Optional ERIC Coversheet - Only for Use with U.S. Department of Education Grantee Submissions

This coversheet should be completed by grantees and added to the PDF of your submission if the information required in this form **is not included on the PDF to be submitted**.

INSTRUCTIONS

- Before beginning submission process, download this PDF coversheet if you will need to provide information not on the PDF.
- Fill in all fields—information in this form **must match** the information on the submitted PDF and add missing information.
- Attach completed coversheet to the PDF you will upload to ERIC [use Adobe Acrobat or other program to combine PDF files]—do not upload the coversheet as a separate document.
- Begin completing submission form at https://eric.ed.gov/submit/ and upload the full-text PDF with attached coversheet when indicated. Your full-text PDF will display in ERIC after the 12-month embargo period.

GRANTEE SUBMISSION REQUIRED FIELDS

Title of article, paper, or other content

Assessing students' competencies with mathematical models in virtual science inquiry investigations.

All author name(s) and affiliations on PDF

Last Name, First Name	Academic/Organizational Affiliation	ORCID ID
Adair, Amy		
Sao Pedro, Michael		
Gobert, Janice		
Owens, Jessica		

Publication/Completion Date—(if *In Press* enter year accepted or completed) 2023

Check type of content being submitted and complete one of the following:

- o If article: Name of journal, volume, and issue number if available
- X If paper: Name of conference, date of conference, and place of conference
- o If book chapter: Title of book, page range, publisher name and location
- o If book: Publisher name and location
- o If dissertation: Name of institution, type of degree, and department granting degree

2023 International Conference of the Learning Sciences, Montréal, QC, Canada, June 10-15, 2023

DOI or URL to published work (if available)

Acknowledgement of Funding— Grantees should check with their grant officer for the preferred wording to acknowledge funding. If the grant officer does not have a preference, grantees can use this suggested wording (adjust wording if multiple grants are to be acknowledged). [Fill in Department of Education funding office, grant number, and name of grant recipient institution or organization.]

"This work was supported by **Institute of Education Sciences** U.S. Department of Education, through Grant R305A210432 to **Rutgers University.** The opinions expressed are those of the authors and do not represent views of the **Institute of Education Sciences** or the U.S. Department of Education.

International Society of the Learning Sciences

Assessing Students' Competencies with Mathematical Models in Virtual Science Inquiry Investigations

Amy Adair, Rutgers University, amy.adair@gse.rutgers.edu
Michael Sao Pedro, Apprendis, mikesp@apprendis.com
Janice Gobert, Rutgers University & Apprendis, janice.gobert@gse.rutgers.edu
Jessica A. Owens, Rutgers University, jo503@rutgers.edu

Abstract: Developing models and using mathematics are two key practices in internationally recognized science education standards such as the Next Generation Science Standards (NGSS, 2013). In this paper, we used a virtual performance-based formative assessment to capture students' competencies at both *developing* and *evaluating* mathematical models in science inquiry contexts aligned with the NGSS (2013). Results show that model development and evaluation competencies are correlated, but students who demonstrate proficiency with model development often struggle with evaluation. Nuanced data illustrate how components of modeling competencies differ and how they may be related.

Introduction

Standards such as the Next Generation Science Standards (NGSS, 2013) emphasize that students must become competent at key science practices, including "Developing and Using Models" (Practice 2) and "Using Mathematics and Computational Thinking" (Practice 5). Here, we define *mathematical models* as mathematical representations, such as equations and graphs, that illustrate and predict scientific phenomena (Harrison & Treagust, 2000). The NGSS (2013) expects that students will "develop and/or use [models]…including mathematical models…to generate data that support explanations, predict phenomena, analyze systems, and/or solve problems" and evaluate the limitations of those models because they "contain approximations and assumptions that limit the range of validity and predictive power" (NGSS 2013, Appendix F, p. 6). However, students often struggle in many ways, such as: labeling axes and selecting data for graphical models (Lai et al., 2016), explaining the underlying assumptions and limitations (Nixon, et al., 2016), and ascertaining validity by comparing models to real-world data (Wilensky & Reisman, 2006).

To help students become more proficient at mathematical models, we are iteratively designing virtual labs that require students to build mathematical models as part of their investigative process within the Inquiry Intelligent Tutoring System (Inq-ITS; Dickler et al., 2021; Olsen et al., 2022). Inq-ITS labs are designed to be performance-based formative assessments, using educational data-mined and knowledge-engineered algorithms to automatically assess students' competencies (Gobert et al., 2013; Sao Pedro et al., 2013b). In recent work, we developed algorithms to auto-score how students create graphs and build best bit curves through data.

