Extending and Investigating the Impact of the High School Model-based Educational Resource

BSCS SCIENCE LEARNING

Cari Herrmann-Abell, Brian Donovan, Emily Harris, Jeffery Snowden, Dennis Lee, Molly Stuhlsatz (Co-PI), & Christopher Wilson (PI)

UC DAVIS

Cynthia Passmore (PI), Hessam Ghanimi, Rich Hedman, Chris Griesemer

Abstract

We are collaborating on a project to examine the efficacy of high school biology instructional materials that support teachers' understanding and practice of model-based reasoning as an approach to support students in developing an integrated, multidimensional understanding of science. This poster summarizes our efforts to develop assessment tasks that measure students' ability to use model-based reasoning (MBR) to make sense of biological phenomena and describes our use of crowdsourced adults to pilot test the tasks.

Project Description

The **Model-Based Educational Resources (MBER)** materials were designed to provide students opportunities to engage in modeling to generate scientific understanding. The curriculum is built around a clear storyline with explicit connections between biological ideas across the full year sequence and within each instructional unit. It provides teachers with pedagogical supports that outline how to engage students in the intellectual work of the classroom. This includes the use of an instructional framework that highlights the connections between the three main components of any modeling lesson: the phenomenon, the question about the phenomenon, and the model that explains the phenomenon. Finally, the curriculum gives teachers the freedom to modify the sequence to accommodate their needs.

The efficacy study seeks to answer the following **research questions**:

1. How do the impacts on student achievement of the MBER program compare to those of Business-as-Usual biology programs?
2. How do the impacts on teacher outcomes of the MBER program compare to those of Business-as-Usual PD biology programs?
3. How much variance in student achievement can be explained by each of the potential mediating (teacher outcomes) and moderating (student demographics) variables (see Figure 1)?

The development of new assessments that will be used as student and teacher outcome measures during the efficacy study is a key part of the project. The goal of the assessment development effort is to create tasks that can be used to measure students' and teachers' ability to use modeling and model-based reasoning to make sense of biological phenomena. As recommended by the NRC (2014), we took a construct-centered approach to assessment development (Mislevy et al., 2003; Wilson, 2005; DeBoer et al., 2008). We also followed the framework put forth by Achieve, Inc. (2019), which requires tasks to focus on real-world phenomena and engage students in sense making while using both disciplinary core ideas and science practices.

