Interpreting Graphs to Distinguish Factors That Impact Climate Change

Elizabeth McBride, SRI International, beth.mcbride@sri.com
Marcia C. Linn, University of California, Berkeley, mclinn@berkeley.edu
Jonathan Vitale, School District of Philadelphia, jonvitale@gmail.com

Abstract: Scientists use models and graphs to distinguish among factors that impact a phenomenon (for example, the impact of CO2 accumulation on climate change) and factors that do not impact the phenomenon (for example the role of ozone depletion on climate change). In this paper, we compare two forms of exploration of time series line graphs: *plan* and *typical*. In the *plan* condition, students plan an experiment with a model by graphing the level of a system parameter (e.g., concentration of greenhouse gases) and the predicted response of an outcome variable (e.g., temperature). They then run the model to observe the accuracy of their predictions. In the *typical* condition, students run the simulation immediately and adjust the parameter level as they see fit. Students produced more informative experiments in the *plan* condition than the *typical* condition. Students in the plan condition made inferences by comparing their prediction to the outcome.

Introduction

Across many scientific fields, interpreting trends in longitudinal data is a primary means of identifying potential relationships and testing hypotheses. Data collected over time and displayed graphically (e.g. line graphs), may provide insight into relationships between variables. For science students, learning to identifying trends and relationships, or the lack thereof, from time series line graphs is an authentic inquiry skill and a means to learn important disciplinary concepts. In this paper, we conduct a study to support this skill in an online curriculum.

In particular, we investigate trend analysis in the context of climate change, where historical data measuring both greenhouse gas emissions and global temperature represent an important source of evidence for global warming. Although a large majority of climate scientists believe that this data presents clear evidence for warming, others, bolstered by non-scientific "interpretive communities" (Leiserowitz, 2005), cite evidence from the same data to undermine the warming hypothesis (e.g., a stable period in the early 2000's).

Research on middle school students' graph reading skills indicate that students have difficulty observing general patterns, making connections with the contextual domain, and making interpolations and projections that "go beyond the data" (Curcio, 1987). Indeed, interpreting graphs has been shown to be challenging for most people (OECD, 2006), a trend which is well- documented in mathematics education research (e.g., Cheung & Slavin, 2013; Rakes, Valentine, MaGatha, & Ronau, 2010).

In this research we sought ways to support students to build an integrated, scientifically accurate, and coherent understanding of a phenomenon while using a complex, interactive simulation featuring time series line graphs. Using the knowledge integration framework we supported students in intentionally testing their own ideas (e.g., predictions), gathering clear evidence of the impact of their tests, and distinguishing between valid and invalid ideas (Linn & Eylon, 2011; Raes, Schellens, & De Wever, 2013). We conducted this research using two conditions. In the *typical* condition, students adjust a parameter by manipulating a slider as the models run, and then interpret data generated in the graphs. In the *plan* condition students plan how the parameter will change over the course of the model run and predict its impact on the outcome variable by drawing graphs of each. In this *plan* approach, once both graphs (parameter, prediction) are constructed, the student runs the model and compares the predicted outcome to authentic data. We investigate two research questions in this study:

1. **Research Question 1**: Are students who generate graphs in the *plan* condition more likely than students in the *typical* condition to produce informative experiments?
2. **Research Question 2**: Are students who generate graphs in the *plan* condition more likely than students in the *typical* condition to make valid interpretations of relationships implicit in models?

Methods

Curricular materials

All instructional materials were implemented in the Web-based Inquiry Science Environment (WISE). WISE is a an open-source, online platform for developing inquiry materials and assessments (Linn, Clark, & Slotta, 2003).

