Real-Time AI-Driven Assessment and Scaffolding that Improves Students' Mathematical Modeling during Science Investigations

Amy Adair[1](✉), Michael Sao Pedro[2], Janice Gobert[1,2], and Ellie Segan[1]

[1] Rutgers University, New Brunswick, NJ 08901, USA
amy.adair@gse.rutgers.edu
[2] Apprendis, Berlin, MA 01503, USA

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 at the intersection of these practices, i.e., developing mathematical models about scientific phenomena. In this paper, we present the design and initial classroom test of AI-scaffolded virtual labs that help students practice these competencies. The labs automatically assess fine-grained sub-components of students' mathematical modeling competencies based on the actions they take to build their mathematical models within the labs. 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 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 data can help students improve on mathematical modeling.

Keywords: Scaffolding · Intelligent Tutoring System · Science Practices · Performance Assessment · Formative Assessment · Science Inquiry · Mathematical Modeling · Developing and Using Models · Virtual Lab · Online Lab · Pedagogical Agent · Next Generation Science Standards Assessment

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, i.e., the ways in which scientists study phenomena. Standards, such as the Next Generation Science Standards (NGSS) [1], define several such practices, including NGSS Practice 2 (Developing and Using Models) and Practice 5 (Using Mathematics and Computational Thinking). However,

N. Wang et al. (Eds.): AIED 2023, LNAI 13916, pp. 202–216, 2023.
https://doi.org/10.1007/978-3-031-36272-9_17

the difficulties that students experience with the scientific practices related to using mathematics and developing models can be barriers for students' access to and success in high school science coursework and future STEM careers [3–5]. Specifically, students often have difficulties developing mathematical models (i.e., graphs) with quantitative data in science inquiry contexts [6] because they struggle to properly label the axes of their graphs [7], interpret variables on a graph [8], make connections between equations and graphs [4], or choose the functional relationship to create a best-fit line or curve [9, 10]. Thus, students need resources capable of formatively assessing and scaffolding their competencies in a rigorous, fine-grained way as they work [12, 13] so that they can develop these critical competencies and, in turn, transfer them across science contexts [11].

In this paper, we evaluate the design of virtual labs in the Inquiry Intelligent Tutoring System (Inq-ITS), which are instrumented to automatically assess and scaffold students' competencies as they conduct investigations and develop mathematical models to represent and describe science phenomena [14–16]. To do so, we address the following research question: Did 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) [17], drawing qualitative graphs of science phenomenon (e.g., WISE) [18], and physics problem solving (e.g., Andes) [19]. However, these environments do not assess students' mathematical modeling competencies within the context of a full science inquiry investigation. Further, they do not 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 an authentic investigation that uses mathematical models to make inferences about science phenomena.

Scaffolding in online learning environments for both math and science has yielded student improvement on competencies by breaking down challenging tasks into smaller ones [20], providing hints on what to do next for students who are stuck on a task [21], and reminding students about the progress and steps taken thus far [20]. While scaffolding strategies have been applied to the online learning environments for modeling in science [19, 22], there are no studies, to our knowledge, that investigate the efficacy of AI-driven scaffolds for *mathematical modeling associated with 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 the Inq-ITS labs to improve students' competencies on science inquiry and mathematical modeling practices.

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 Fall 2022. Thirty-one percent of students qualify for free or reduced-price lunch; 71% identify as White, 16% as Hispanic, and 6% as two or more races.

All students completed two Inq-ITS mathematical modeling virtual labs on the disciplinary core idea of Forces and Motion (NGSS DCI PS2.A). Both labs were augmented with automated scaffolding. Students completed the two labs during their regularly scheduled classes. 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 (see Sect. 2.2 for more details). Inq-ITS automatically assessed students' competencies using previously validated educational data-mined and knowledge-engineered algorithms [14, 16, 23], which triggered scaffolds to students as they worked.

2.2 Inq-ITS Virtual Lab Activities with Mathematical Modeling

The virtual labs consisted of six stages that structured the investigation and captured different aspects of students' competencies at several NGSS practices (Table 1, Fig. 1). The goal of each activity was to develop a mathematical model (i.e., a best-fit curve represented by a graph and corresponding equation) that can explain how changing one factor (e.g., roughness of a ramp/road) impacted 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 [24, 25], 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, 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.

