# Using AI-Based Assessment and Scaffolds to Identify Student Difficulties with Plotting Data and Modeling in Virtual Science Labs

Ellie Segan, Rutgers University, ers173@scarletmail.rutgers.edu
Janice Gobert, Rutgers University & Apprendis, janice.gobert@gse.rutgers.edu
Michael Sao Pedro, Apprendis, mikesp@apprendis.edu
Amy Adair, Rutgers University, amy.adair@gse.rutgers.edu
Jessica A. Owens, Rutgers University, jo503@rutgers.edu

**Abstract:** Developing proficiency in science practices, including using mathematics, outlined in the Next Generation Science Standards is essential for success in STEM courses and future careers. However, students often struggle with developing mathematical models, which limits their ability to understand scientific phenomena. To improve students' learning and teachers' assessment, we extended Inq-ITS to automatically assess and scaffold students' competencies in developing mathematical models of scientific phenomena. We analyzed student data from six virtual science labs in Inq-ITS at both the practice level and the sub-practice level to determine if they maintained their mathematical competencies with scaffolding. By operationalizing and analyzing data at the sub-practice level, the results provide valuable formative data regarding the challenges students face when developing mathematical models about scientific phenomena, which in turn, can inform future scaffolds across science domains.

## Introduction

Most science classrooms in the United States are guided by the Next Generation Science Standards (NGSS, 2013) to develop students' understanding of phenomena and science practices including Using Mathematics (Practice 5) and Developing Models (Practice 2). Prior research has shown that students struggle with the mathematical modeling competencies needed to develop a deep understanding of scientific phenomena (McDermott et al., 1987). Since a lack of proficiency in math is a barrier to understanding science (Basson, 2002), addressing these challenges, particularly at the high school level when science and math are deeply intertwined, is crucial for better preparing students for future STEM majors and careers (Schuchardt & Schunn, 2015; Gottfried & Bozick, 2016).

In this study, we address two areas of difficulty for students: graphing data collected in science labs and determining the proper functional relationship of the model that best fits graphed data. Students commonly struggle with constructing, interpreting, and describing graphs while also connecting them to the underlying science concepts (Lai et al., 2016; Potgieter et al., 2008; Nixon et al., 2016). The current study investigates students' challenges with two types of graphical relationships: linear and inverse square. For tasks in which the inverse square law underlies science phenomena, students can struggle to transfer their understanding of the algebraic equation to graphical representation (Moynihan et al., 2019). Given that graphical relationships in science contexts can vary, and understanding graphing is an essential part of science, students need to understand each type of graphical relationship and the differences between them (McKenzie & Padilla, 1986) to deeply understand science, particularly physics, phenomena (Angell et al., 2008).

Though some online learning environments target the mathematical competencies used in science, such as dynamic systems modeling (VanLehn et al., 2016) and qualitative graphing (Matuk et al., 2019), these systems do not assess and scaffold the full range of NGSS practices necessary for developing deep understandings of science phenomena (NGSS, 2013). The platform in the current study, Inq-ITS (Gobert et al., 2013), engages students in performance-based formative assessments (i.e., virtual science investigations) covering a wide range of NGSS practices, including Carrying Out Investigations (running controlled trials; Sao Pedro et al., 2013) and Using Mathematics and Developing Models (mathematical modeling; Adair et al., 2023). Our approach operationalizes the NGSS practices into fine-grained sub-practices and leverages AI techniques to both assess students while they are working and support them on the specific sub-practices with which they are struggling in real-time when support is optimal for learning (Koedinger & Corbett, 2006). This work is grounded by the evidence-centered design (ECD) and learning progressions analytics (LPA) frameworks, which describe the use of fine-grained data from digital learning systems and ITSs to better understand students' understanding and skill level in science and mathematics (Kubsch et al., 2022).

Recent research on Inq-ITS scaffolds for mathematical modeling showed that students receiving scaffolds improved their performance on these practices across two labs (Adair et al., 2023). In the present study, we extended this research in two ways: RQ1) We analyzed students' competencies across multiple labs, testing for the transfer of students' competencies (thereby testing our scaffolds) when the mathematical relationship was the same (linear) and when the relationship changed to a more complex mathematical relationship (from linear to an inverse square) and, RQ2) We identified specific sub-practices for which students had challenges when transferring their competencies from a linear to inverse square mathematical relationship.