In this study, we explore what kinds of challenges students have when building mathematical models to better understand what kinds of support students may need to be successful. We expect that students with the highest proficiency can: (1) effectively develop mathematical models of scientific phenomena, (2) articulate the assumptions they made when building their model, (3) articulate how the assumptions impose limitations on how well the model can make valid predictions, and (4) richly integrate understanding of the scientific phenomenon (Schwarz et al., 2009; Windschitl et al., 2008). Given the complexity of mathematical modeling, many students may fall in the "messy middle," where they have "some pieces of knowledge and ability to respond to complex science tasks, but not all" (Gotwals & Songer, 2010, p. 277). To better understand students' competencies with mathematical modeling, we address the following research questions: (1) To what extent are students' model development and evaluation competencies correlated, and (2) When students' model development and evaluation competencies differ, what difficulties do students seem to demonstrate?

Method

Forty-one middle and high school students from four different teachers completed an Inq-ITS momentum virtual lab (NGSS DCI PS2.A). Its goal was to have students develop a mathematical model (i.e., graph and corresponding equation) that illustrates how the mass of a moving car affects its momentum before colliding into a stationary car. During the first three stages of the investigation (Collecting Data, Plotting Data, and Building Models), students collect data using a simulation, select the x- and y-axes for their graph, select which data to plot on their graph, determine the type of mathematical function (i.e., linear, inverse, square, inverse square, or horizontal) that best fits the shape of the plotted data, and select a coefficient and constant for that mathematical function that best fits the data on the graph. Students' scores on these stages were calculated using previously developed algorithms (Dickler et al., 2021). The activity then prompts students to write in their own words a reflection on their model development process. Specifically, students are asked the following question:

> "Your mathematical model makes predictions about the momentum of Car #1 before the collision when you change the mass of Car #1. Do other conditions need to be met in order for your mathematical model to make good predictions? For example, do other variables like the mass of Car #2, or the velocity of Car #1 before collision need to be specific values? Can they be different values? Please explain and provide enough detail so that a friend who did not build your mathematical model could reconstruct it and could understand how to use it."

Measures

We refer to students' performance on the first part of activity as their *model development competencies* because the tasks require students to construct a mathematical model (best-fit curve) from data they collect. We measure these competencies as the sum of fine-grained, sub-practice scores, which were automatically assessed as either 0 (incorrect) or 1 (correct) using educational data-mined and knowledge-engineered algorithms that generate scores based on students' interactions within the Inq-ITS environment (Sao Pedro et al., 2013b; Olsen et al., 2022). We note that performance on these tasks may also indicate competency in other practices as well.

We refer to how students evaluate their models through writing as their *model evaluation competencies* because this task requires students to evaluate the limitations of the mathematical model they developed. Automated scoring was not available for this prompt as this activity is still in pilot testing. Thus, two of the paper's authors independently hand-scored all students' responses across two dimensions: correctness (0 for incorrect or non-answer, 1 for partially correct, 2 for fully correct) and relevance (0 for not relevant, 1 for relevant). Authors agreed 100% for the *relevance* dimension and 82.9% (unweighted Cohen's kappa = 0.70) for the *correctness* dimension. Disagreements were discussed, and the agreed-upon scores were used for analyses.

Results

RQ1: Comparing students' model development and evaluation competencies

We first examined the relationship between students' scores on the two types of competencies using a Pearson's correlation. Students' model development and model evaluation competencies were moderately positively correlated, $r(39) = .58, p < .001$, suggesting that students tend to be proficient (or not) at both competencies together. Though a correlation was performed, it is likely that model evaluation is, at least in part, dependent on students' competencies with model development, as suggested by learning progressions (Schwarz et al., 2009).

Table 1

Raw count and percentages of students (N = 41) with each categorical competency level. Note: The "Medium" (7-8) range for model development had no students and is omitted.

