Assessing Students' Ability to Create and Use Models to Explain Energy-Related Phenomena

Joseph Hardcastle[1], Cari Herrmann Abell[2], George DeBoer[1]

[1]American Association for the Advancement of Science/ Project 2061 [2]BSCS Science Learning

Paper Presented at the 2019 NARST Annual Conference

Baltimore, MD

Abstract:

Students' ability to explain phenomena were compared when they were provided a model versus asked to draw their own model. As part of a pilot test, 1,405 students in the fourth through twelfth grades from across the United States responded to one of three different modeling tasks. Each task presented students with a phenomenon related to energy and asked them to either draw a model or use a provided model to explain the phenomenon. Students were then prompted to write explanations of the phenomenon and answer several multiple-choice questions assessing their knowledge of the topic of energy. Students' explanations, drawn models, and responses to the multiple-choice questions were then scored and compared. Our results indicated that students who were provided a model wrote better explanations than students who had to draw their own model, and that students' drawn models added little to their explanation that wasn't captured in their written responses. When comparing students' drawn models, we found that students who drew the most sophisticated models had significantly more understanding of the energy concept then students who drew the least sophisticated models.

Introduction

Models are essential to the production, dissemination, and acceptance of science (Gilbert, 2004). Their central role is to help develop knowledge about the world. Science education has sought to reflect the importance of models in science with their inclusion in past and present science education standards (National Research Council, 2012; NGSS Lead States, 2013; American Association for the Advancement of Science, 1993). These frameworks focus on students learning to develop and use conceptual models as tools for thinking, making predictions, and explaining phenomena. Several curriculum materials engaging students in modeling have resulted in gains in students' science and modeling content knowledge (Chinn & Samarapungavan, 2009; Lehrer & Schauble, 2006; Schwartz & White, 2005; Schwarz et al., 2009), however, there are still questions regarding how to assess students' modeling ability.

Assessments of modeling have largely focused on measuring students' ability to create and evaluate models, and students' metamodeling knowledge (for a review see Namdar & Shen, 2015). These epistemological measures are useful for understanding what students' general knowledge of models and modeling is but are distinct from measuring the practice of using and creating models to explain a phenomenon. While a student's metamodeling knowledge influences their modeling of a phenomenon, other factors like science content knowledge and familiarity with the context in which the student is asked to model may play important roles (Fortus, Shwartz, & Rosenfeld, 2016; Ruppert, Duncan, & Chinn, 2017).

One area which has received little attention is how *using a model* and *constructing a model* are different concepts or explanatory tools. Using and creating models are commonly linked in modeling language and it is unclear whether this is a tautology or two different concepts being mixed. Schwarz *et al.* distinguished constructing and using models with both practices being linked to explaining phenomena (Schwarz et al., 2009). Fortus *et al.* suggested that students' knowledge of models and use of models as tools for explanation may be a different dimension of learning from students' knowledge and skill in modifying/revising models to improve their understanding (Fortus et al., 2016).Creating and improving models may be more a tool for students to improve or create new knowledge, while using models could be more an application of a student's knowledge of the model and how it can be used to explain a phenomena.

In this work we seek to examine students' ability to write explanations when *using* a diagrammatic model versus *drawing* a model. Specifically, our research goals can be stated as (1) How do students' explanations differ when using a diagrammatic model versus when drawing their own? (2) What additional explanatory power do students' drawn models have compared to their written explanation? (3) How are the models students draw contingent on students' content knowledge?

To address these questions, we developed a set of modeling assessment tasks that prompted students with a scenario and asked them to either use a provided model or draw their own model to help them write explanations of phenomena. In addition, students answered several multiple-choice science content knowledge questions and responded to follow-up questions requesting their feedback on the task.

Methodology

Assessment tasks. Three modeling assessment tasks were developed, one targeting each grade band (elementary, middle, and high school). For each task, two versions were created; one in which students were given a model and one in which students were asked to draw their own model. Provided models ranged in their sophistication from relatively simply flow charts (used in the elementary school task) to a complex model using mathematical and molecular representations (used in the high school task). After seeing the provided model or drawing their own model, students were prompted with a series of questions that required them to write explanations. After completing the task, students answered 10 multiple-choice questions assessing their content knowledge of the targeted disciplinary core idea (DCI) of energy. Finally, students were asked for their written feedback about the task.

