Eliciting Middle School Students' Ideas About Graphs Supports Their Learning from a Computer Model

Eliane Stampfer Wiese (eliane.wiese@berkeley.edu)
Graduate School of Education, 4407 Tolman Hall, Berkeley, CA 94720 USA

Anna N. Rafferty (arafferty@carleton.edu)
Computer Science Department, One North College Street, Northfield, MN 55057 USA

Marcia C. Linn (mclinn@berkeley.edu)
Graduate School of Education, 4611 Tolman Hall, Berkeley, CA 94720 USA

Abstract

When middle school students learn science content with graphs, the graphing and science knowledge may be mutually reinforcing: understanding the science content may help students interpret a related graph, and information from a graph may illustrate a scientific concept. We examine this relationship between graphing and science by studying how students learn from interactive computer models with accompanying data graphs. The computer models provide an animated simulation that illustrates an unobservable phenomenon, while the data graph tracks one or more quantities over time. This ordering study, on middle school students learning about photosynthesis, indicates that engaging with novel graph concepts helped students interpret their data as they experimented with the computer model. The study also provided some support for the opposite direction: experimenting with the model first helped students make sense of the graphs.

Keywords: Graphing; Photosynthesis; Knowledge Integration

Learning From Graphs in Science

Graphs are important in science and improving students' integration of graph and science concepts is often neglected, especially at the middle school level (Lai et al., 2016). Graphs can help students test hypotheses to make sense of new science concepts (e.g., Vitale, Madhok, & Linn, 2016). When linked to an animated simulation, time-series graphs can record and summarize information that points to key relationships. Identifying the main points of a computer simulation is not trivial for students: animations can be deceptively clear, giving a false sense of understanding (Chiu, Chen, & Linn, 2013), and students may focus on discrete events instead of looking for an underlying pattern (Vitale et al., 2016). While animations provide visualizations of mechanisms and processes, graphs provide visualizations of relationships between variables. Thus, graphs may help students focus on the underlying quantitative relationships in the animations.

While graphs have the potential to be extremely useful for learning science content, middle school students often have difficulty interpreting them (Lai et al., 2016). Graphs are generally taught in math classes, and students may not spontaneously transfer their knowledge to science contexts (Grant, 2013). Prior work notes common student problems in graph interpretation, many stemming from misunderstanding what the graph axes represent (e.g., Clement, 1985). This type of confusion may result from shallow reasoning based on superficial similarity (e.g., Janvier, 1981) or the ease of extracting irrelevant data from the graph (e.g., Clement, 1985; McDermott, Rosenquist, Popp, & van Zee, 1983). Therefore, while graphs can be powerful tools, novices need support to work with them. Misinterpretations of graphs could lead students to misunderstand the science being conveyed. On the other hand, if students have a firm grasp of the science content, that knowledge may help them interpret the graph by constraining potential interpretations to ones that are consistent with their science knowledge (Ainsworth, 2006). It is therefore plausible that graph and science knowledge could be mutually reinforcing. To examine if this reinforcement is uni- or bi-directional, we assigned students to see graph-focused steps either before or after an animated computer model of plant growth. Graph First students learned more science from the simulation, with some advantage for Model First students in interpreting a specific graph feature.

Photosynthesis in the WISE Platform

The ordering experiment on graphing was implemented via a seventh-grade unit on photosynthesis, which was designed with the Knowledge Integration (KI) framework (Linn, Lee, Tinker, Husic, & Chiu, 2006), using the Web-Based Inquiry Science Environment (WISE: http://www.wise.berkeley.edu). The unit focuses on processes of energy transformation and aligns with the Next Generation Science Standards (NGSS Lead States, 2013). In photosynthesis, glucose is created and stores energy. The energy is released during cellular respiration. A key difference between photosynthesis (making glucose) and cellular respiration (using glucose) is that plants can only make glucose when there is light, but they must use glucose all the time to perform basic cellular functions. This concept is targeted in the plant growth activities (Figure 1).

