# 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**

Real-time AI-driven assessment & scaffolding that improves students' mathematical modeling during science inqu

**All author name(s) and affiliations on PDF. If more than 6 names, ERIC will complete the list from the submitted PDF.**



| Last Name, First Name | Academic/Organizational Affiliation | ORCID ID |
|---|---|---|
| Adair, Amy | | |
| Segan, Ellie | | |
| Gobert, Janice | | |
| Sao Pedro, Michael | | |
| | | |
| | | |



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

**Check type of content being submitted and complete one of the following in the box below:**

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

2023 International Conference on Artificial Intelligence in Education (AIEd), Tokyo, Japan, July 3-7, 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 U.S. Department of Education **[Office name]** Institute of Education Scier through **[Grant number]** R305A21043 to **Institution]** Rutgers Univer. The opinions expressed are those of the authors and do not represent views of the **[Office name]** Institute of Education Scie or the U.S. Department of Education.

# Real-time AI-Driven Assessment & Scaffolding that Improves Students' Mathematical Modeling during Science Inquiry

Authors[1,2]

[1] Affiliation 1
[2] Affiliation 2

**Abstract.** Developing models and using mathematics are two key practices in internationally recognized science education standards, such as the Next Generation Science Standards (NGSS) [1]. However, students often struggle with these two intersecting practices, particularly when developing mathematical models about scientific phenomena. Formative performance-based assessments designed to elicit fine-grained data about students' competencies on these practices can be leveraged to develop embedded AI scaffolds to support students' learning. In this paper, we present the design and initial classroom test of virtual labs that automatically assess fine-grained sub-components of students' mathematical modeling competencies based on their actions within the learning environment. We describe how we leveraged underlying machine-learned and knowledge-engineered algorithms to trigger scaffolds, delivered proactively by a pedagogical agent, that address students' individual difficulties as they work. Results show that the students who received automated scaffolds for a given practice on their first virtual lab *improved* on that practice for the next virtual lab on the same science topic in a different scenario (a near-transfer task). These findings suggest that real-time automated scaffolds based on fine-grained assessment can foster students' mathematical modeling competencies related to the NGSS.

**Keywords:** Scaffolding, Intelligent Tutoring System, Science Inquiry, Mathematical Modeling, Virtual Lab.

## 1 Introduction

To deepen students' understanding of scientific phenomena and ensure that students are fully prepared for future careers related to science and mathematics [2], students must become proficient at key science inquiry practices, such as those outlined in the Next Generation Science Standards [1]. However, the difficulties that students experience with practices, particularly NGSS Practice 2 (Developing and Using Models) and Practice 5 (Using Mathematics and Computational Thinking), can be barriers for students' access and success in high school science coursework and future STEM careers [3, 4, 5]. Specifically, students often struggle to develop mathematical models (i.e., graphs) with quantitative data in science inquiry contexts [6] because they cannot properly label the axes of their graphs [7], interpret the variables on the graph [8], make connections between equations and graphs [4], or choose the functional relationship to create a best-fit line or curve [9, 10]. To help students develop and hone these critical competencies so that they can transfer them across science contexts [11], students need resources and tools capable of formatively assessing their competencies in a rigorous, fine-grained way, and in turn, enabling automated, targeted scaffolding that supports their learning while they work [12, 13].

In this paper, we evaluate the design of a virtual labs instrumented to automatically assess and scaffold students' competencies as they conduct investigations and develop mathematical models to explain science phenomena [14, 15]. To do so, we address the following research question: Did the real-time, individualized scaffolding triggered by automated assessment algorithms help students improve on their mathematical modeling competencies from the first virtual lab activity to a second virtual lab activity on the same topic in a different scenario (i.e., a near-transfer task)?

### 1.1 Related Work

Some online environments seek to assess and support students' competencies related to mathematical modeling for science, such as constructing and exploring computational models (e.g., Dragoon, [16]), drawing qualitative graphs of science phenomenon (e.g., WISE, [17]), and physics problem solving (e.g., Andes, [18]). However, these environments do not assess students' mathematical modeling competencies within the context of a full science inquiry investigation, nor do they provide AI-driven real-time scaffolding on the full suite of other NGSS practices (e.g., Planning and Conducting Investigations), all of which are needed for conducting a full authentic inquiry investigation that uses mathematical models to make inferences about science phenomena.