Rubrics were created to score student explanations based on the task's targeted disciplinary core ideas (chemical energy), and crosscutting concepts (flow of energy and matter, and systems and system models). Rubrics ranged from 0 to 1, 2, or 3 points with the levels of the rubric being based on a learning progression. Students were awarded points based on the level of progression that they demonstrated in their response. For example, students whose explanations included the fifth-grade conception of energy (energy comes from food) may receive one point while students whose explanations included the middle school conception that energy is released through the process of cellular respiration may receive two points.

Scoring of student response was done in three stages. First, students' written explanations were scored with scorers being blind to the specific type of modeling task the student was administered. Second, students who were required to draw a model were given a second explanation score based on scoring both their writing explanation and drawn model holistically. Lastly, to measure how much information students' models communicated drawn models were scored based on their relevance and coherence. The rubric for scoring the models was done independent of the targeted content knowledge and was instead based on the inclusion of relevant elements, connections, and relationships between elements. A description of each level of the model rubric is shown in Table 1.

Table 1: *Summary of Modeling Rubric*

Level	*Description*
0	Student did not draw a model or drew something irrelevant to the task.
1	Student drew a model using relevant elements, but elements have no connections between them.
2	Student drew a model using relevant elements with some connections, but relationships are unclear so that the overall coherence of the model is weak.
3	Student drew a model using relevant elements, connections, and clear relationships so that the overall coherence of the model is strong.

Participants. 1,405 students in the fourth through twelfth grades from across the United States participated. The sample was 16% elementary school, 39% middle school, and 45% high school students with 48% female and 52% male students. A small percentage of the sample (4%)

indicated that English was not their primary language. All students were enrolled in a science class at the time of testing, but not necessarily in a physical science class. Each student was randomly assigned an assessment task resulting in each task being administered to approximately 200 students. Elementary school students were excluded from being assigned the high school task due to it requiring content knowledge that was not grade appropriate (an atom level understanding of chemical reactions).

Rasch Analysis. WINSTEPS software (Linacre, 2016) was used to estimate Rasch student and item measures. The measures are expressed on the same interval scale, are measured in logits, and are mutually independent. The average item difficulty was set to zero logits. Explanation tasks were modeled using the Andrich Rating-Scale Model (Andrich, 1978)..

Statistical Analysis. Students explanation scores were compared using chi-squared tests while Rasch measures were compared using t-tests with Bonferroni corrections.

Results

Comparing the explanations for students who used a provided model versus students who drew their own model. Table 2 shows how many points students received on each question for each version of each task. For all but two questions we found a statistically significant difference ($p<0.01$) between students' explanation scores for the two versions of the tasks. Where significant differences existed, students who were provided a model were more likely to receive higher explanation scores than students who were asked to draw a model. Higher explanation scores were predominantly due to more students receiving two or three points while the number of students who received zero points was similar for the two versions of the tasks. This suggests that providing a model may have helped students who would receive some explanation points receive more points, but may not be as helpful to students who are receiving no points.

We did not find a statistically significant difference in scores for two questions (elementary task question 1 and high school task question 1) suggesting that the provided model may not have been helpful to students when writing their explanations to these questions. Most students received the maximum amount of points on question 1 of the elementary school task, while most students received no points on question 1 of the high school task. In the case of the elementary school task, it may be that students didn't need a model to successfully write an explanation to question 1. In contrast, the model presented in the high school task may not have been helpful to students because it required a level understanding of the chemical energy that most student didn't have and thus they were not able to properly interpret and use

Table 2: *Summary of Explanation Scores for the tasks*

Scores for elementary school task

	Version	0 points	1 point	2 points	3 points	χ2	P
Question 1	Provided a model	21 (12%)	151 (88%)	X	X	0.3	0.66
	Drew a model	28 (14%)	169 (86%)	X	X		
Question 2	Provided a model	33 (19%)	90 (52%)	49 (28%)	X	48.4	<0.001
	Drew a model	27 (14%)	159 (81%)	11 (6%)	X		
Question 3	Provided a model	78 (45%)	35 (20%)	59 (34%)	X	79.1	<0.001
	Drew a model	88 (44%)	104 (53%)	6 (3%)	X		