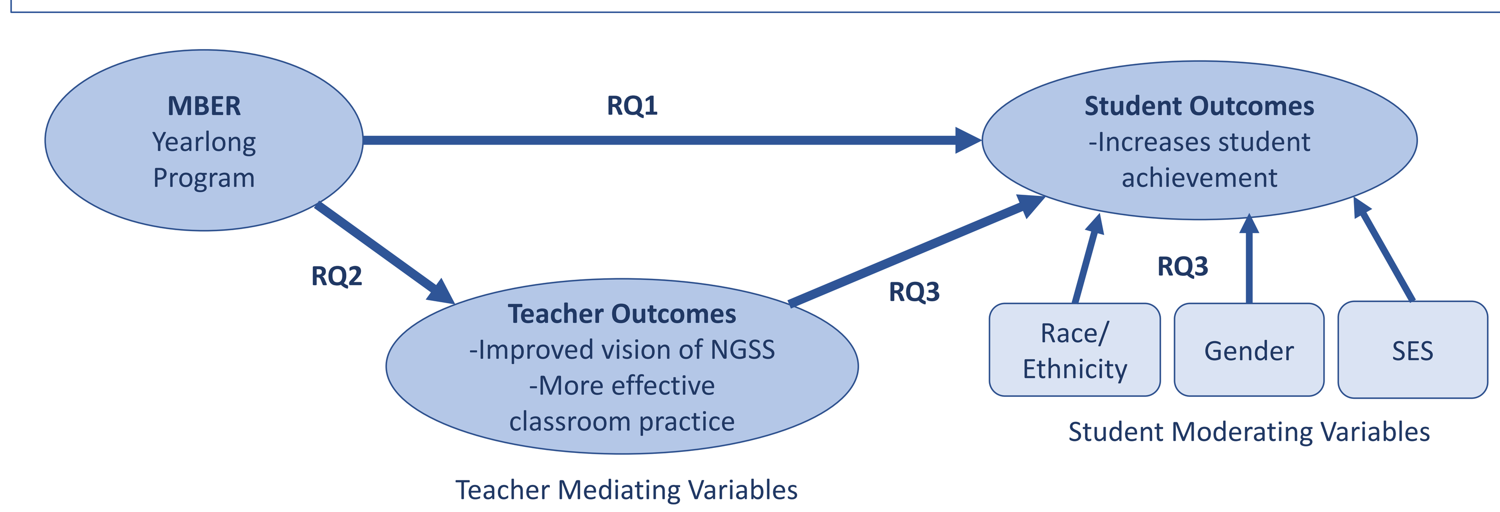

Figure 1. Project theory of action

Task Development

Task template. We developed 4 MBR tasks using the task structure depicted in Figure 2. Each task begins with a description of a real-world phenomenon related to natural selection or genetics. After the phenomenon or scenario is described, a set of data is provided, and the respondent is asked to analyze data and propose an initial model that would explain those data (inductive reasoning). Next, respondents are presented with an alternative model and additional data. They consider the alternative model and data and evaluate whether these new data support the model (abductive reasoning). Finally, respondents are provided with data that contradict the alternative model and asked whether these data support the model. They are then prompted to write a revised model that fits all the data presented in the task (hypothetico-deductive reasoning).

NGSS alignment. While the focus of the assessments was model-based reasoning using biology core ideas (LS3.A & LS4.B), students were required to use aspects of all three NGSS dimensions when completing the tasks. For example, students had to analyze graphical data to identify patterns that could be used as evidence to support a model-based explanation.

Task contexts. Two of the tasks were aligned to assess understanding of natural selection using the contexts of rock pocket mice and lactose tolerance. The other two tasks assess understanding of genetics with one using Tay-Sachs as a context targeting both Mendelian and molecular genetics and the other using polydactyl cats as a context targeting only Mendelian genetics.