Nonetheless, scaffolding in online learning environments for both math and science has yielded student improvement on competencies by breaking down challenging tasks into smaller ones [19], providing hints on what to do next for students who are stuck on a task [20], and reminding students about the progress and steps taken thus far [19]. While scaffolding strategies have been applied to the online learning environments for modeling in science [21, 18], there are no studies, to our knowledge, that investigate the efficacy of AI-driven scaffolds for *mathematical modeling* (i.e., graphing) *in the context of science inquiry,* as envisioned by the practices outlined in the NGSS. Thus, the goal of the current study is to evaluate the use of real-time automated scaffolding in an intelligent tutoring system environment, [ITS], to improve students' competencies on science inquiry and mathematical modeling practices outlined in science and 21$^{st}$ century standards such as the NGSS.

# 2 Methods

## 2.1 Participants and Materials

Participants included 70 students across four eighth grade science classes taught by the same teacher from the same school in the northeastern region of the United States during the Fall 2022 semester. In terms of demographics of the participating school population, 31% of students qualify for free or reduced-price lunch; 71% identify as White, 16% as Hispanic, and 6% as two or more races.

For this study, all students completed two [ITS] mathematical modeling virtual labs on the disciplinary core idea of Forces and Motion (NGSS DCI PS2.A) that were augmented with automated scaffolding. In these labs, students used simulations to collect data and develop mathematical models to demonstrate the relationship between the roughness/friction of a surface and the acceleration of a moving object on that surface. The scenario of the first lab was a truck driving along a road, and the scenario of the second lab, which students completed about a week later, involved a sled sliding down a ramp (Fig. 2). [ITS] automatically assessed students' competencies using previously validated educational data-mined and knowledge-engineered algorithms [14, 15], which triggered the individualized scaffolds to support students on their science inquiry and mathematical modeling competencies.

## 2.2 Scaffolded [ITS] Virtual Labs with Mathematical Modeling

The virtual labs consisted of six stages that structured the investigation and captured different aspects of students' competencies associated with NGSS practices (Table 1, Fig. 1). The goal of each activity was to develop a mathematical model (a graph and corresponding equation) that can explain how changing one factor (e.g., roughness of a ramp/road) impacts an outcome (e.g., acceleration of a sled sliding down that ramp, or acceleration of the truck on the road). Descriptions of each stage and how each stage aligned to NGSS practices are shown in Table 1.

We consider the tasks presented in both labs as isomorphic, near-transfer tasks [22, 23], since they consisted of the same stages and focused on the same physical science concept (i.e., the relationship between friction/roughness of a surface and acceleration of a moving object on that surface). However, the scenarios depicted in the simulations differed. In the first lab ("Truck"), students investigated the mathematical relationship between the roughness/friction of a *flat road* and the acceleration of the *truck* on that road (Fig. 2, left). In the second lab ("Ramp"), students investigated the mathematical relationship between the roughness/friction of a *ramp* and the ending acceleration of a *sled* sliding down the ramp (Fig. 2, right). In both cases, the students learn that, when they only change the roughness/friction of the surface (i.e., road/ramp) and keep all other variables constant, there is a negative linear relationship between the friction of the surface and the acceleration of the object moving along that surface.