## Method

Participants included 41 high school science students in an Applied Physics class taught by one high school teacher in the Northeastern region of the United States. The participating school represents students from diverse backgrounds, including 29% of students receiving free and reduced lunch and 61% identifying as students of color. As part of this research, students worked individually and completed two sets of three Inq-ITS labs (6 labs total) that involved mathematical modeling in science during their normally scheduled classes in the Spring of 2023. Students worked on the second set of labs approximately three weeks after the first set.

Each lab involves six stages: (1) asking questions about the relationship between variables; 2) collecting data; 3) plotting data by selecting axes labels and data points for the graph; 4) building models by creating a best-fit curve through the data points; 5) forming and warranting claims about the mathematical relationship between variables; and 6) explaining findings in a claim, evidence, and reasoning format. Students were auto scaffolded in each of the first four stages when the algorithm detected students were struggling (Adair et al., 2023).

The two sets of Inq-ITS labs focused on Momentum (NGSS DCI PS2.A) and Electricity and Magnetism (NGSS DCI PS2.B), respectively. The students had received one lecture on Momentum before completing the labs, but no prior instruction was given on Electricity and Magnetism. The goal for each lab was for students to develop a mathematical model (i.e., graph and corresponding equation) to describe a scientific phenomenon. In the Momentum labs, the investigation involved a toy car moving toward and colliding with another stationary car. The goals for the Momentum labs included determining how the: (Lab 1) mass of the moving car affected the momentum of the moving car before collision, (Lab 2) velocity of the moving car before collision affects the momentum of the system after collision, and (Lab 3) velocity of the moving car before collision affects the velocity of the system after collision. In the Electricity & Magnetism labs, the investigation involved a large electromagnet picking up scrap metal in a junkyard. The goals for those labs included determining how the: (Lab 4) number of coil turns in the electromagnet affects the strength of the magnetic field, (Lab 5) amount of current through the electromagnet affects the strength of the magnetic field, and (Lab 6) distance of the electromagnet from the metal affects the lifting force of the electromagnet. The type of mathematical relationship that best demonstrates the scientific phenomena was linear for Labs 1 through 5 and inverse square for Lab 6.

### Measures

In the present study, we only examined student performance on the third and fourth stages (Plotting Data and Building Models), as these stages target students' mathematical modeling competencies related to our NGSS practices of interest (i.e., NGSS Practices 2 & 5). Students' competencies in these stages were assessed according to fine-grained sub-practices (Adair et al., 2023). Sub-practices for Plotting Data included: selecting appropriate variables for the axes, selecting axes that align with the lab goal, selecting axes that align with the hypothesis the student generated, plotting only controlled data, and plotting enough data points to see the model trend in the data. Sub-practices for Building Models included: building a model with the correct mathematical relationship (i.e., linear, inverse square), building a model that fit the data with at least a 70% fit, and building a model that had both the correct math relationship and a good fit. Students' sub-practice scores were automatically assessed as either 0 (incorrect) or 1 (correct) using knowledge-engineered algorithms that generate scores based on students' interactions within the Inq-ITS environment. Students' stage scores were calculated as an average of their sub-practice scores associated with that stage (Adair et al., 2023).

## Analyses and results

To determine whether there was a difference between the competency scores for the Plotting Data and Building Models stages across the six labs (RQ1), we performed a one-way Multivariate Analysis of Variance (MANOVA) with the independent variable as the 6 labs completed by students and the dependent variables as the Plotting Data and Building Models competency scores. There was a statistically significant overall difference in competency scores across labs $F(10, 478) = 6.07$, $p < .001$; Wilk's $\Lambda = .787$, partial $\eta^2 = .11$. Furthermore, there were significantly different competency scores across labs in the Plotting Data stage $F(5, 240) = 2.54$, $p = .029$, partial $\eta^2 = .05$ and the Building Models stage $F(5, 240) = 7.21$, $p < .001$, partial $\eta^2 = .13$. The mean competency scores for each lab are viewed in Table 1. This sample of participants performed at a high level across labs in both stages, except for Building Models in Lab 6 where performance decreased. Even though students performed well overall, the standard deviations in both stages (i.e., Plotting Data for Labs 1 to 6 and Building Models for Labs 1 to 5) decreased, indicating students continued to improve over time.