The plant growth activities begin with predicting when the plant will make and use glucose (multiple-choice text selection). Next, students interact with the animation from the plant growth model, which shows the plant's glucose stores growing when the light is on and shrinking when the light is off. Students then draw a graph to show their predictions for the cumulative glucose that will be made, used, and stored with the light on for a few weeks and then off, and also select which graph shapes match their predictions (multiple choice: all lines increase with light. In the dark, one is horizontal, one decreases, one increases at the same rate, and one in-
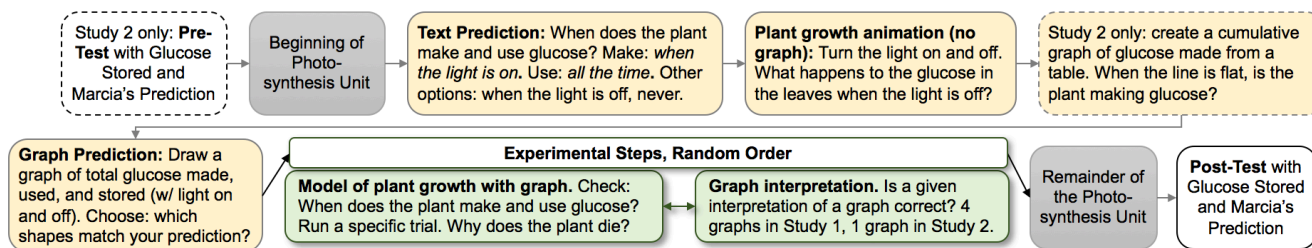
Figure 1: Sequence of activities in Studies 1 and 2.

creases more sharply). Finally, students see the experimental steps: interacting with the full plant growth model (with the graph) and graph interpretation. The two experimental steps, described below, are ordered randomly. Test questions assessed how students connect ideas. As shown in Figure 3a, one prompt presented a graph of glucose stored and asked students to explain how turning the light off affected the shape of the graph (Glucose Stored). Another prompt gave an incorrect prediction: that glucose stored would not change when the light was off (Marcia's Prediction). Students explained if the prediction was correct and matched it to a graph (Figure 3b). These questions were intended to measure how well students can connect a graph's shape to its meaning.

Exploring the Plant Growth Model

The plant growth model allows students to turn a light on or off, while the accompanying time-series graph shows the cumulative amounts of glucose made, used, and stored (Figure 2c). In this model, glucose is made at a constant rate when the light is on, and glucose is used at a constant rate all the time. Students were again asked when the plant makes and uses glucose. The cumulative graph shows these relationships directly: the running total for glucose made only increases while the light is on, but the running total for glucose used increases until the plant dies.

Students were also asked to run a specific trial with the model: leave the light on for 4 weeks in a row, and then turn it off. This trial shows that plants draw on stored glucose to survive in the dark. To structure students' thinking about this trial, students were asked when the plant died, and what best explained why the plant died (Figure 2c). Correctly answering both questions requires reading the graph.

Ordering Study: Graph First vs. Model First

The graph that accompanies the photosynthesis model is complex: it is cumulative rather than instantaneous and it presents three quantities changing over time. Science knowledge may help students avoid the error of reading the graph as instantaneous (e.g., the line for glucose made is not zero in the dark). While science knowledge could help students interpret the graph, the graph may also help students learn science by illustrating a relationship between quantities of glucose made, used, and stored that explains how plants survive periods of darkness. The KI framework proposes that eliciting students' ideas is crucial for learning because ideas

cannot be leveraged or inspected if they are not first brought to the student's attention. We designed a graph interpretation step to elicit students' ideas about graphs and to prompt them to connect their ideas about a graph's shape to its meaning. Questions asked if an interpretation of a given graph was correct (e.g., the plant is making glucose when the glucose made line is flat). The graph interpretation step was more extensive in Study 1, and is described in more detail for each study.