		Model Development Competency Level	
		Low (Scores: 0-6)	High (Scores: 9-10)
Model Evaluation Competency Level	Low (Scores: 0-1)	8 (19.5%)	10 (24.4%) *Messy Middle Group 1*
	Medium (Score: 2)	2 (4.9%) *Messy Middle Group 2*	18 (43.4%) *Messy Middle Group 3*
	High (Score: 3)	0 (0.0%)	3 (7.3%)
	Total	10 (24.4%)	31 (75.6%)

We further disaggregated their performance on each competency into "high," "medium," and "low" categories (Table 1) to examine cases where students performed well at one competency and not another. Most students (75.6%) performed "high" on model development, indicating that these students have the inquiry competencies (e.g., collecting unconfounded data) and mathematical/graphical competencies (e.g., determining the best-fit curve) to complete the activity. However, very few students (7.3%) performed "high" on model evaluation, suggesting that students may not fully understand the importance of collecting and plotting controlled data, or they may struggle with describing how to evaluate models. These findings suggest a need for supports, like embedded real-time scaffolding, that can address these difficulties.

RQ2: Differences between model development and evaluation

As shown in Table 1, there were three groups of students in which competencies did not align. For each group, we triangulated the interaction logs of how students built their models with the written text of how they evaluated their models to identify any commonalities for how students struggle.

Group 1: High model development, low model evaluation

Students in Group 1 (24.4%) were proficient at developing models but stated incorrect and/or irrelevant responses when evaluating their model. For example, Student A, who had no errors when developing their model, stated, "no nothing else needs to be added because its [*sic*] very accurate." Another student, Student B, with no errors simply responded, "The line best fits the dots." These responses suggest that although these students can execute the procedures of developing mathematical models, they may lack conceptual understanding about how that process embeds assumptions and limitations about their model. For example, by stating that their models are "very accurate" and "fits the dots," they may believe that their model does not have any limitations. They may not recognize that their model will only predict well when the other variables for which they controlled have the exact same measurements. Such students may require supports that highlight the importance of the variables they controlled and how those controls impact the general applicability of their models.

Group 2: Low model development, medium model evaluation

Only two students (4.9%) demonstrated poor model development competencies. Student C plotted uncontrolled trials, and Student D selected the incorrect variable for x-axis. Both students did not recognize their errors when looking at the graph, which could suggest that they may be struggling with interpretation (Glazer, 2011). Therefore, students in this group may require more targeted support on specific sub-practices (e.g., plotting controlled trials for Student C and choosing the axes for the graph for Student D).

Group 3: High model development, medium model evaluation

Students in Group 3 (43.4%) were proficient at developing models but gave only partially correct answers when evaluating their models. One consistent error made by these students was stating that the "mass of the stationary car" needed to be a certain value for the model to be used to make predictions. However, students should recognize, either through prior scientific content knowledge or the virtual lab investigation, that only the *velocity of the moving car* needs to remain constant since this is the only other simulation variable that would affect the moving car's momentum *before* the collision with the stationary car. This finding aligns with previous research showing the need to support students in defining the boundaries of their models (Eidin et al., 2020), which may help students to develop a deeper understanding of the scientific phenomenon being modeled (Wilensky & Reisman, 2006; Windschitl et al., 2008), as envisioned by the NGSS (2013).

Discussion

Developing and evaluating mathematical models in science inquiry contexts is important for science learning (NGSS, 2013). However, teachers rarely provide students with the opportunity to construct and evaluate their *own* models (Schwarz et al., 2009). Furthermore, assessing these competencies is challenging (Furtak et al., 2017). Here, we presented a novel design for Inq-ITS virtual labs to formatively assess students' competencies on fine-grained components of NGSS-aligned mathematical modeling practices. By unpacking NGSS Practice 2 (Developing and Using Models) and Practice 5 (Using Mathematics and Computational Thinking), assessments can provide important information on students' modeling competencies. We used the virtual labs to investigate how students develop and evaluate mathematical models, and what specific difficulties they encounter when doing so. Consistent with prior work on constructing and critiquing graphs (Vitale et al., 2015), we found that students' model development and evaluation competencies are correlated, but that many students who performed well on model development still struggled with model evaluation. More specifically, students in the "messy middle" who performed well on model development tended to struggle with understanding how the development process affects the limitations of the model (e.g., Group 1) or how the boundaries of the system can also affect the limitations of the model (e.g., Group 3). Our findings also shed light on the importance of *fine-grained* assessments that can specifically target the sub-components of modeling practices. Namely, we found that students in Group 2 had different difficulties with the model development stages, which teachers would need to address differently when providing support. These findings suggest that students' difficulties with facets of modeling are likely varied. As such, it is vital that formative assessments are operationalized at a fine-grained enough level to identify exactly how students are struggling so that optimum support may be provided.