The design of the lab focuses on students' competencies with inter-related practices including collecting controlled data [26], plotting/graphing the data [27], and determining the informal line/curve of best fit [9] *without* deriving the algebraic equations, which shifts the focus to modeling the phenomenon rather than completing rote "plug-and-chug" methods often taught in physics problem-solving contexts [28]. 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 [29].

Table 1. Stages of the Inq-ITS Mathematical Modeling Virtual Lab Activity

Stage	Primary Related NGSS Practice(s)	Description of Stage
Stage 1: Hypothesizing/Question Formation	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 and select the variables to place on the x-axis and 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.

(*continued*)

Table 1. (*continued*)

Stage	Primary Related NGSS Practice(s)	Description of Stage
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 (linear, inverse, square, inverse square, 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 equation. 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 a best-fit curve/line.
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.

Fig. 1. Screenshots of Inq-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).

2.3 Approach to Automated Assessment and Scaffolding of Science Practices

Inq-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

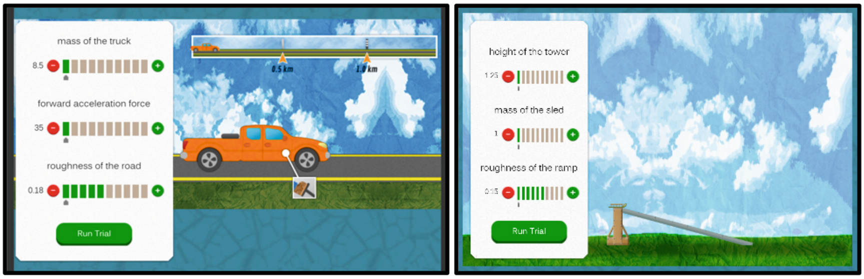

Fig. 2. The simulation in the Collecting Data stage of the Truck lab (left) and Ramp lab (right).

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, 23]. 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) [14–16, 23]. 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 [30] 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) [16]. This approach has shown to be effective in helping students learn and transfer other science inquiry competencies [31] even after many months [32]; 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, Rex. Rex 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 [30–33]. 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 [30, 33] stating the 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 Inq-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.

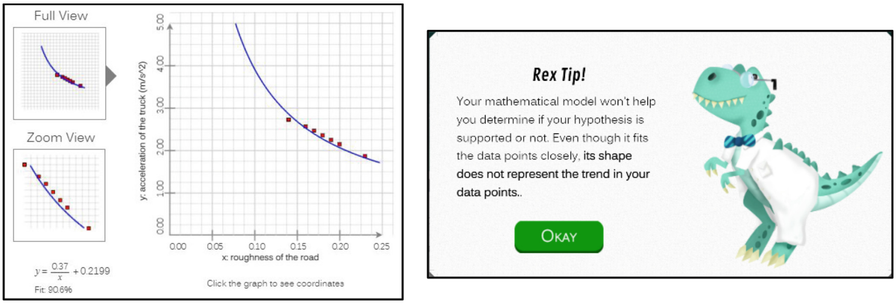

Fig. 3. Example screenshot of a student struggling with "Math Model has Good Form" sub-practice, but not with "Math Model has Good Fit" sub-practice (left; note: the student selected an *inverse* relationship when they should have chosen a *linear* relationship, given the variables on their graph); the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

To illustrate, consider a student who is struggling with choosing the mathematical functional form that best demonstrates the relationship between variables, 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 Inq-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). Rex 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 Rex states, "Your mathematical model won't help you determine if your hypothesis is supported or not. Even though it 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 (i.e., 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 (e.g., choosing the correct functional form) with which the student is struggling in that moment.

2.4 Measures and Analyses

To measure students' competencies, students' stage scores 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 [32]. Because students may receive multiple scaffolds addressing different sub-practices on a single stage and the effect of those

Table 2. Operationalization of Automatically Scored Sub-Practices in the Inq-ITS Virtual Lab

Stage	Sub-Practice	Criteria
Stage 1: Hypothesizing	Hypothesis 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 dependent variable that will be *measured* 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. Assessed by EDM algorithm [23]
	Data Collection is Controlled Experiment	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. Assessed by EDM algorithm [23]
	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

(*continued*)

Table 2. (*continued*)

Stage	Sub-Practice	Criteria
	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 controlled data with 5 unique values for the variable chosen for the x-axis. This number is sufficient to see mathematical trends for Inq-ITS' simulation designs
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. This minimum score balances between students spending too much effort maximizing fit, and not having a useful model. It represents a reasonably strong fit to the data

(*continued*)

Table 2. (*continued*)

Stage	Sub-Practice	Criteria
	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

scaffolds may be entangled, we use the measures of students' overall competencies at the stage level for this study's analyses.