The curriculum contained the following features: **Climate model.** The climate model used in this research allows students to make predictions and then test those predictions, and logs student clickstream data. Students see a visual representation of the energy transfer mechanism (solar radiation turns into heat, which turns into infrared radiation) on one side of the simulation, while the other side of the simulation shows a graphical representation of temperature vs. time. The computer model sequence shown in Figure 1 uses a single climate model as a basis with the same visualization features and energy representations, but changes the parameters that may be changed. **Interpreting graphs of unrelated variables.** To introduce graph interpretation, students completed exercises where they interpreting (unrelated) graphs (e.g., of "pollen count" and "temperature"). These exercises introduced global trends and null relationships. **Analyzing climate models**. Students manipulate parameters in a series of climate models (in NetLogo; Wilensky, 1999). They vary one parameter and observe its effect on an outcome variable. To introduce the activity, students analyze an accessible example of a fire and thermometer. Subsequent models depict relationships related to global phenomena. They alternate between causal and non-causal parameters (Figure 1). In the *typical* condition, students manipulate the parameter with a slider while running the model. In the *plan* condition, students generate a graph prior to running the model (*plan* condition). **Making predictions and analyzing outcomes.** All students made predictions concerning the relationships among the parameter and outcome for each pair of models as part of intentional learning. Students were also prompted to explain, after running the simulation, "how the evidence you gathered from the models supports or contradicts your predictions."

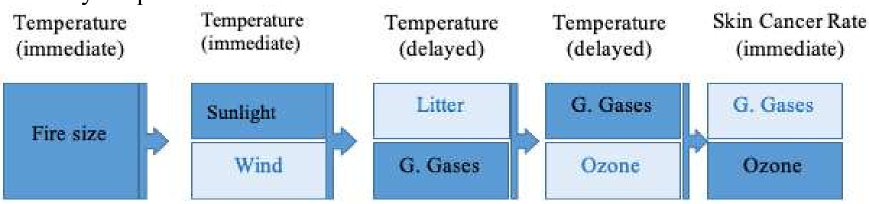

Figure 1. Computer model sequence. Text above boxes indicates the outcome variable of the model. Text in boxes indicates parameters to be manipulated. In some cases the control parameter has a (nearly) immediate effect on the outcome variable, while in others the response is delayed. Dark boxes indicate the variable is related to the outcome variable. Light boxes indicate the variable is unrelated to the outcome variable.

Assessments

The pre- and post-assessment materials measure integrated understanding by assessing evidence generation practices and graph interpretation capabilities using two items: Best Experiment and Explain to Armando.
Best Experiment (posttest only). In this short essay item, students explained which of two graphs was more likely to produce an interpretable result. In this case, Luke's response includes more fluctuations of the control parameter (greenhouse gases). Because of the delay between the greenhouse gas accumulation and temperature change, this parameter profile will not show a clear relationship between variables.
Explain to Armando. In this short essay item students were asked to help a fictional student understand the impact of greenhouse gas concentration and/or the ozone layer in global warming.

Analysis

We use logged data to analyze how students interpret the relationships in six embedded models (Figure 1). We omitted sunshine and temperature and greenhouse gases and temperature (with ozone) due to ceiling effects. We analyzed the process data for each group to recreate the parameter graph produced in the model. Based on our prior research developing automated scoring algorithms for student generated graphs (Vitale, Lai, & Linn, 2015), we designed classifiers to analyze this data.

We use the following as measures of student actions: ***Number of changes.*** In our system, students derive important information about the models by changing a parameter. This metric is a count of the number of drastic changes in slope students make when changing the parameter in the model. ***Sustained-number of changes***. Although *changes* are necessary to show relationships between model parameters and outcome variables, because of noise in the system and temporal delay between parameter changes and outcomes, *changes* that are sustained over time, provide clearer information about relationships. ***Parsimony***. Although there is a limit to the number of *sustained-changes*, the number of *changes*, is only limited to the amount of times the student can move the slider back and forth within the simulation run time. A large number of these motions, without sustaining the direction, would provide difficult-to-interpret information. A parsimonious graph with a minimal number of non-sustained changes would provide stronger information. To measure *parsimony* we calculate the ratio of *sustained changes* to overall *changes*. ***Trend analysis in model steps***. At the bottom of each page displaying a model, students are prompted to describe the relationships presented in the model. Although this is an open response item, we coded the scores dichotomously. We use logistic regression to analyze factors determining correct recognition of relationships, including experimental condition and graph features. ***Pretest-posttest***. The pretest and posttest include both multiple-choice and open response items. We focus on the open response items and score responses using a knowledge integration rubric.