Overall, the design work presented in this paper sheds light on how online formative, performance-based assessments can be used to capture rigorous, rich, and nuanced data on students' competencies. With this data, systems, like Inq-ITS, can be further augmented to provide targeted, real-time support to students on the specific sub-practice for which they are struggling as well as to teachers on how best to support their students. Furthermore, this work contributes to and deepens the existing literature on general scientific modeling (e.g., Schwarz et al., 2009; Windschitl et al., 2008) by drawing a deeper attention to *mathematical* modeling done in the context of science inquiry, important to NGSS (2013) Practices 2 and 5. By developing scalable formative assessments that address students' competencies with both *developing* and *evaluating* mathematical models in science inquiry contexts, we can begin to address inter-dependencies between these aspects of modeling to develop a learning progression that includes components of scientific mathematical modeling competencies. Furthermore, unlike previous work on science graphing (e.g., Vitale et al., 2015), we assess students' competencies with evaluating models they developed themselves, capturing a more authentic, integrated version of the science practice envisioned by the NGSS (2013). Although the virtual lab in this study focused on only one physical science topic (i.e., momentum) and one type of mathematical relationship (i.e., linear), future work will investigate mathematical modeling across a broader range of science topics.

Acknowledgements

This material is based upon work supported by the National Science Foundation GRFP under Grant No. DGE-1842213 (Amy Adair) and the Institute of Education Sciences, U.S. Department of Education (Grant No. R305A210432 & 91990019C0037; Janice Gobert & Michael Sao Pedro). Any opinions, findings, and conclusions expressed in this material are those of the authors and do not necessarily reflect the views of the NSF or the IES.

References

Dickler, R., Sao Pedro, M., Adair, A., Gobert, J., Olsen, J., Kleban, J., Betts, C., Staudenraus, C., & Roughan, P. (2021). Supporting students remotely: Integrating mathematics and sciences in virtual labs. In, *15th International Conference of the Learning Sciences* (pp.1013-1014). ISLS.

Eidin, E., Bielik, T., Touitou, I., Bowers, J., McIntyre, C., & Damelin, D. (2020). Characterizing advantages and challenges for students engaging in computational thinking and systems thinking through model construction. In, *14th International Conference of the Learning Sciences, Volume 1* (pp. 183-190). ISLS.

Furtak, E. M. (2017). Confronting dilemmas posed by three-dimensional classroom assessment: Introduction to a virtual issue of science education. *Science Education, 101*(5), 854–867.

Glazer, N. (2011). Challenges with graph interpretation: A review of the literature. *Stud Sci Educ, 47*(2), 183-210.

Gobert, J., Sao Pedro, M., Raziuddin, J., & Baker, R. (2013). From log files to assessment metrics: Measuring students' science inquiry skills using educational data mining. *J Learn Sci, 22*(4), 521-563.

Gotwals, A. W., & Songer, N. B. (2010). Reasoning up and down a food chain: Using an assessment framework to investigate students' middle knowledge. *Science Education, 94*(2), 259-281.

Harrison, A. G., & Treagust, D. F. (2000). A typology of school science models. *Int J Sci Educ, 22*(9), 1011-1026.

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.

NGSS Lead States. (2013). *Next generation science standards: For states, by states.* National Academies Press.

Nixon, R. S., Godfrey, T. J., Mayhew, N. T., & Wiegert, C. C. (2016). Undergraduate student construction and interpretation of graphs in physics lab activities. *Physical Review Physics Education Research, 12*(1).

Olsen, J., Adair, A., Gobert, J., Sao Pedro, M., & O'Brien, M. (2022). Using log data to validate performance assessments of mathematical modeling practices. In, *International Conference on Artificial Intelligence in Education* (pp. 488-491). Springer, Cham.

Sao Pedro, M., Baker, R., Gobert, J., Montalvo, O., & Nakama, A. (2013b). Leveraging machine-learned detectors of systematic inquiry behavior to estimate and predict transfer of inquiry skill. *User Modeling and User-Adapted Interaction*, *23*, 1-39.

Schwarz, C. V., Reiser, B. J., Davis, E. A., Kenyon, L., Achér, A., Fortus, D., Schwartz, Y., Hug, B., & Krajcik, J. (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.

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., & Reisman, K. (2006). Thinking like a wolf, a sheep, or a firefly: Learning biology through constructing and testing computational theories—an embodied modeling approach. *Cognition and Instruction, 24*(2), 171-209.

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