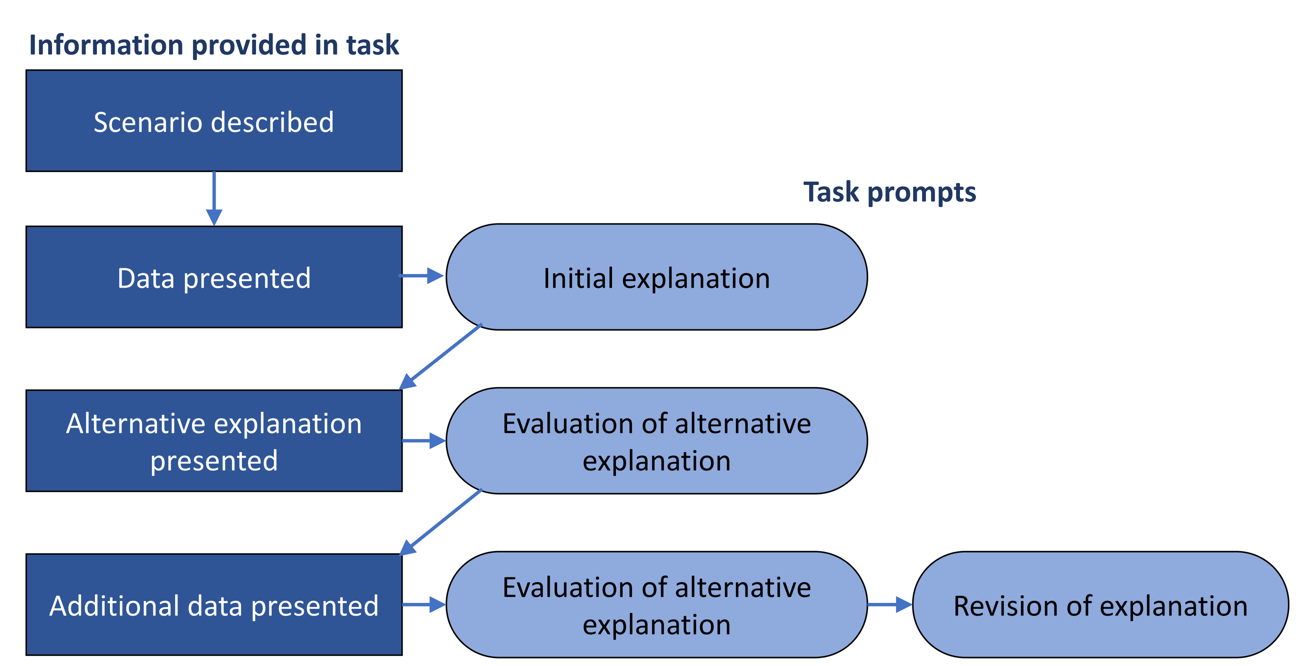

Figure 2. Model-based reasoning task template.

Pilot Test

Four pilot test forms were constructed, each being made up of 2 MBR tasks and 2 sets of content-focused (CF) items. The CF items were selected from the Conceptual Inventory of Natural Selection (CINS) (Anderson, et al., 2002) and the Genetics Literacy Assessment Instrument (GLAI) (Bowling, et al., 2008).

Because COVID-19 made it impossible to pilot test with high school students, we utilized Amazon's Mechanical Turk (MTurk) to collect data from 234 adults with at least a high school diploma from the U.S. Although this population is not an exact match to the intended student population, it provides us with data that can be used to inform item revisions and evaluate scoring rubrics. The population is a closer match to the teacher population and had the potential to give us more correct responses at the upper end of the distribution, which might be difficult to obtain in a high school sample.