To examine the effects of graph and science knowledge on each other, we sequenced the graph interpretation step either directly before or after the computer model. Two seventh grade teachers at Bay Area public schools used this WISE unit to teach photosynthesis, and their students were randomly assigned to either the Graph First or Model First sequence. Students did pre- and post-tests individually, and did the unit individually or in teacher-assigned pairs.[1]

Study 1: Extensive Elicitation of Graph Ideas

We ran the first iteration of the study in six 7th grade classes, all taught by Mrs. R.[2] Students completed a pretest, but it is not included in the analyses because it was not matched to the post-test. Since the pre-test and the unit took longer than Mrs. R. expected, she requested a revised and shortened post-test.

Materials. The post-test[3] included the questions Glucose Stored and Marcia's Prediction (Figure 3). Students could access the plant growth model on both the pre- and post-test. The experimental step to elicit students' graph ideas asked students to interpret four cumulative graphs, all showing glucose made, used, and stored when the light was shining (three also showed darkness). Two graphs were consistent with the model and two were incorrect. For each graph, students were given one or two interpretations and asked if those interpretations were correct. In total, the elicitation step included five true/false items (one shown in Figure 2a) and three open-ended prompts for students to explain their reasoning.

---

[1] A separate, unrelated study comparing forms of scaffolding for essays was run concurrently in the same unit. The target essays did not involve the computer model or graphs, and condition assignment was independent for the two studies. Our results do not show condition effects of the unrelated study.

[2] The full unit is available at http://wise.berkeley.edu/ preview-project.html?projectId=18309

[3] The post-test is available at http://wise.berkeley.edu/ preview-project.html?projectId=18463

(A) True or False: This graph says that the plant makes and uses some glucose when the light is on, and makes and uses 0 glucose when the light is off.

(B) Sasha says "In week 4, the plant *is not* making glucose. I know because the line is flat for that week." Malia says: "In week 4, the plant *is* making glucose. I know because the line is not at 0." Who is right? **Sasha**, Malia, both, or neither?

(C) Prompted trial: Reset the model, and leave the light on for four weeks. Then turn it off. When does the plant die? **Week 10.** The plant died then because that was when… the plant stopped making glucose; the light was turned off; the line for total glucose made was at 0; **the line for total glucose stored was at 0**; the line for total glucose used was at 0.
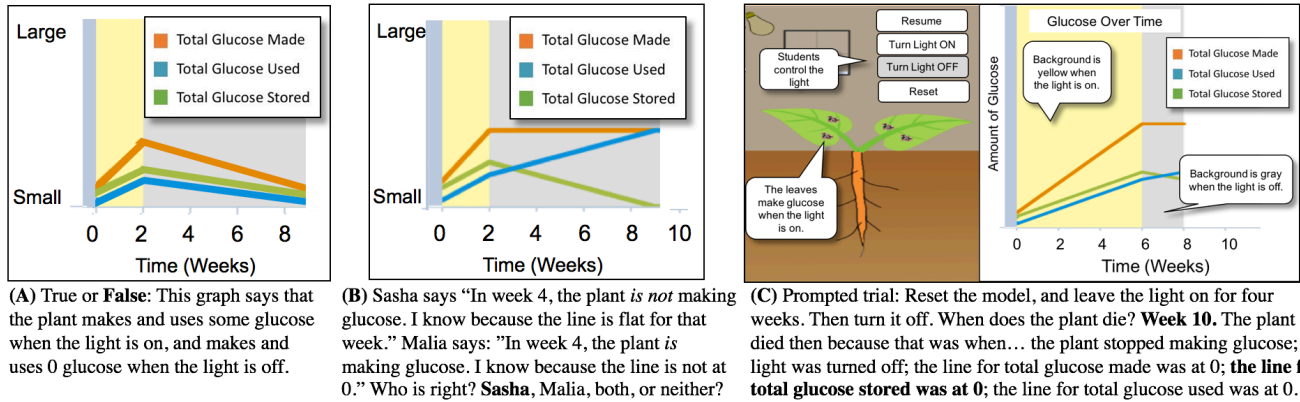
Figure 2: Screenshots. (a) A true/false graph interpretation item in Study 1. (b) The graph interpretation item in Study 2. (c) The plant growth model, with an animation on the left and graphed the output on the right. Correct answers in bold.