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 examined students' competencies on each stage separately to determine students' improvement on the competency for which they were helped. Furthermore, to isolate whether each type of scaffolding improved students' performance on the respective competency, we ran four two-tailed, paired samples *t*-tests with a Bonferroni correction (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 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 [16]. However, unpacking the correlation between competencies and how the scaffolding may affect performance on multiple competencies was outside the scope of this study.

3 Results

We found that, for all 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 stages (i.e., Hypothesizing, Collecting Data, Plotting Data, Building Models; Table 3). Further, the effect sizes (Cohen's d) were large, suggesting that the automated scaffold was effective at helping students to improve at those competencies within the Inq-ITS labs.

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

Stage	N	Lab 1: M (SD)	Lab 2: M (SD)	Within-Subjects Effects
Hypothesizing	27	.41 (.30)	.74 (.27)	$t(26) = -5.45, p < .001, d = 1.05$
Collecting Data	37	.58 (.22)	.84 (.24)	$t(36) = -6.05, p < .001, d = 1.00$
Plotting Data	24	.63 (.21)	.86 (.21)	$t(23) = -4.12, p < .001, d = .84$
Building Models	31	.24 (.20)	.59 (.43)	$t(30) = -4.40, p < .001, d = .79$

4 Discussion

Students' competencies with mathematical modeling practices during science inquiry are critical for deep science learning [1, 2] and for future STEM courses and careers [4, 5]. However, students have difficulties with many aspects of mathematical modeling crucial to analyzing scientific phenomena [6] including those of focus in this study (e.g., identifying the functional form in plotted data [9, 10]). When students struggle with constructing and interpreting graphs in mathematics, it hampers their ability to transfer those competencies to science contexts [4, 34]. Further, even though these mathematical modeling competencies are necessary for developing deep understanding of science phenomena [6, 22, 29], they are not often addressed in science classrooms [7]. Thus, there is a need for resources that provide immediate, targeted support on the specific components for which students struggle, when it is optimal for learning [30, 31].

In this study, we found that students who received AI-driven real-time scaffolds during a virtual lab improved their mathematical modeling competencies when completing a near-transfer (i.e., isomorphic; [24]) task on the same physical science topic in a different scenario. These results suggest that 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' learning and transfer of their mathematical modeling competencies. We speculate that students improve because the lab design operationalized the mathematical modeling practices (e.g., NGSS Practices 2 & 5) into *fine-grained* sub-practices. More specifically, students were given support based on their *specific* difficulties with these fine-grained sub-practices (e.g., labeling the axes of the graph, identifying the functional form in a graph, etc.). Furthermore, by addressing concerns raised by others who have articulated the lack of specificity in the NGSS for assessment purposes [11], we have evidence that our approach toward operationalizing, assessing, and scaffolding the sub-practices associated with the NGSS practices can positively impact students' learning.

Though promising, 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. Future work will also disentangle how the scaffolding on one practice can impact students' competencies on other practices and examine students' ability to transfer their mathematical modeling competencies across physical science topics and assessment contexts outside of Inq-ITS, all of which are critical to achieve the vision of NGSS [1].

Acknowledgements. This material is based upon work supported by an NSF Graduate Research Fellowship (DGE-1842213; Amy Adair) and the U.S. Department of Education Institute of Education Sciences (R305A210432; Janice Gobert & Michael Sao Pedro). Any opinions, findings, and conclusions or recommendations expressed are those of the author(s) and do not necessarily reflect the views of either organization.