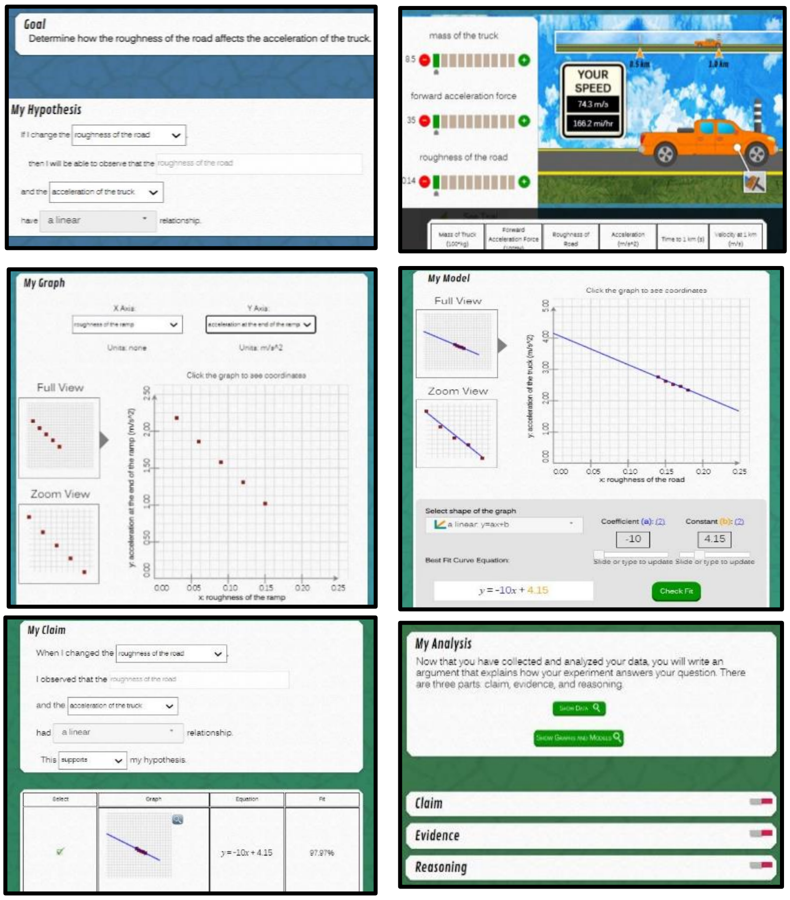


**Fig. 1.** Screenshots of [ITS] mathematical modeling virtual lab; Stages include (1) Hypothesizing (top left), (2) Collecting Data (top right), (3) Plotting Data (middle left), (4) Building Models (middle right), (5) Analyzing Data (bottom left), (6) Communicating Findings (bottom right).


**Fig. 2.** The simulation in the Collecting Data stage of the Truck lab (left) and Ramp lab (right).

**Table 1.** Stages of the [ITS] Mathematical Modeling Virtual Lab Activity

| Stage | Related NGSS Practice(s) | Description of Stage |
|---|---|---|
| Stage 1: Hypothesizing | Practice 1: Asking Questions & Defining Problems | Students form a question about the mathematical relationship between an independent and dependent variable based on a given goal (e.g., If I change the roughness of the ramp, then I will be able to observe that the roughness of the ramp and the acceleration of the sled at the end of the ramp have a linear relationship). |
| Stage 2: Collecting Data | Practice 3: Planning & Carrying Out Investigations | Students collect data using a simulation that can be used to investigate the relationship between the variables outlined in their hypothesis (e.g., roughness of the ramp and acceleration of the sled at the end of the ramp). The data that they collect are automatically stored in a data table. |
| Stage 3: Plotting Data | Practice 2: Developing & Using Models Practice 5: Using Mathematics & Computational Thinking | Students select trials from their data table to plot on a graph. Students select the variable to place on the x-axis of their graph and the y-axis of their graph. Ideally, students should place their independent variable (e.g., roughness of the ramp) on the x-axis and their dependent variable (e.g., acceleration of the sled at the end of the ramp) on the y-axis, and students should only plot *controlled* data. |
| Stage 4: Building Models | Practice 2: Developing & Using Models Practice 5: Using Mathematics & Computational Thinking | Students select the type of mathematical relationship that best fits the shape of the plotted data (e.g., linear, inverse, square, inverse square, or horizontal). Students also determine the coefficient and constant for the equation of the best-fit curve/line as well as check the fit (i.e., coefficient of determination, $R^2$), which is automatically calculated and stored in their table along with a snapshot of their graph and the equation that they built. Ideally, students should create a model that fits the data points and demonstrates the mathematical relationship between the two variables. Students are *not* expected to calculate the coefficient and constants for the equation of their model, but rather they are expected to use the slider to create an “informal” line/curve of best fit. |
| Stage 5: Analyzing Data | Practice 4: Analyzing & Interpreting Data | Students interpret the results of their graphs by making a claim about the relationship between the variables, identifying if it was the relationship that they had initially hypothesized, and selecting the graphs and corresponding equations that best demonstrated this relationship. |
| Stage 6: Communicating Findings | Practice 6: Constructing Explanations | Students write an explanation of their findings in the claim, evidence, and reasoning (CER) format. |