As mentioned in the Methods, the phenomena explored in Labs 1 through 5 demonstrated a linear graphical relationship, whereas the phenomenon in Lab 6 had an underlying inverse square relationship. To understand whether students transferred their mathematical competencies when the underlying relationship changed from a linear to inverse square (RQ2), we performed two paired samples t-tests for the Plotting Data and Building Model stage, comparing those competency scores on Lab 5 (which explored the linear relationship between the amount of current in an electromagnet and the strength of the magnetic field) versus Lab 6 (which explored the inverse square relationship between the distance of the electromagnet and the lifting force of the electromagnet). One t-test was performed per competency (i.e., Plotting Data and Building Models), at an alpha of .025 (see Table 2). There was no significant difference found for Plotting Data, but a significant difference was found for Building Models such that students' performance decreased in Lab 6 when the underlying relationship changed from linear to inverse square.

To better understand the challenges that students experienced with the inverse square task (Lab 6), we examined the sub-practices scores for the 12 (out of 41) students who struggled (i.e., required scaffolds) with the Building Models stage in Lab 6. Here, upon further investigation of the students' log data (that recorded all student actions for each lab), we noted that all 12 students first selected "inverse" ($1/x$) to describe the relationship of their data, rather than "inverse square" ($1/x^2$). After receiving scaffolds, all 12 students were then able to correctly identify the correct inverse square relationship (i.e., where the best-fit model forms a rapidly decreasing curve). Though they were incorrect in selecting "inverse" rather than "inverse square" to describe the relationship, they were all able to recognize that it was not a linear relationship, as had been the case in the previous 5 labs. Yet, they were not yet able to recognize the best-fit curve as "inverse square" upon their first observation of the data.

**Table 1**
*Average Competency Scores in the Plotting Data and Building Models Stages across Inq-ITS Labs*

| Description of Lab (Lab #) | Plotting Data *M* (*SD*) | Building Models *M* (*SD*) |
|---|---|---|
| Momentum: Mass vs. Momentum (Lab 1) | .90 (.23) | .93 (.22) |
| Momentum: Velocity vs. Momentum (Lab 2) | .92 (.17) | .98 (.10) |
| Momentum: Velocity Before vs. After Collision (Lab 3) | .96 (.15) | .98 (.10) |
| Magnetism: Coil Turns vs. Magnetic Field Strength (Lab 4) | .97 (.13) | .98 (.16) |
| Magnetism: Current vs. Magnetic Field Strength (Lab 5) | .99 (.05) | 1.00 (.00) |
| Magnetism: Distance vs. Lifting Force (Lab 6) | .99 (.06) | .79 (.35) |

**Table 2**
*Results of Paired Samples t-Tests for Lab 5 Competency Scores (Magnetism: Current vs. Magnetic Field Strength) vs Lab 6 Competency Scores (Magnetism: Distance vs. Lifting Force)*

| Math Stage | *N* | Lab 5 (Linear): *M* (*SD*) | Lab 6 (Inverse Square): *M* (*SD*) | Within-Subjects Effects |
|---|---|---|---|---|
| Plotting Data | 41 | .99 (.05) | .99 (.06) | $t(40) = -.334$, $p = .740$, $d = .06$ |
| Building Models | 41 | 1.0 (.00) | .79 (.35) | $t(40) = 3.89$, $p < .001$, $d = .35$ |

## Discussion

In this study, we tested whether students were able to transfer their mathematical modeling competencies across multiple labs that explored scientific phenomena with linear and then inverse square relationships. We determined that students maintained their high competencies in mathematical modeling stages (i.e., Plotting Data and Building Models) across all 6 labs, except for the Building Models stage during Lab 6 where performance decreased. Further, the students who struggled to transfer their competencies in the Building Models stage when the mathematical relationship changed from linear to inverse square (Lab 6) incorrectly identified the type of graphical relationship before receiving scaffolds. These findings suggest that students may need additional practice (i.e., more labs) containing an inverse square graph and scaffolding that highlights the differences (i.e., different equations and shape of best-fit curve) between an inverse and inverse square graph. In general, knowing how to integrate different graphical relationships into scientific investigations contributes to students' understanding of scientific phenomena (Angell et al., 2008). Graphing and modeling competencies are particularly relevant in high school physics classrooms (McDermott et al., 1987), when students integrate those skills in topics such as electricity and magnetism, kinematics, and conservation.

