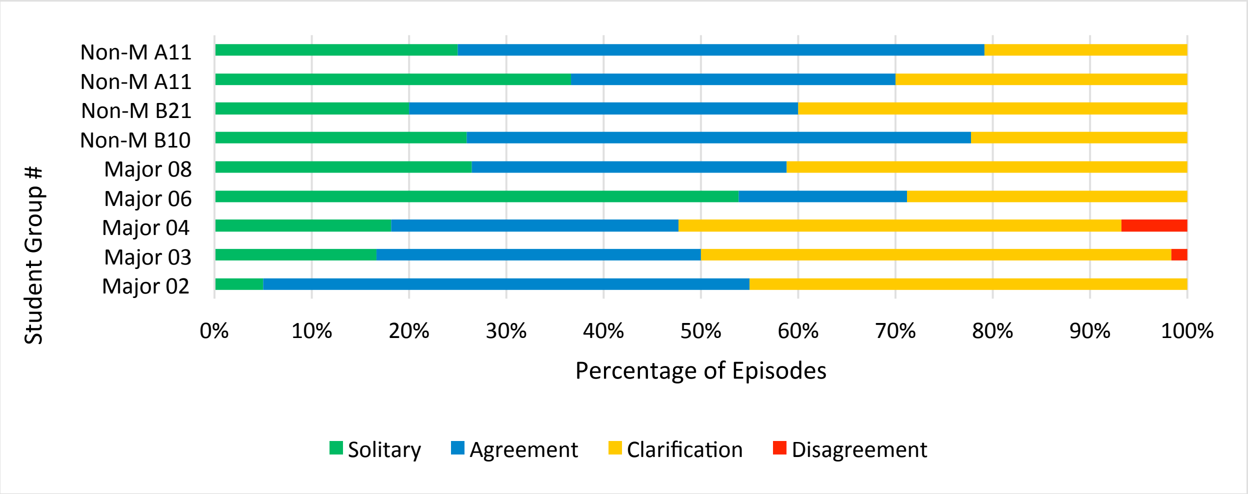


Figure 3. The percentage of episodes in each interaction level per group. An episode includes discourse, if any, and resulting drawing. Non-M = non-majors group.

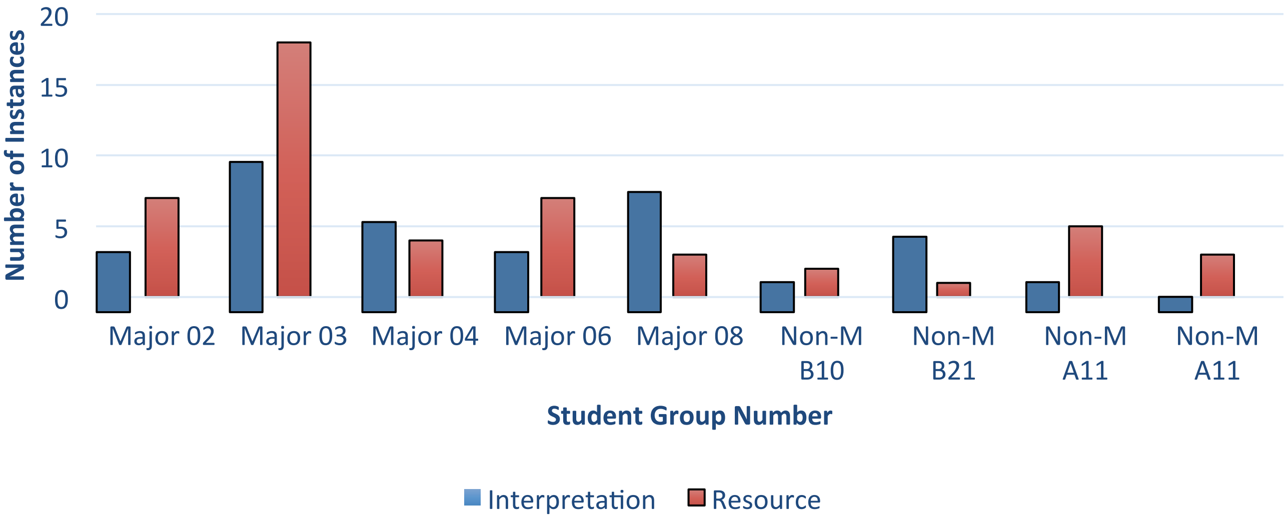


Figure 4. The number of justification instances of each type of justification per group. Non-M = non-majors group.

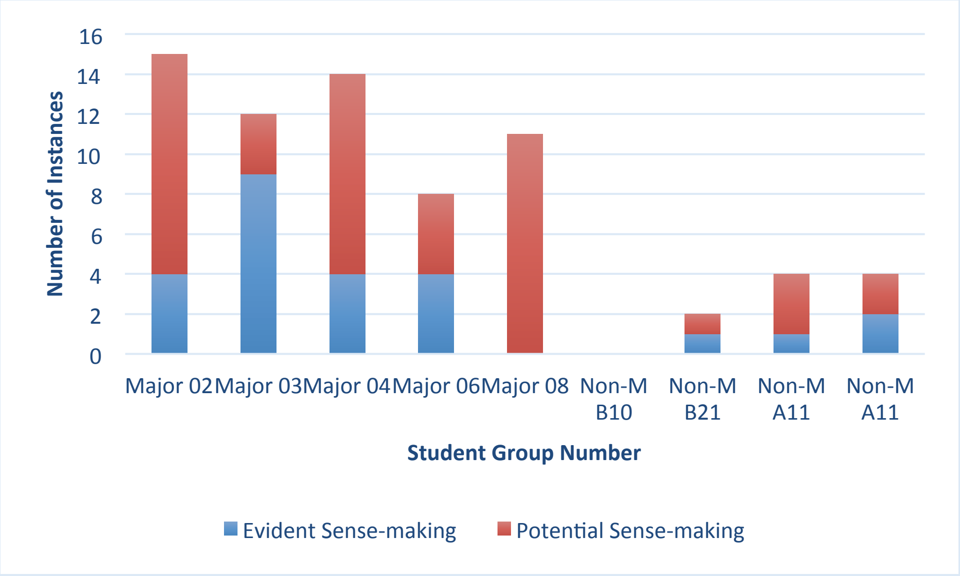


Figure 5. The number of sense-making instances per group. Non-M = non-m