The design of the lab foregrounds students' competencies with collecting controlled data [24], plotting/graphing the data [25], and determining the informal line/curve of best fit [9] *without* deriving the algebraic equations, as is typical in physics contexts [26]. This design not only helps students more readily identify the similarities in the mathematical and scientific relationship between the variables in the two scenarios (i.e., the friction of the road/ramp vs. the acceleration of the truck/sled), but also helps students develop more sophisticated understandings of the scientific meaning in the graphs, a task with which students often struggle [27].

### 2.3 Automated Assessment and Scaffolding of Science Practices

[ITS] automatically assesses and scaffolds their competencies on fine-grained components, or "sub-practices," of the related NGSS practices elicited in each stage of the lab activity (Table 2). For this study, the automated scoring algorithms were active for the first four stages of the lab (Hypothesizing, Collecting Data, Plotting Data, and Building Models); automated scoring algorithms for the other stages are in development and thus out of scope of this study.

Assessment and scaffolding are executed as follows. Each sub-practice is automatically scored as either 0 (incorrect) or 1 (correct) using previously validated educational data-mined and knowledge-engineered algorithms [14]. The algorithms take as input the work products created by the student (e.g., their graphs or mathematical models), and/or distilled features that summarize the steps they followed (e.g., the processes they used to collect data) [14, 15]. If the student completes the task correctly (i.e., receives 1 for all sub-practices), they can proceed to the next stage. If not, individualized scaffolding is automatically triggered based on the sub-practices on which the student was correct or incorrect, and they are prevented from moving forward to the next stage. This proactive approach was chosen because students often cannot recognize when to ask for help [28] and because making errors on earlier stages make subsequent stages fruitless to complete (e.g., it does not make sense to graph data that are completely confounded) [15]. This approach has shown to be effective in helping students learn and transfer other science inquiry competencies [14] even after many months [29]; however, to date, we had not tested this approach with the mathematical modeling competencies described in this study.

The automated scaffolding appears as an on-screen pop-up message delivered from a pedagogical agent, [AGENT NAME]. [AGENT NAME] scaffolding messages are specifically designed to orient and support students on the sub-practice for which they are struggling, explain how the sub-practice should be completed, and elaborate on why the sub-practice is completed in that way [28, 30]. Students also have the option to ask further predefined questions to the agent to receive definitions for key terms and further elaborations on how to complete the sub-practice. If students continue to struggle, the student will eventually receive a bottom-out hint [28, 30] stating the exact actions they should take within the system to move forward in the activity. If the student needs support on multiple sub-practices, the scaffolds are provided in the priority order that was determined through discussions with domain experts and teachers familiar with the task and the [ITS] system. For example, if a student is struggling with both the "Good Form" and “Good Fit” sub-practices for the “Building Models” stage (Table 2), the student will receive scaffolding on the “Good Form” sub-practice *first* since the student must be able to identify the shape of the data before fitting the model to the data.

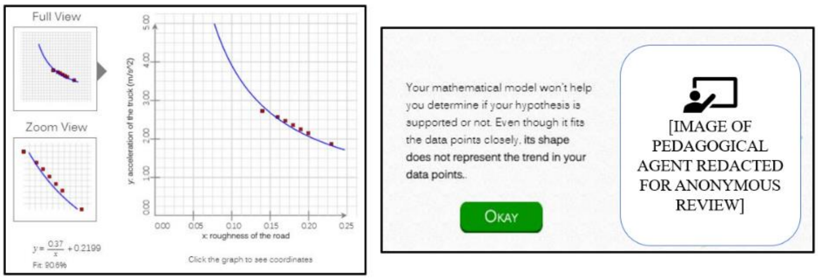


**Fig. 3.** Example screenshot of a model built by a student who is struggling with “Math Model has Good Form” sub-practice, but not with “Math Model has Good Fit” sub-practice (left; note: the student selected an *inverse* relationship when they should have chosen a *linear* relationship, given the variables on their graph), and the first scaffold the student would receive to remediate this difficulty (right).