Method and Participants. The pretest was administered over 1-2 class days before students began the unit. Students then worked on the unit for seven class days. Most students worked in teacher-assigned pairs, and they were randomly assigned to the Graph First or Model First conditions (within-class). Most students completed the model-related steps over two class periods. The posttest was administered two school days after the unit, and took one class period.

177 students across Mrs. R's six classes worked on the WISE unit, with 24 to 35 students per class. Five students worked individually, with the rest in pairs. The 89 groups (176 students; 43 Graph First and 46 Model First) who did the experimental activities are included in the analyses of embedded items. Of those, 173 did the posttest and are included in the post-test analyses.

Results: Embedded Assessments. 96% of groups correctly selected that the plant would make glucose only when the light was on (Graph First: 98%, Model First: 93%). 71% of groups correctly selected that the plant would use glucose all the time (Graph First: 72%, Model First: 70%). Students selected which of four shapes (or none) best matched their ideas for glucose made, used, and stored (all lines increase when the light is on. When the light is off, one increases at the same rate, one increases more sharply, one decreases, and one is horizontal.). Most students selected the correct shapes for glucose made and stored, but only 43% did so for glucose used. The most common incorrect shape was flat after the light was turned off (22%). Of students who predicted, in text, that plants would use glucose all the time, only 55% selected the correct graph. Thus either students hold multiple conflicting ideas about when plants use glucose or are unable to connect their science ideas to the graph shape.

Students were engaged with the model; 93% of groups ran at least one trial where the plant died. However, students' choice to run the suggested trial (light on for four weeks, then off) differed by condition: 88.4% of Graph First groups vs. 71.7% of Model First groups. 70% of Graph First groups and 45% of Model First groups chose the correct explanation for why the plant died. A logistic regression on answer correctness (with factors for conditions in this study and the unrelated study) confirmed that there was a main effect of condition on whether they responded correctly ($t(86) = 2.27$, $p = .023$). A logistic regression on whether students ran the suggested trial found a marginally significant effect of condition ($t(86) = 1.92$, $p = .055$). To determine if running the trial mediated the effect of condition, we re-ran the previous logistic regression on students' answers with running the trial as an additional factor. Running the trial was the only significant factor ($t(85) = 3.20$, $p = .0014$; 67.6% correct if trial was run vs. 16.7% correct without this trial), indicating that running the trial mediated the effect of condition. Thus, structuring exploration with a specific trial is more effective for learning than non-structured exploration, even when both uncover the key information (i.e. the plant dying).

To determine if lack of engagement in general led students to ignore the directions, we coded all five open-response items within the activity (but before the experimental steps) as on-task (answering the question, correctly or not) or off-task (blank, *I don't know*, or nonsense answers). 20 responses were coded by two authors with 100% agreement; remaining responses were coded by one of the authors. Both conditions showed high engagement, with means of 98% on-task responses (means were 4.91/5 for Graph First, 4.89/5 for Model First, $t(87) = .16$, $p = .87$). We repeated the logistic regression on students' answers with all covariates above and engagement as an additional covariate ($t(84) = 0.68$, $p = .50$) and found that only running the suggested trial was significant ($t(84) = 3.39$, $p = .0007$). Repeating the logistic regression on whether students ran the suggested trial with factors for conditions and engagement also showed the same effect: ordering condition was marginally significant ($t(85) = 1.90$, $p = .058$), and engagement was not ($t(85) = 0.54$, $p = .58$).