Scores for middle school task

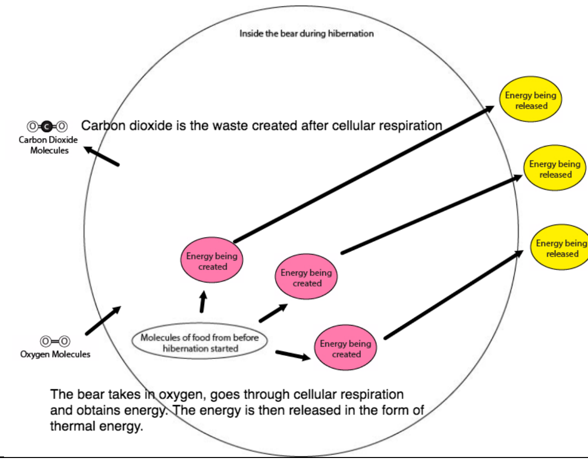
	Version	0 points	1 point	2 points	3 points	χ2	P
Question 1	Provided a model	74 (40%)	87 (47%)	13 (7%)	13 (7%)	11	<0.01
	Drew a model	72 (42%)	91 (53%)	8 (5%)	1 (1%)		
Question 2	Provided a model	58 (31%)	71 (38%)	39 (21%)	19 (10%)	48.4	<0.001
	Drew a model	48 (28%)	60 (35%)	62 (36%)	0 (0%)		

Scores for high school task

	Version	0 points	1 point	2 points	3 points	χ2	P
Question 1	Provided a model	101 (71%)	30 (21%)	11 (8%)	X	1.1	0.33
	Drew a model	70 (65%)	28 (26%)	10 (9%)	X		
Question 2	Provided a model	103 (77%)	22 (16%)	9 (7%)	X	5.9	0.03
	Drew a model	84 (88%)	11 (11%)	1 (1%)	X		

For students who were asked to draw a model, we also compared scores given when scoring their written explanation and when scoring their written explanation and drawn model together. Figure 1 shows an example of a student's written explanation, drawn model, and the scores they received when scoring the written explanation with and without the drawn model. In the example shown, the student received additional points when scoring their written explanation with their model as their written explanation implied "burning" of food for energy while they model showed a more sophisticated process including the use of oxygen and cellular respiration.

Figure 1: *Example written explanation and drawn model*

Written explanation: *The energy that the bear needed to stay alive came from the food it had previous eaten before hibernation. Throughout hibernation, the bear slowly burned the extra body weight it had gained prior to hibernating*	**Drawn Model**: 
Written Explanation Score: 2	**Written explanation + Model Score:** 3

We found that for all but one question there no statistical difference ($p<0.05$) between the scores given when scoring students written explanations with and without their drawn model. While we found individual cases where students models added information or context to their writing (Figure 1), overall, the drawn model did not add significantly to the explanation score they receive when only scoring their written work.

Students' Created Models. Table 3 summarizes the percentage of students who received a specific model score for each task. We found that most students received zero or one point for their models corresponding to either not creating a model or creating a model with no interconnected elements and no clear relationships. This may be due to students being unfamiliar with drawing models as some students commented on "being confused" or "not understanding" what kind of model they were supposed to draw.

Table 3: *Percentage of students who received a specific model score for each task*

	0 points	1 point	2 points	3 points
Description	No Model or drew something irrelevant to the task.	Model using relevant elements, but elements have no connections between them.	Model using relevant elements with some connections, but overall coherence of the model is weak	Model using relevant elements, connections, and clear relationships so that the overall coherence is strong.
Elementary Task	11%	40%	41%	7%
Middle Task	36%	35%	23%	6%
High Task	40%	37%	28%	5%

Students were more likely to score two points on the elementary school task than the middle school and high school tasks. This may be due to the elementary task asking about energy transfer in an ecosystem and the food web model being a relatively common ecosystem model taught to students on how matter and energy transfer in an ecosystem. In addition, the middle and high school tasks required connecting elements across different scales, for example linking an organism to the chemical reactions happening inside the organism, which may make it more difficult for students to include connections in their model. This suggests further work looking at how students map models they are familiar with onto different scenarios and how students think about modeling when the phenomena span different scales.