A limitation of this study was that this sample of students were high-performing in their competencies; therefore, there were fewer opportunities to identify student difficulties in plotting data and modeling practices, and, in turn, less opportunity to test the efficacy of our scaffolds for these practices. We only tested these six labs with one classroom teacher. Future work will involve using multiple classrooms from different school districts and collecting more data to find student weaknesses and further test the efficacy of Inq-ITS scaffolds.

Overall, by using Inq-ITS, students are provided with the unique opportunity to engage in authentic science investigations involving important mathematical modeling competencies, operationalized into fine-grained sub-practices, which provide greater specificity to NGSS practices, including 2 and 5 (NGSS, 2013). The autoscoring in Inq-ITS generated formative performance-based assessment data that exposed specific areas of challenge for students as they transferred their mathematical competencies to a lab containing a different, more difficult graphical relationship. The automated scaffolds addressed student difficulties in real-time, which allowed them to refine their competencies in a rich inquiry context. As we continue to evaluate students on these practices and identify other challenges with mathematical practices embedded in science, we can continue to develop and refine our scaffolds to better support students. This level of support is especially important for students who are falling behind in math at the high school level, which is a major barrier to understanding science (Basson, 2002). In future work, we will also continue to develop, test, and refine assessments and scaffolds for other mathematical competencies used in science investigations so that teachers can leverage this technology and its scalable assessments to achieve the visions set forth by the NGSS.

## References


Adair, A., Sao Pedro, M., Gobert, J., & Segan, E. (2023). Real-time AI-driven assessment and scaffolding that improves students' mathematical modeling during science investigations. In, *AIED* (pp. 202–216).

Angell, C., Kind, P. M., Henriksen, E. K., & Guttersrud, Ø. (2008). An empirical-mathematical modelling approach to upper secondary physics. *Physics Education, 43*(3), 256-264.

Basson, I. (2002). Physics and mathematics as interrelated fields of thought development using acceleration as an example. *Int J Math Educ Sci Technol, 33*(5), 679-690.

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.

Gottfried, M. A., & Bozick, R. (2016). 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.

Koedinger, K., & Corbett, A. (2006). Cognitive Tutors: Technology bringing learning sciences to the classroom. In, *The Cambridge Handbook of the Learning Sciences* (pp. 61-77). Cambridge University Press.

Kubsch, M., Czinczel, B., Lossjew, J., Wyrwich, T., Bednorz, D., Bernholt, S., Fiedler, D., Strauß, S., Cress, U., Drachsler, H., Neumann, K., & Rummel, N. (2022). Toward learning progression analytics—Developing learning environments for the automated analysis of learning using evidence centered design. *Frontiers in Education*, *7*, 981910.

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.

Matuk, C., Zhang, J., Uk, I., & Linn, M. C. (2019). 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.

McDermott, L. C., Rosenquist, M. L., & van Zee, E. H. (1987). Student difficulties in connecting graphs and physics: Examples from kinematics. *American Journal of Physics, 55*(6), 503-513.

McKenzie, D. L., & Padilla, M. J. (1986). The construction and validation of the Test of Graphing in Science (TOGS). *Journal of Research in Science Teaching, 23*(7), 571–579.

Moynihan, R., van Kampen, P., Finlayson, O. E., & McLoughlin, E. (2019). Developing second level students' understanding of the inverse square law and electric fields. *J. Phys.: Conf. Ser.*, *1286*(1), 012014.

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).

Potgieter, M., Harding A., & Engelbrecht J. (2008). Transfer of algebraic and graphical thinking between mathematics and chemistry. *Journal of Research in Science Teaching, 45*(2), 197–218.

Schuchardt, A. & Schunn, C. (2015). Modeling scientific processes with mathematics equations enhances student qualitative conceptual understanding and quantitative problem solving. *Sci. Educ., 100*(2), 290-320.

Sao Pedro, M., Baker, R., Gobert, J., Montalvo, O., & Nakama, A. (2013). 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.

Van Lehn, K., Wetzel, J., Grover S., & Sande, B. (2016). 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.


 