Performance on the graph items did not show differences across conditions, suggesting that interacting with the model first did improve graph interpretation in general. Two questions focused on interpretation of a flat line in a cumulative

2: Vague, incorrect, or not focused on stored glucose. (1) *When the light was turned off less glucose was used.* (2) *no because, if you leave the light off it will die*

3: Partially correct or simple. (1) *The total glucose stored was steadily going up then when the light turned off it went down.* (2) *Marcia's prediction is not correct. Even though the plant will not make glucose or use anymore, the amount of glucose stored would go down and so will the glucose made.*

4S: Plants survive in the dark by using stored glucose. (1) *when the light turned off the plant did not make glucose but only used the glucose that the plant stored.* (2) *No, a plant to use glucose needs glucose. And plant don't get the needed sunlight when the light is off so has to eat into the glucose stored.*

4G: Explanation of graph shapes. (1) *Turning the light off in addition to causing the glucose stored to go down, it would also cause the glucose made to stop increasing and just stay the same. Turning the light off would not, however, effect the total glucose used, which would keep going up.* (2) *Marcia's prediction is incorrect because the glucose stored line in the graph goes down and doesn't stay level.*

5: Relating the graph's shape to the plant's use of stored glucose. (1) *It affected the shape of the graph because the plant cannot make glucose, so the plant has to use the stored glucose. This makes the stored glucose go down in quantity.* (2) *No, because if the plant doesn't make any at night and still uses it where is it coming from. It must be the storage and that must make it go down.*

Table 1: Rubric for scoring (1) Glucose Stored and (2) Marcia's Prediction (0 is blank, 1 is off-topic or "don't know").

graph. Groups in the Graph First condition had lower average scores on these items than groups in the Model First condition (average score 55% vs. 80%; ANOVA with factors for conditions in both studies found a main effect of ordering condition: $F(1,88) = 13.23$, $p < .001$). Both conditions scored near chance for the other items, 46%-63%.

Results: Post-Test. The test items Glucose Stored and Marcia's Prediction each have one question with an objectively correct answer (when is the light turned off and which graph matches Marcia's prediction) and one open-response that encourages connections between graph and science ideas. The objective questions were scored as correct or not. For the open-ended questions, we developed KI rubrics that required linking the graph shape to processes of plant growth to get the highest score (a 5). Full explanations of either the graph or the science concepts were scored a 4, and were coded for graph or science ideas (see Table 1). Two of the authors coded an initial set of responses for each question, reconciling differences and revising the rubric. The same two authors coded a second test set for each question (39 answers for Glucose Stored, with 85% agreement and Cohen's kappa of .79; 31 answers for Marcia's Prediction, with 90% agreement and Cohen's kappa of .85). For both questions, the second test sets included responses from both studies.

For Glucose Stored, most students correctly identified that the light was turned off in week 6 (Graph First: 88%, Model First: 90%). Mean scores for their explanations of how turning the light off affected the shape of the graph were 2.9 overall (2.94 for Graph First and 2.85 for Model First). All students scored a 2 or higher, indicating engagement with the task (0s are blank and 1s are off task). 24% of the Graph First and 21% of the Model First condition scored a 4 or 5.

For Marcia's Prediction, students were below ceiling but above chance for selecting the graph that matched her incorrect prediction (Graph First: 62% correct; Model First: 57% correct). Mean scores for their explanations of why the prediction was correct or not were 3.4 overall (3.5 for Graph First and 3.2 for Model First). 98% of students gave on-task answers. 60% of the Graph First and 50% of the Model First

condition scored a 4 or 5. A MANOVA on the two explanation items with conditions in both studies as factors showed no significant difference by condition for either study (ordering condition: $F(2,168) = 1.2$, $p = .3$).

Study 2: Targeted Elicitation of Graph Ideas

Materials. Since Mrs. R's students took longer than we anticipated to complete the experiment, we shortened the materials before repeating the study with Mr. W's classes. The first questions on the pre- and post-test were the same as on the post-test in Study 1, with additional questions afterward. In the unit, on steps that all students saw in the same order, we removed open-response explanation prompts or replaced them with multiple-choice questions. Since Mrs. R's students were confused about the flat segment of the glucose made line, we added a cumulative graph activity. Students were given a table of glucose made each week and filled in a column with the running total. This running total was then shown as a graph, and students were asked to interpret the flat segment. In Study 1, students drew prediction graphs and then characterized the shapes afterward. In Study 2, we reversed this sequence to help students plan their predictions. We also shortened the steps in the ordering experiment. Instead of eliciting ideas about several graphs, we asked students to interpret the flat segment of the glucose made graph with a multiple choice question and an open-response explanation (see Figure 2b). In the modeling activity we removed all open-ended prompts, and replaced them with one prompt that asked students if their initial predictions on when the plant would make and use glucose had been correct. We retained all multiple choice questions.