Rasch Analysis. To further examine students drawn models, we examined whether students with different levels of content knowledge scored differently on their drawn models. To obtain a measure of students' overall content knowledge we fit students' responses to the explanation questions and multiple-choice content questions to a Rasch Model. This produced individual measures for each student representing their general understanding of the topic of chemical energy.

Table 4: *Summary of Rasch Fit Statistics*

	Item			Student		
	Min	Max	Median	Min	Max	Median
Standard error	0.06	0.17	0.08	0.52	1.28	0.61
Infit mean-square	0.84	1.10	1.01	0.01	3.25	0.92
Outfit mean-square	0.78	1.21	0.95	0.05	6.14	0.90
Point-measure correlation	0.17	0.62	0.39			
Separation index (Reliability)	11.40 (0.99)			0.56 (0.24)		

Table 4 shows the fit to the Rasch model. The high reliability (>0.7) and low mean square error values (<1.4) for items indicates that items had a reliable fit to the Rasch mode. Person reliability was found to be low indicating the tasks and multiple-choice items together are not sensitive enough to discriminate between students of different ability. This is to be expected since students took only a single task and ten multiple choice questions. Significant overlap was found between the distributions of the person measures and multiple-choice item difficulties; however, the item difficulties were higher than most students' ability measures, with the average student measure being -0.66 and the average item difficulty 0. The items being relatively difficult for these students also likely contributed to the poor person reliability.

Table 5 compares the student measures of students who were provided a model and students who had to draw a model. We found no statistical difference between students who were provided a model and students who drew a model indicating the both groups had similar understanding of the topic of chemical energy.

Table 5: *Comparison of Rasch measure of student who were provide a model with students who drew a model*

	Model Type		Mean Difference in student ability	SE	t-Statistic	p	Bonferroni Corrected p
Provided vs. drawn	Provided	0	0.24	0.09	2.65	0.009	0.09
	Provided	1	0.1	0.07	1.28	0.203	1
	Provided	2	-0.11	0.07	-1.53	0.128	1
	Provided	3	-0.36	0.14	-2.48	0.018	0.18

Table 6 also compares the student measures for students who received different scores on their drawn models. Students who received three points on their models had higher student measures than students who received zero points (Bonferroni corrected $p<0.001$) and marginally higher measures than students who received one point (Bonferroni corrected $p = 0.05$). In addition, students who received two points for their drawn model had higher student measures than students who received zero points (Bonferroni corrected $p = 0.01$). These results indicate that students who drew models that included relevant elements, connections, and relationships between elements had more content knowledge on the topic of chemical energy than students who may not have drawn a model or drew something irrelevant to the task.

Table 6: *Comparison of Rasch measure of student who were provide a model with students who drew a model*

	Drawn model score		Mean Difference in student ability	SE	t-Statistic	p	Bonferroni Corrected p
Different Drawn Models	0	1	-0.14	0.1	-1.42	0.158	1
	0	2	-0.35	0.1	-3.51	<0.001	0.01
	0	3	-0.6	0.16	-3.74	<.001	<.001
	1	2	-0.21	0.09	-2.4	0.017	0.17
	1	3	-0.45	0.15	-3	0.005	0.05
	2	3	-0.24	0.15	-1.61	0.115	1

Conclusions

Our results show that students who are provided a model write more advanced explanations then students who must draw their own model. While some students commented that the model contained “all the information needed to complete the task,” many others commented that “it didn’t help explain the phenomena” or they didn’t understand it because “it wasn’t detailed enough.” It has been suggested that whether a model is useful to someone in explaining a phenomenon depends on the modeler’s metamodeling knowledge and prior experience and familiarity with the specific representation (Ruppert, Duncan, & Chinn, 2017). While this work suggests providing a student a model may lead to improved explanations relative to having them draw their own model, we did not assess what students understanding of the information communicated by the provided models was or how familiar they were with the model’s representation. Future work in asking students what elements of a model they find helpful in writing their explanation and why those elements are helpful should lead to a better understanding of how to design models that students find to be helpful explanatory tools.