To illustrate, consider a student who is struggling with choosing the mathematical function that best demonstrates the relationship between the variables while building mathematical models, a common difficulty for students [9, 10, 15]. In this case, the student creates a mathematical model that appears to fit the data points plotted on the graph, but the function chosen for the model does *not* best represent the shape of the data in the graph (Fig. 3, left). When the student chooses to move on to next stage, the [ITS] assessment algorithms use features of the student’s mathematical model, including the shape of the mathematical model (e.g. linear, square), the numerical values chosen for their coefficients and constants, and their fit scores, to determine that the student built a mathematical model with a “good fit” but *not* a “good form” (see Table 2 for sub-practice criteria). [AGENT NAME] then provides feedback to help the student ensure their model has the correct functional form expected between the variables selected for their graph. In this example, the first scaffold the student receives from [AGENT NAME] states, “Your mathematical models won't help you determine if your hypothesis is supported or not. Even though one of your models, model number 1, fits the data points closely, its shape does not represent the trend in your data points.” (Fig. 3, right). If the student continues to struggle on this sub-practice, the student will receive the next level of scaffold (a procedural hint), stating “Let me help you some more. Look at what kind of shape your data points make. Then, when you select the shape of the graph, choose the option that looks most like the shape your data points make.” If the student continues to struggle after receiving the first two scaffolds, the student will receive a bottom-out hint stating, “Let me help you some more. The shape of your data looks most like linear.” As illustrated, scaffolds are designed to support students in building their mathematical modeling competencies by focusing on the *fine-grained* sub-practice with which the student is struggling in that moment.

**Table 2.** Operationalization of Automatically Scored Sub-Practices in the [ITS] Virtual Lab

| Stage | Sub-Practice | Criteria |
|---|---|---|
| Stage 1: Hypothe-sizing | Hypothesis IV | A potential independent variable (IV; a variable that can be *changed* by the experimenter) was chosen as the IV in the hypothesis drop-down menu. |
| | Hypothesis IV Goal-Aligned | The *goal-aligned* IV (the IV from the investigation goal) was chosen as the IV in the hypothesis drop-down menu. |
| | Hypothesis DV | A potential dependent variable (DV; a variable that will be *measured* by the experimenter) was chosen as the DV in their hypothesis drop-down menu. |
| | Hypothesis DV Goal-Aligned | The *goal-aligned* DV (the DV from the investigation goal) was chosen as the DV in the hypothesis drop-down menu. |
| Stage 2: Collecting Data | Data Collection Tests Hypothesis | The student collected controlled data that can be used to develop a mathematical model demonstrating the relationship between the IVs and DVs stated in the investigation goal. Detected by EDM algorithm [14]. |
| | Data Collection is Controlled Ex-periment | The student collected controlled data that can be used to develop a mathematical model demonstrating the relationship between *any* of the changeable variables and the DV stated in the investigation goal. Detected by EDM algorithm [14]. |
| | Data Collection has Pairwise-IV CVS | The student collected at least two trials, where only the goal-aligned IV changes and all other variables are held constant (i.e., controlled variable strategy; CVS). |
| Stage 3: Plotting Data | Graph's X-Axis is an IV & Y-Axis is a DV | Using the drop-down menus, the student selected one of the potential IVs for the x-axis of their graph and one of the potential DVs for the y-axis of their graph. |
| | Axes of Graph are Goal Aligned | Using the drop-down menus, the student selected the *goal-aligned* IV for x-axis and the *goal-aligned* DV for y-axis. |
| | Axes of Graph are Hypothesis Aligned | The student selected the *hypothesis-aligned* IV (i.e., the IV that the student chose in hypothesis) as the x-axis of their graph and the *hypothesis-aligned* DV (i.e., the DV that the student chose in hypothesis) as the y-axis of their graph. |
| | Graph Plotted Controlled Data | The student only plotted controlled data with respect to the variable chosen for the x-axis. |
| | Graph Plotted Minimum for Trend | The student plotted at least 5 controlled data points with respect to the variable chosen for the x-axis. |
| Stage 4: Building Models | Math Model has Good Form | The student built a model with the correct mathematical relationship, based on the variables selected for the graph's axes. |
| | Math Model has Good Fit | The student built a model that fits the plotted data with at least 70% fit. |
| | Math Model has Good Fit *and* Form | The student built a model that *both* has the correct mathematical relationship based on the variables selected for the axes of the graph *and* fits the plotted data with at least 70% fit. If the student has one model with good fit but not good form and another model with good form but not good fit, the student does not get credit. |