Participants and procedure

One teacher and his 186 sixth grade students, in a suburban area, participated in this study (2% English language learners, 12% free or reduced-price lunch). Students completed the pretest using individual computers in a computer lab in one class section. On the following day, students were paired into workgroups according to intact seating assignments. Students were automatically assigned to *plan* or *typical* conditions by the software.

Results

We found that students in the *plan* condition typically produced generally linear, rising or falling *changes* in direction, while those in the *typical* condition generally produced step-wise graphs.

Focusing on only the final graphs students made in each activity, each condition displays different distribution characteristics for all three measures of interest (*number of changes*, *sustained-number of changes*, *parsimony*). For the *plan* condition, the most common *number of changes* was one (e.g., a single rising or falling segment), while one was never the most common *number of changes* in the *typical* condition (e.g., a single shift of the parameter). Likewise, regarding *parsimony*, students in the *plan* condition most commonly made perfectly parsimonious graphs, with no non-informative, un-sustained changes. On the other hand, students in the *typical* condition often produced graphs with unnecessary segments (Table 1). For the first five selected models (all except oz-cancer), more *parsimony* emerged in the *plan* condition than the *typical* condition. In the final model (oz-cancer), although those in the *typical* condition produced more turns, this was offset by more sustained turns as well.

Table 1: Displays non-parametric analyses of two measures of student graphs. Significant differences are bolded (Wilcoxon rank sum test)

Model	*Number of changes*						*Parsimony*				
	Median		z	p	r		Median		z	p	r
	Plan	*Typ.*					*Plan*	*Typ.*			
wind-temp	2	3	-1.8	0.07	-0.2		**1**	**0.5**	**2.8**	**0.01**	**0.31**
litter-temp	2	3	-1.1	0.28	-0.11		**0.67**	**0.4**	**2.7**	**0.01**	**0.29**
ghg-temp	2	2	-0.8	0.41	-0.09		**1**	**0.5**	**2.8**	**0.00**	**0.29**
oz-temp	2	2.5	-0.8	0.43	-0.08		**1**	**0.5**	**2.1**	**0.04**	**0.22**
ghg-cancer	2	2.5	-1.6	0.1	-0.17		**1**	**0.5**	**2.9**	**0.00**	**0.3**
oz-cancer	**1**	**2**	**-2.7**	**0.01**	**-0.29**		1	0.67	1.9	0.06	0.2

To determine whether the *plan* procedure facilitates more accurate interpretation of relationships than the *typical* condition we focused on the accuracy of trend analyses made immediately following interaction with the models. As Table 2 displays, the percent of workgroups that provided accurate interpretations of model relationships was greater in the *plan* condition than the *typical* condition in all but the final model. A mixed-model logistic regression with condition as a fixed effect and workgroup and model as random effects, demonstrates a significant effect for the *plan* condition [B = 0.69, odds-ratio: 2.0, SE = 0.32, z = 2.1, p = .03], indicating that, overall, students in the *plan* condition were more likely than those in the *typical* condition to recognize the correct relationship between two variables in a model.

Table 2: Percent of workgroups that correctly identified relationship in models, by condition

Condition	*wind-temp*	*litter-temp*	*ghg-temp*	*oz-temp*	*ghg-cancer*	*oz-cancer*
Related?	No	No	Yes	No	No	Yes
Plan	0.73	0.81	0.66	0.75	0.75	0.71
Typical	0.44	0.58	0.63	0.68	0.57	0.8

To determine whether there was a sustained impact of alternate designs on graphing practices (i.e., what is valid evidence) we compared posttest KI scores on the *Best Experiment* item with independent samples t-tests.