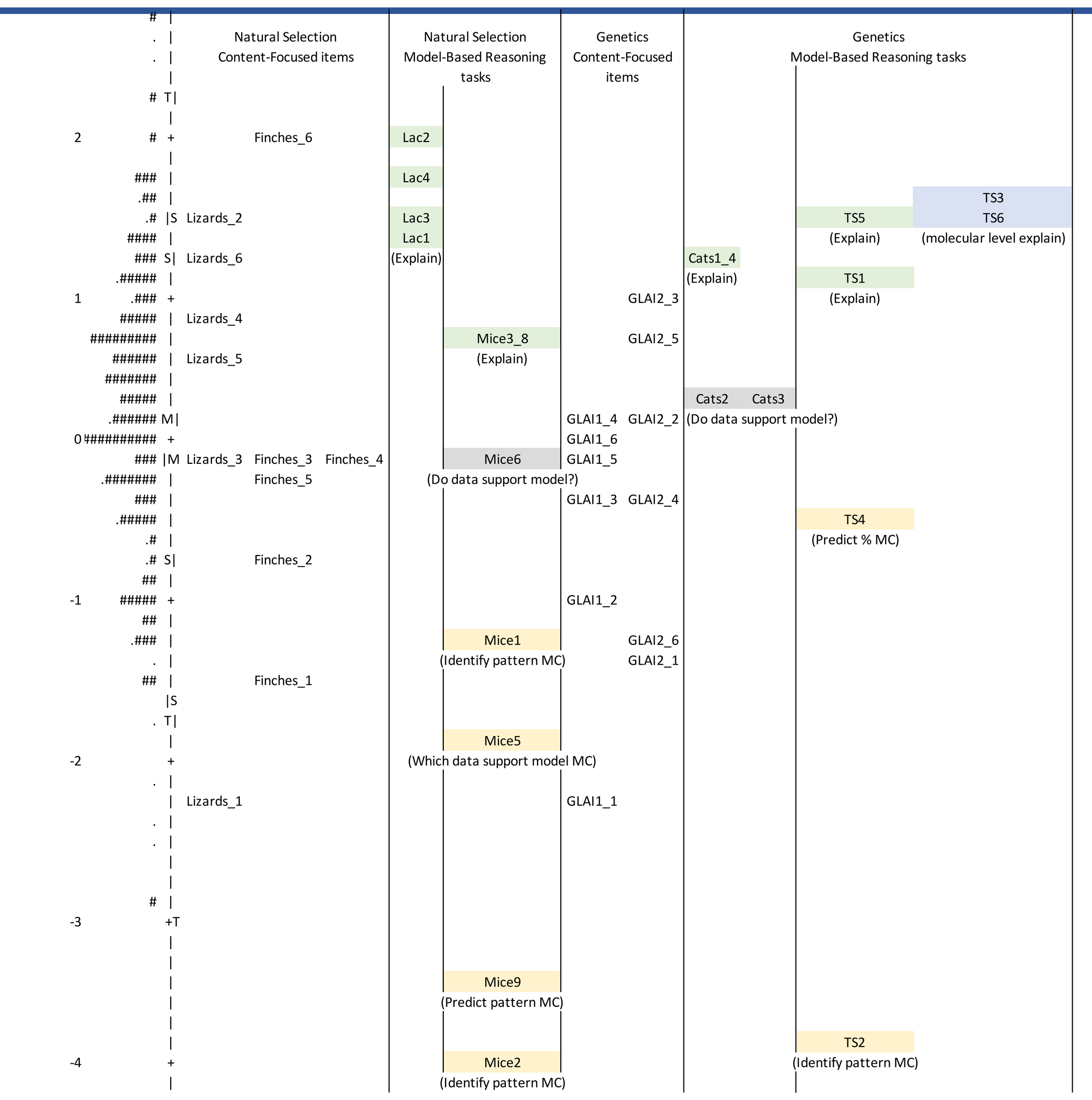

Figure 3. Wright map showing the item difficulties (right side) and person measures (left side).

Results and Discussion

The MTurk data fit the partial credit Rasch model well with person and item reliabilities of .80 and .97, respectively. The average person measure was small and positive (0.32), indicating that the items were well matched to the MTurk workers' ability. As shown in the Wright map in Figure 3, the multiple-choice items within the tasks that ask respondents to predict or identify patterns (shaded yellow) were easier than the constructed-response items (shaded gray, green & blue). The items that asked respondents to evaluate whether data support a proposed model (shaded gray) were easier than the items that required respondents to write a model-based explanation (shaded green and blue).

A principal component analysis of the Rasch residuals showed that the first component of the correlation matrix was 2.5 indicating the MBR tasks and CF items were predominantly unidimensional, although the construct may be broad. Cross-plots plot the person measures based only on the MBR tasks against the person measures based only on the CF items and show a small but statistically significant correlation between performance on the CF items and MBR tasks ($\rho = 0.59$, $p < .001$ for genetics; $\rho = 0.42$, $p < .001$ for natural selection). This suggests that the MBR tasks measure a different aspect of the construct than the CF items, which is to be expected given that the MBR tasks require both content knowledge of biology and model-based reasoning, while the CF items focus solely on content knowledge.

The results suggest that instruments made up of content-focused items and MBR tasks measure the intended outcome and will be appropriate for use with the teacher population. We are currently pilot testing the assessments with a small sample of high school students to ensure that they perform as expected for the targeted student population. We will also be conducting a second MTurk pilot test of these assessments along with the assessments to measure self-efficacy and tolerance to ambiguity. The MTurk respondents will be narrowed to 18-year-old, high school graduates to better mimic the target student population.

References

1. Achieve. (2019). Task annotation project in science. Retrieved from https://www.achieve.org
2. Anderson, D.L., Fisher, K.M., & Norman, G.J. (2002). Development and Evaluation of the Conceptual Inventory of Natural Selection. *Journal of Research in Science Teaching*, 39, 952-978.
3. Bowling, B.V., Acra, E.E., Wang, L., Myers, M.F., Dean, G.E., Markle, G.C., Moskalik, C.L., & Huether, C.A. (2008). Development and Evaluation of a Genetics Literacy Assessment Instrument for Undergraduates. *Genetics*, 178(1), 15-22.
4. DeBoer, G. E., Herrmann-Abell, C. F., Gogos, A., Michiels, A., Regan, T., & Wilson, P. (2008a). Assessment linked to science learning goals: Probing student thinking through assessment. In J. Coffey, R. Douglas, & C. Stearns (Eds.), *Assessing student learning: Perspectives from research and practice* (pp. 231-252). Arlington, VA: NSTA Press.
5. Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2003). *A brief introduction to evidence-centered design* (Research Report 03-16). Princeton, NJ: Educational Testing Service.
6. NRC. (2014). *Developing Assessments for the Next Generation Science Standards*. Washington DC: The National Academies Press.
7. Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Mahwah, NJ: Lawrence Erlbaum Associates.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant 1813538. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.