### 2.4 Measures & Analyses

To measure students' competencies with the practices associated with each stage, students' scores for each stage are calculated as the average of the sub-practice scores for that stage (Table 2) before scaffolding was received (if any), as has been done in previous studies [29]. Because students may receive multiple scaffolds addressing different sub-practices on a single stage and the effect of those scaffolds may be entangled, we use the measures of students' overall competencies at the stage level, rather than the sub-practice level, for the analyses of this study.

To determine the impact of the real-time AI-driven scaffolding, we analyzed how the scaffolded students' competencies from the first virtual lab activity ("Truck") to the second virtual lab activity ("Ramp"). We note that students who received scaffolding on one stage (e.g., Collecting Data) did not necessarily receive scaffolding on another stage (e.g., Plotting Data). As such, we examine students' competencies on each stage separately to determine students' improvement on the competency for which they were helped. Thus, because we are interested in isolating if each type of scaffolding improved students' performance on the respective competency, we ran four one-tailed, paired samples t-tests with a Bonferroni correction [31] (i.e., one for each competency to account for the chance of false-positive results when running the multiple t-tests).

We recognize that our analytical approach does not account for the effects of scaffolding on one competency possibly leading to improvements on other competencies (despite the student only having received scaffolding on one of the competencies). For example, a student may receive scaffolding on plotting controlled data on the Plotting Data stage, which in turn potentially impacts their performance with fitting the mathematical model to the plotted data on the subsequent Building Models stage [15]. However, unpacking the correlation between competencies and how the scaffolding can affect performance on multiple competencies is outside the scope of this study.

## 3 Results

We found that, for all four stages, the scaffolded students' competencies *increased* from the first lab ("Truck") to the second ("Ramp"; Fig. 2). With a Bonferroni corrected alpha (0.05/4 = .0125), the differences were significant for all four of the competencies (i.e., Hypothesizing, Collecting Data, Plotting Data, and Building Models; Table 3). Further, the effect sizes (Cohen's *d*) were moderately large, especially considering that most students completed the second activity at least one week after completing the first [32]. These findings suggest that scaffolding was effective for all four scientific inquiry and mathematical modeling practices assessed in the [ITS] virtual lab.

**Table 3.** Average inquiry practice scores across activities and results of paired samples t-tests

| Stage | # of scaffolded students | Lab 1: Truck M (SD) | Lab 2: Ramp M (SD) | Within-Subjects Effects |
|---|---|---|---|---|
| Hypothesizing | 27 | .41 (.30) | .74 (.27) | $t(26) = -5.45$, $p < .001$, $d = 1.05$ |
| Collecting Data | 37 | .57 (.22) | .82 (.28) | $t(36) = -4.72$, $p < .001$, $d = .78$ |
| Plotting Data | 24 | .63 (.21) | .82 (.27) | $t(23) = -2.68$, $p = .007$, $d = .55$ |
| Building Models | 31 | .24 (.20) | .52 (.44) | $t(30) = -3.22$, $p = .002$, $d = .58$ |

## 4 Discussion

Students' competencies with mathematical modeling practices during science inquiry are critical for deep science learning [1,2], and middle and high school science courses are a key opportunity for remediation so that students are better prepared for future STEM courses and careers [3, 4, 5]. However, as previously discussed in the introduction, students have difficulties with many mathematical modeling competencies crucial to science [6] including those of focus in this study (e.g., identifying the functional form in plotted data [9, 10]), and when students struggle with constructing and interpreting graphs in mathematics, it hampers their ability to transfer those competencies to science contexts [4, 33]. Further, even though these mathematical modeling competencies are necessary for developing deep understanding of science phenomena [6, 21, 27], they are not often addressed in science classrooms [7]. Thus, there is a need for formative assessment resources that provide targeted support on the specific components for which they are experiencing difficulties in real time, when it is optimal for learning [28] so that they can in turn, transfer these competencies to future tasks [11, 23, 29].