Method and Participants. As in Study 1, the pre- and post-test were done individually, each on a single class day, with the pretest two days before students started the unit and the posttest right after. Students did the unit in groups of 1-2 over five class days, with a one-week gap between the first three days and the final two days. As in Study 1, most students took 1-2 class days for the experimental steps.

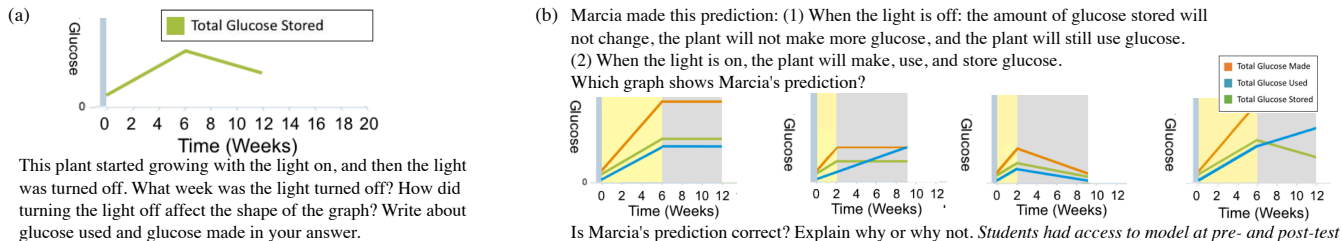79 students across Mr. W's 7th grade classes used the

Figure 3: (a) Glucose Stored requires graph interpretation and science knowledge of when a plant makes and uses glucose. (b) Marcia's Prediction requires finding a graph shape to show an idea, and critiquing that idea with science knowledge.

WISE unit, with 23-29 students per class. As in Study 1, groups were assigned randomly to conditions (within-class). 24 groups did not reach the experimental steps and are not included in the analysis. The 10 Graph First groups and 18 Model First groups are included in the analysis below (43 students total. 11 did the unit individually, the rest in pairs). Of those students, 39 did both the pre- and post-test and are included in those analyses (13 Graph First and 26 Model First).

Results: Embedded Assessments. 79% of groups correctly predicted when the plant would make glucose (Graph First: 90%, Model First: 72%) and 29% correctly predicted when the plant would use glucose (Graph First: 30%, Model First: 27%). Students were not above chance (11 − 25%) in selecting which graph shapes matched glucose made, used, and stored. In the most common shape for total glucose made (35%), the line decreased in the dark, a nonsensical answer for *cumulative* glucose made. Students with the correct text predictions were somewhat more likely to select the correct graphs, but still performed poorly (14% correct for glucose made and 25% for stored). These predictions suggest that students struggled both with the science content and the graphs.

Only 75% of groups interacted with the model. To measured engagement, we coded for on-task responses on earlier step. On the two earlier steps, overall 75% of responses were on-task (means were 1.6/2 for Graph First, 1.4/2 for model first, $t(26) = .56$, $p = .57$). Despite lower engagement, we ran the same analyses as in Study 1. Graph First groups ran the suggested trial more frequently than Model First (50% vs. 27.8%), but were not more correct in explaining why the plant died (10% vs. 11%). These trends were not significant for condition (all analyses here are logistic regressions with factors for conditions in both studies; ordering condition $t(25) = 1.17$, $p = .24$ for suggested trial and $t(25) = 0.0$, $p = 1.0$ for answering correctly). Of the 10 groups who ran the trial, 3 answered correctly, compared with *none* of the 18 groups who did not run the trial, but this trend was not significant ($t(24) = .0003$, $p = .99$). We repeated the logistic regression on running the trial with engagement as an additional covariate, which was not significant ($t(23) = 0.96$, $p = .33$). The replication of trends but not reliable effects may be due to small sample sizes, lack of engagement, or lack of prior knowledge, leading to very low performance on many items. Performance on the graph interpretation item

was above chance but below ceiling; 54% of groups correctly said that the plant was not making glucose when the line was flat. There was no significant difference by condition.