Our results also suggest that when students are asked to draw a model and write an explanation, their models and written explanations evaluated together usually don’t improve their explanation score relative to simply scoring just their written explanation. While some student’s created models illustrated a more sophisticated understanding of the phenomena then their written work, overall this was rare with most students drawn models adding little if anything to their explanation. This seems to indicate that students are much better at communicating their explanation of phenomena in writing than they are drawing a model. It is worth highlighting that our scoring of students' models and explanations together was done largely in a summative way and did not take into account student misconceptions or seek to provide students with feedback. Students’ drawn models have the potential to be powerful tools for diagnosing misconceptions and providing feedback as it allows students to represent their thinking in an alternative format.

Lastly, our results indicate that students who drew the most sophisticated models had a better understanding of energy than students who didn’t draw a model or drew something irrelevant to the task. While these results seem to indicate there is a link between the sophistication of a student’s model and their content knowledge, the nature of this link remains unclear. One possibility is more content knowledge allows students to draw more coherent models, while another possibility is students’ content knowledge and ability to draw models are both linked to a third unmeasured variable such as how serious they took the assessment or general intelligence. Our work highlights the need for additional research and discussion on how students’ knowledge of a domain is related to their ability to create a coherent model using that knowledge.

Implication and Importance. This work is of interest to the NARST research community as it provides insights into the use and creation of models in assessment and curriculum. The distinction between using a model and drawing a model has been given little attention in research literature, and our work suggests that students do write different explanations depending on whether they are using a model or drawing a model. Our work also suggests that if the purpose of the model is to explain a phenomenon, a holistic approach to evaluating students’ models and written explanations may not result in an improved explanation score and that students’ writing is a good indicator of their ability to explain a phenomenon. Lastly, our work highlights the link between content knowledge and the sophistication of a students’ model and calls for further research examining this link.

Acknowledgements. The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A180512 to the Biological Sciences Curriculum Study. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

References

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*. https://doi.org/10.1007/BF02293814

Chinn, C. A., & Samarapungavan, A. (2009). Conceptual change-multiple routes, multiple mechanisms: A commentary on Ohlsson (2009). *Educational Psychologist*. https://doi.org/10.1080/00461520802616291

Fortus, D., Shwartz, Y., & Rosenfeld, S. (2016). High School Students' Meta-Modeling Knowledge. *Research in Science Education*. https://doi.org/10.1007/s11165-015-9480-z

Lehrer, R., & Schauble, L. (2006). Cultivating model-based reasoning in science education. *Cambridge Handbook of the Learning Sciences*. https://doi.org/http://dx.doi.org/10.1017/CBO9780511816833.023

Linacre, J. M. (2016). Winsteps ® Rasch measurement computer program. Beaverton, Oregon. Retrieved from Winsteps.com

Namdar, B., & Shen, J. (2015). Modeling-Oriented Assessment in K-12 Science Education: A synthesis of research from 1980 to 2013 and new directions. *International Journal of Science Education*. https://doi.org/10.1080/09500693.2015.1012185

National Research Council. (2012). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. (C. on a C. F. for N. K.-12 S. E. S. B. on S. E. D. of B. and S. S. and Education, Ed.). Washington DC: The National Academies Press.

NGSS Lead States. (2013). *Next Generation Science Standards: For States, By States*. Washington DC: The National Academies Press.

Project 2016 (American Association for the Advancement of Science). (1993). *Benchmarks for science literacy*. New York: Oxford University Press.

Ruppert, J., Duncan, R. G., & Chinn, C. A. (2017). Disentangling the Role of Domain-Specific Knowledge in Student Modeling. *Research in Science Education*. https://doi.org/10.1007/s11165-017-9656-9

Schwartz, C. V, & White, B. Y. (2005). Metamodeling knoweldge: Developing students' understanding of scientific modeling. *Cognition and Instruction*.

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*. https://doi.org/10.1002/tea.20311