In this study, we found that students who received AI-driven real-time scaffolds for the mathematical modeling competencies during a virtual lab investigation improved on those competencies when completing a near-transfer (i.e., isomorphic; [22]) task on the same physical science topic in a different scenario. These results suggest that the scaffolds that address the sub-practices associated with each of the four stages in the virtual lab (i.e., Hypothesizing, Collecting Data, Plotting Data, and Building Models) are beneficial for students' transfer of their competencies with the mathematical modeling practices important to science inquiry. We speculate that this improvement in competencies occurred because the [ITS] assessment approach involved operationalizing the mathematical modeling practices (e.g., NGSS Practices 2 & 5) into fine-grained sub-practices upon which the targeted scaffolds are based. This approach to unpacking the NGSS practices, which are often considered underspecified for assessment purposes [11], is necessary when developing NGSS-aligned formative assessments [13]. More specifically, by operationalizing the NGSS practices into fine-grained sub-practices (see Table 2), [ITS] can identify, in real time, precisely how the students are struggling on the mathematical modeling sub-practices (e.g., labeling the axes of the graph, identifying the functional form in a graph, etc.) in the virtual environment. In turn, this fine-grained assessment allows [ITS] to provide targeted support to students on the specific sub-practice(s) on which they are struggling and subsequently help them in transferring these integrated mathematical modeling and science inquiry competencies.

The data presented in this study shows promising evidence that our assessment and scaffolding approach can improve students' mathematical modeling competencies in science inquiry contexts. However, to better understand the generalizability of students' improvement as well as whether the improvement occurred because of the scaffolding or because of the practice opportunities, a randomized controlled experiment with a larger sample size comparing students' improvement with scaffolding versus without scaffolding in the virtual labs will be conducted. Other future work will also disentangle how the scaffolding on one practice can impact students' competencies on other practices as well as examine students' ability to transfer their mathematical modeling competencies across physical science topics and assessment contexts outside of [ITS], all of which are critical to achieve the vision of NGSS [1].

## References


1. Next Generation Science Standards Lead States: Next generation science standards: for states, by states. National Academies Press, Washington (2013).
2. National Science Board.: Science and engineering indicators digest 2016 (NSB-2016-2). National Science Foundation, Arlington, VA (2016).
3. Gottfried, M. A., Bozick, R.: Supporting the STEM pipeline: Linking applied STEM course-taking in high school to declaring a STEM major in college. Education Finance and Policy 11, 177-202 (2016).
4. Potgieter M., Harding A., Engelbrecht J.: Transfer of algebraic and graphical thinking between mathematics and chemistry. J Res Sci Teach 45(2), 197-218 (2008).
5. Sadler, P. M., Tai, R. H.: The two high-school pillars supporting college science. Science Education 85(2), 111-136. (2007).
6. Glazer, N.: Challenges with graph interpretation: A review of the literature. Studies in Science Education (47), 183-210 (2011).
7. Lai, K., Cabrera, J., Vitale, J. M., Madhok, J., Tinker, R., & Linn, M. C.: Measuring graph comprehension, critique, and construction in science. Journal of Science Education and Technology 25(4), 665-681 (2016).
8. Nixon, R. S., Godfrey, T. J., Mayhew, N. T., Wiegert, C. C.: Undergraduate student construction and interpretation of graphs in physics lab activities. Physical Review Physics Education Research 12(1), (2016).
9. Casey, S. A.: Examining student conceptions of covariation: A focus on the line of best fit. Journal of Statistics Education, 23(1), 1-33 (2015).
10. De Bock, D., Neyens, D., Van Dooren, W.: Students' ability to connect function properties to different types of elementary functions: An empirical study on the role of external representations. International Journal of Science Mathematics Education 15(5), 939-955 (2017).
11. Penuel, W. R., Turner, M. L., Jacobs, J. K., Van Horne, K., Sumner, T.: Developing tasks to assess phenomenon-based science learning: Challenges and lessons learned from building proximal transfer tasks. Science Education 103(6), 1367-1395 (2019).
12. Furtak, E. M.: Confronting dilemmas posed by three-dimensional classroom assessment. Science Education 101(5), 854–867 (2017).