We found that students in the *plan* condition, on average, earned higher scores than those in the *typical* condition [*plan*: $M = 2.3$, $SD = 0.69$; *typical*: $M = 2.1$, $SD = 0.50$; $t(156) = 2.2$, $p = .03$, $d = 0.35$]. Thus, students in the *plan* condition were more likely than those in the *typical* condition to correctly explain why the more parsimonious graph represented a better experiment incorporating concepts such as temporal delay and the impact of noise. To determine the impact of our curriculum on content learning we measured changes in the *Armando* item from pretest to posttest. Overall, students significantly increased their understanding from pretest to posttest [*pretest:* $M = 2.5$, $SD = 0.8$;1 *posttest*: $M = 2.8$, $SD = 1.1$; $t(159) = 3.6$, $p < .001$, $d = 0.3$]. Comparing posttest scores by condition (with pretest as a covariate), does not demonstrate a significant difference between condition [$F(1, 157) = 0.24$, $p > .2$].

Discussion and conclusions

In alignment with our first research question (Are students who generate graphs in the *plan* condition more likely than students in the *typical* condition to produce informative experiments?), our results indicate that overall, students in the *plan* condition were more likely to generate simpler, more parsimonious graphs than those in the *typical* condition. Creating simple graphs gave students the opportunity to test their ideas directly.
Simple graphs mirrored graphs in the preceding prediction step. In contrast, on the posttest *Best Experiment* item, students in the *typical* condition were more likely to assert that rapid changes would provide stronger evidence than gradual changes. Thus, more students in the *plan* condition achieved the subtle insight that less frequent change produces more accurate interpretation.

In response to the second research question (Are students who generate graphs in the *plan* condition more likely than students in the *typical* condition to make valid interpretations of relationships implicit in models?), students in the *plan* condition were more likely than those in the *typical* condition to identify correct relationships in the models. This difference may reflect the difference in parsimonious graphs: with simpler graphs, students in the *plan* condition were more likely to recognize valid trends. However, it is also possible that the plan condition led to a more intentional approach to the model activity and facilitated better interpretations, regardless of the quality of the graphs.

Together, these results show that the *plan* condition, compared to the *typical* condition, enabled students to distinguish causal factors from the climate model during instruction. Although other features of instruction (including a concept mapping activity with automated guidance) may have mitigated the difference between conditions on posttest outcomes, these results suggest that generating parameter and prediction graphs can increase insight into the mechanisms in the climate model and the design of experiments to detect these mechanisms.

References

Cheung, A. C. K., & Slavin, R. E. (2013). The effectiveness of educational technology applications for enhancing mathematics achievement in K-12 classrooms: A meta-analysis. *Educational Research Review*, *9*, 88–113.

Curcio, F. R. (1987). Comprehension of mathematical relationships expressed in graphs. *Journal for research in mathematics education*, 382-393.

Leiserowitz, A. (2005). Communicating the risks of global warming: American risk perceptions, affective images and interpretative communities. In S. Moser & L. Dilling (Eds.), *Creating a climate for change: Communicating climate change - facilitating social change* (Vol. 25, pp. 1–32). Cambridge Univ. Press.

Linn, M. C., Clark, D., & Slotta, J. D. (2003). WISE design for knowledge integration. *Science Education*, *87*, 517–538.

Linn, M. C., & Eylon, B.-S. (2011). *Science learning and instruction: Taking advantage of technology to promote knowledge integration*. New York: Routledge.

Organization for Economic Cooperation and Development. (2006). *PISA 2006 Science Competencies for Tomorrow's World*. Accessed May 7th, 2018, from www.oecd.org/fr/education/scolaire/programmeinternationalpourlesuividesacquisdeselevespisa/pisa2006results.htm

Raes, A., Schellens, T., & De Wever, B. (2013). Web-based Collaborative Inquiry to Bridge Gaps in Secondary Science Education. Journal of the Learning Sciences, 22. doi:10.1080/10508406.2013.836656

Rakes, C. R., Valentine, J. C., McGatha, M. B., & Ronau, R. N. (2010). Methods of instructional improvement in algebra: A systematic review and meta-analysis. *Review of Educational Research*, *80*(3), 372–400.

Vitale, J. M., Lai, K., & Linn, M. C. (2015). Taking advantage of automated assessment of student-constructed graphs in science. Journal of Research in Science Teaching, 52(10), 1426-1450.

Wilensky, U. (1999). {NetLogo}.