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. Educ. Fin. Pol. **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. Sci. Educ. **85**(2), 111–136 (2007)
6. Glazer, N.: Challenges with graph interpretation: a review of the literature. Stud. Sci. Educ. **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. J. Sci. Educ. Technol. **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. J. Stat. Educ. **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. Int. J. Sci. Math. Educ. **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. Sci. Educ. **103**(6), 1367–1395 (2019)
12. Furtak, E.M.: Confronting dilemmas posed by three-dimensional classroom assessment. Sci. Educ. **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. Gobert, J.D., Sao Pedro, M., Raziuddin, J., Baker, R.S.: From log files to assessment metrics: measuring students' science inquiry skills using educational data mining. J. Learn. Sci. **22**(4), 521–563 (2013)
15. Dickler, R., et al.: Supporting students remotely: Integrating mathematics and sciences in virtual labs. In: International Conference of Learning Sciences, pp. 1013–1014. ISLS (2021)

16. Olsen, J., Adair, A., Gobert, J., Sao Pedro, M., O'Brien, M.: Using log data to validate performance assessments of mathematical modeling practices. In: Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners' and Doctoral Consortium: 23rd International Conference, AIED 2022, Durham, UK, July 27–31, 2022, Proceedings, Part II, pp. 488–491. Springer International Publishing, Cham (2022). https://doi.org/10.1007/978-3-031-11647-6_99
17. 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 Trans. Learn. Technol. **10**(2), 154–167 (2016)
18. 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. J. Res. Sci. Teach. **56**(7), 905–936 (2019)
19. VanLehn, K., et al.: The Andes physics tutoring system: lessons learned. Int. J. Artif. Intell. Educ. **15**(3), 147–204 (2005)
20. Koedinger, K.R., Anderson, J.R.: The early evolution of a Cognitive Tutor for algebra symbolization. Interact. Learn. Environ. **5**(1), 161–179 (1998)
21. Aleven, V., McLaren, B.M., Roll, I., Koedinger, K.R.: Help helps, but only so much: research on help seeking with intelligent tutoring systems. Int. J. Artif. Intell. Educ. **26**(1), 205–223 (2016)
22. 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)
23. Sao Pedro, M., Baker, R., Gobert, J., Montalvo, O., Nakama, A.: Leveraging machine-learned detectors of systematic inquiry behavior to estimate and predict transfer of inquiry skill. User Model. User-Adap. Inter. **23**, 1–39 (2013)
24. Bassok, M., Holyoak, K.J.: Interdomain transfer between isomorphic topics in algebra and physics. J. Exp. Psychol. **15**(1), 153–166 (1989)
25. Bransford, J.D., Schwartz, D.L.: Rethinking transfer: a simple proposal with multiple implications. Rev. Res. Educ. **24**(1), 61–100 (1999)
26. Siler, S., Klahr, D., Matlen, B.: Conceptual change when learning experimental design. In: International Handbook of Research on Conceptual Change, pp.138–158. Routledge (2013)
27. 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)
28. Uhden, O., Karam, R., Pietrocola, M., Pospiech, G.: Modelling mathematical reasoning in physics education. Sci. Educ. **21**(4), 485–506 (2012)
29. Jin, H., Delgado, C., Bauer, M., Wylie, E., Cisterna, D., Llort, K.: A hypothetical learning progression for quantifying phenomena in science. Sci. Educ. **28**(9), 1181–1208 (2019)
30. Aleven, V., Koedinger, K.R.: Limitations of student control: do students know when they need help? In: Gauthier, G., Frasson, C., VanLehn, K. (eds.) Intelligent Tutoring Systems, pp. 292–303. Springer Berlin Heidelberg, Berlin, Heidelberg (2000). https://doi.org/10.1007/3-540-45108-0_33
31. Sao Pedro, M., Baker, R., Gobert, J.: Incorporating scaffolding and tutor context into Bayesian knowledge tracing to predict inquiry skill acquisition. In: Educational Data Mining, pp. 185–192 (2013)
32. Li, H., Gobert, J., Dickler, R.: Evaluating the transfer of scaffolded inquiry: what sticks and does it last? In: Isotani, S., Millán, E., Ogan, A., Hastings, P., McLaren, B., Luckin, R. (eds.) AIED 2019. LNCS (LNAI), vol. 11626, pp. 163–168. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-23207-8_31

33. Wood, H., Wood, D.: Help seeking, learning and contingent tutoring. Comput. Educ. **33**, 153–169 (1999)
34. Rebello, N.S., Cui, L., Bennett, A.G., Zollman, D.A., Ozimek, D.J.: 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)