Results: Pre- and Post-tests. On Glucose Stored, identification of when the light was turned off rose from pre- to posttest (from 4/13 to 9/13 for Graph First, and from 12/26 to 16/26 for Model First; McNemar test from pre- to posttest, $p = .02$). A logistic regression on posttest scores with pretest score and study conditions as covariates was not significant for ordering condition. Mean explanation scores also rose from pre- to post-test (Graph First: 1.2 to 2.3, Model First: 1.4 to 1.9). For selecting which graph matched Marcia's Prediction, students were below chance at pretest (2/13 correct for Graph First, 5/26 for Model First) and just above chance at post-test (Graph First: 4/13 correct, Model First: 7/26). Explanation scores rose from pre- to post-test for the Graph First group (.6 to 1.6) but not for Model First (.8 to .7). 82% of responses at pre-test were blank or off task, with the remaining students answering with incorrect or overly simple ideas. At post-test, 1 Model First and 3 Graph First students had a complex idea about the graph or the science.

A full factorial repeated measures ANOVA with test time, question, and conditions in both studies indicates improvement from pre- to posttest on the explanation items, with significant factors for test time ($F(1, 36) = 10.3, p = .003$), question ($F(1, 36) = 36.7, p < .001$), and a time by ordering condition interaction ($F(1, 36) = 4.8, p = .035$). While estimated marginal means for Graph First were lower than Model First at pretest (.9 vs. 1.1) and higher at posttest (1.9 vs. 1.3), ordering condition was not significant in a MANCOVA on the two post-test explanation scores (with pretest scores as covariates and condition in the unrelated study as a factor).

Discussion

This study provided support for graph understanding helping students to interpret a science model and some support for science understanding helping students interpret specific graph features. While students had difficulty interpreting cumulative graphs, eliciting students' ideas about those graphs helped them make sense of glucose production, use, and storage in the photosynthesis model.

Graph First students were more likely to figure out that the plant dies when it runs out of stored glucose. Interestingly,

this effect was strongly mediated by whether students ran the suggested trial leading to plant death (significant in Study 1, with the same pattern in Study 2). The finding that elicitation of graph ideas, without feedback, improved learning from the model is consistent with other research on knowledge integration (Linn et al., 2006). The cumulative graph in the model was more familiar to Graph First students since they had just been examining cumulative graphs. To run the suggested trial, students needed to read the graph to turn the light off at the right week, and Graph First students may have extracted this information better than Model First students. Furthermore, students in both conditions were equally engaged on prior steps and engagement was not predictive of following the directions to run the suggested trial. We note that condition differences were significant in Mrs. R's classes, and the pattern was similar in Mr. W's classes.

In both studies, performance was not mediated by running *any* trial where the plant died, only the suggested trial, which was followed by questions to structure students' thinking. This result is consistent with prior work showing that without structure, students may uncover the necessary information but miss key patterns (Reiser, 2004; Vitale et al., 2016).

Study 1 provides some evidence that exploring the model helped students interpret flat lines in cumulative graphs, as Model First groups performed better than Graph First groups on those questions. Science knowledge may have helped constrain students' interpretations of those graphs. However, there were no condition differences when considering all of the graph interpretation questions together; exploring the model may help only with particular graph features.

This research only varied the order of activities. All students were prompted to think about cumulative graphs before the experimental steps (Figure 1). The treatment elicited students' ideas about relevant graphs by suggesting interpretations that connected graph shapes to their meanings. This treatment may have prepared students to distinguish among alternative interpretations while exploring the model.