13. Harris, C. J., Krajcik, J. S., Pellegrino, J. W., McElhaney, K. W.: Constructing assessment tasks that blend disciplinary core Ideas, crosscutting concepts, and science practices for classroom formative applications. SRI International, Menlo Park, CA (2016).
14. Authors: (2013).
15. Authors: (2021).
16. Vanlehn, K., Wetzel, J., Grover S., Sande, B.: Learning How to Construct Models of Dynamic Systems: An Initial Evaluation of the Dragoon Intelligent Tutoring System. IEEE Transactions on Learning Technologies 10(2), 154-167 (2016).
17. Matuk, C., Zhang, J., Uk, I., Linn, M. C.: Qualitative graphing in an authentic inquiry context: How construction and critique help middle school students to reason about cancer. Journal of Research in Science Teaching 56(7), 905-936 (2019).
18. VanLehn, K., Lynch, C., Schulze, K., Shapiro, J. A., Shelby, R., Taylor, L., Treacy, D., Weinstein, A., Wintersgill, M.: The Andes physics tutoring system: Lessons learned. International Journal of Artificial Intelligence in Education 15(3), 147-204 (2005).
19. Koedinger, K. R., Anderson, J. R.: The early evolution of a Cognitive Tutor for algebra symbolization. Interactive Learning Environments 5(1), 161–179 (1998).
20. Aleven, V., McLaren, B. M., Roll, I., Koedinger, K. R.: Help helps, but only so much: Research on help seeking with intelligent tutoring systems. International Journal of Artificial Intelligence in Education, 26(1), 205-223 (2016).
21. Fretz, E.B., Wu, H.K., Zhang, B., Davis, E.A., Krajcik, J.S., Soloway, E.: An investigation of software scaffolds supporting modeling practices. Res Sci Educ 32(4), 567-589 (2002).
22. Bassok, M., & Holyoak, K. J.: Interdomain transfer between isomorphic topics in algebra and physics. Journal of Experimental Psychology 15(1), 153 (1989).
23. Bransford, J. D., Schwartz, D. L.: Rethinking transfer: A simple proposal with multiple implications. Review of Research in Education 24(1), 61-100 (1999).
24. Siler, S., Klahr, D., Matlen, B.: Conceptual change in experimental design: From engineering goal to science goals. 2nd edn. In: International Handbook of Research on Conceptual Change, pp.150-170. Routledge, New York (2013).
25. Koedinger, K. R., Baker, R. S., Corbett, A. T.: Toward a model of learning data representations. In: Proceedings of the Twenty-Third Annual Conference of the Cognitive Science Society, pp. 45-50. Erlbaum, Mahwah, NJ (2001).
26. Uhden, O., Karam, R., Pietrocola, M., Pospiech, G. Modelling mathematical reasoning in physics education. Science & Education 21(4), 485-506 (2012).
27. Jin, H., Delgado, C., Bauer, M. I., Wylie, E. C., Cisterna, D., & Llort, K. F.: A hypothetical learning progression for quantifying phenomena in science. Science & Education 28(9), 1181-1208 (2019).
28. Aleven, V., Koedinger, K.R.: Limitations of student control: Do students know when they need help? In: Intelligent Tutoring Systems, Springer, Berlin, Heidelberg (2000).
29. Authors: (2018).
30. Wood, H., Wood, D. Help seeking, learning and contingent tutoring. Computers and Education 33, 153-169 (1999).
31. Etymologia: Bonferroni correction. Emerg Infect Dis. 2015 Feb [Jan 16, 2023].
32. Kraft, M. A. Interpreting effect sizes of education interventions. Educational Researcher 49(4), 241-253 (2020).
33. Rebello, N. S., Cui, L., Bennett, A. G., Zollman, D. A., Ozimek, D. J. (2017). Transfer of learning in problem solving in the context of mathematics and physics. In: *Learning to solve complex scientific problems*, pp. 223-246. Routledge, New York (2017).