Further, this study captures difficulties that students have with cumulative graphs in science. In the graph prediction step, many students chose graph shapes that did not match their text predictions, and in the graph interpretation step, many students agreed with interpretations that did not follow from the given graph. Of the 71% of Mrs. R's groups who said that plants use glucose all the time, only 55% of them selected a graph shape that matched this prediction. The most common incorrect answer suggested that students were interpreting the graph as instantaneous instead of cumulative. Mr. W's students also seemed to demonstrate this error: the most common graph selected for glucose made decreased when the light was off. This would have been a more reasonable choice if the graphs were instantaneous instead of cumulative, as it would have been the only shape consistent with the plant making less glucose in the dark. While these examples show difficulties going from meaning to graph, students also had difficulty with the reverse in the graph interpretation

steps. Students in Study 1 were near chance for those steps, and almost half of the groups in Study 2 did not recognize that no glucose is made when that line is flat. Graph misconceptions documented in prior work have focused on instantaneous graphs. Our results show that the type of misinterpretations noted in prior work also affect cumulative graphs (specifically, interpreting the graph as if the axes were different, i.e., instantaneous vs. cumulative). Though students exhibited difficulties, Study 2 shows pre- to post-test improvement on graph interpretation and on explanations.

Conclusion. When learning science from computer models with complex graph output, students may benefit from interpreting graphs before interacting with the model. This study demonstrates the power of eliciting student ideas, a component of the knowledge integration framework that is intended to prepare students to distinguish among alternatives in the experiment (Linn et al., 2006). While graphs are currently under-utilized in middle school science, this study shows how graphs can be leveraged to help students connect graph and science ideas.

References

Ainsworth, S. (2006). Deft: A conceptual framework for considering learning with multiple representations. *Learning and instruction*, *16*(3), 183–198.

Chiu, J. L., Chen, J. K., & Linn, M. C. (2013). Overcoming deceptive clarity by encouraging metacognition in the web-based inquiry science environment. In *International handbook of metacognition and learning technologies* (pp. 517–531). Springer.

Clement, J. (1985). Misconceptions in graphing. In *Proceedings of the Ninth International Conference for the Psychology of Mathematics Education* (Vol. 1, pp. 369–375).

Grant, S. (2013). Graph Interpretation: How a Teacher Learns Pedagogical Content Knowledge. *Presented at the 2013 Annual Meeting of the American Educational Research Association*. San Francisco, CA.

Janvier, C. (1981). Use of situations in mathematics education. *Educational Studies in Mathematics*, *12*(1), 113–122.

Lai, K., Cabrera, J., Vitale, J. M., Madhok, J., Tinker, R., & Linn, M. C. (2016). Measuring graph comprehension, critique, and construction in science. *Journal of Science Education and Technology*, *25*(4), 665–681.

Linn, M. C., Lee, H.-S., Tinker, R., Husic, F., & Chiu, J. L. (2006). Teaching and assessing knowledge integration in science. *Science*, *313*(5790), 1049–1050.

McDermott, L., Rosenquist, M., Popp, B., & van Zee, E. (1983). Student difficulties in connecting graphs, concepts and physical phenomena. *Presented at the 1983 Annual Meeting of the American Educational Research Association*. Montreal, Quebec.

NGSS Lead States. (2013). *Next Generation Science Standards: For States, by States (Matter and Energy in Organisms and Ecosystems)*. http://www.nextgenscience.org/topic-arrangement/msmatter-and-energy-organisms-and-ecosystems.

Reiser, B. J. (2004). Scaffolding complex learning: The mechanisms of structuring and problematizing student work. *The Journal of the Learning Sciences*, *13*(3), 273–304.

Vitale, J. M., Madhok, J., & Linn, M. C. (2016). Designing a data-centered approach to inquiry practices with virtual models of density. In *Proceedings of the International Conference of the Learning Sciences* (pp. 591–598).