The Impacts of Automatic Scaffolding on Students' Acquisition of Data Collection Inquiry Skills

Michael A. Sao Pedro[1,2] Janice D. Gobert[1,2] Ryan S. Baker[3]
mikesp@wpi.edu jgobert@wpi.edu ryan@educationaldatamining.org

[1]Learning Sciences and Technologies Program, Worcester Polytechnic Institute
[2]Apprendis LLC
[3]Teacher's College, Columbia University

Abstract: We explore in this paper if automated scaffolding delivered via a pedagogical agent within a simulation can help students acquire data collection inquiry skills. Our initial analyses revealed that such scaffolding was effective for helping students who initially did not know two specific skills, designing controlled experiments and testing stated hypotheses, acquire those skills. These results provide evidence towards realizing rigorous, scalable, performance-based assessment of scientific inquiry skills and the efficacy of an automated scaffolding approach.

Introduction

Science educators and reseachers agree that cultivating inquiry and critical thinking skills are necessary for students to become scientifically literate [1], [2], [3], [4] and to be well-poised for the demands of the knowledge-based economy of the 21st century [5]. However, learning inquiry skills is challenging for students (e.g. [6], [7], [8], [9], [10]). These challenges can lead to many false starts [11], misconceptions [12], and failure to learn targeted science principles [13]. Given the importance of inquiry and these challenges, it is important to understand how best to foster learning of these skills (cf. [13], [14]).

Relevant to this paper, we consider skills at designing and conducting experiments (cf. [1]). Some studies showed that explicitly teaching strategies like controlling for variables [15] can lead to acquisition [16], [17], [18], [19], retention [18], [20], and transfer [16], [17], [18] of the strategy over pure discovery methods. Others showed that long-term, repeated practice also promotes skill acquisition, without instruction [21], [22], [23]. Scaffolding-based approaches, on the other hand, may strike a balance between these extremes by providing help *only* when students need it (cf. [24], [25]). For example, providing structure (scaffolding) during open-ended inquiry activities can foster learning (e.g. [26]). Similarly, individualized, real-time feedback may also help students learn inquiry skills [27].

Towards promoting the learning of data collection inquiry skills, we determine if scaffolding can help students acquire two such skills [1]: testing one's stated hypothesis, and designing controlled experiments [28], [29]. This work is conducted within the context of a web-based inquiry learning environment, called Inq-ITS[1], in which students conduct inquiry using simulations [29]. The learning environment was augmented to provide real-time, automated scaffolding as students collect data with the simulation. This scaffolding is driven, in part, by data-mined models that determine when students are off-track [30], [28]. A randomized, controlled experiment was conducted with two groups of students, those who received data collection scaffolds and those who did not, to determine if scaffolding impacted data collection skill acquisition in one set of inquiry activities on phase change, a middle school physical science topic. We hypothesize that our automated scaffolding approach that provides just-in-time feedback can improve students' data collection skills.

[1] http://slinq.org

Inq-ITS

Inq-ITS (**Inq**uiry **I**ntelligent **T**utoring **S**ystem) is a web-based, virtual science lab environment that aims to automatically assess and provide students personalized feedback on their inquiry and critical thinking skills [29]. In this environment, students conduct inquiry by forming hypotheses, collecting data, analyzing data, and communicating findings using interactive simulations. The simulations provide a focal point around which students conduct their inquiry, and are designed to address concepts and misconceptions aligned to the NGSS standards [2]. Currently, simulations have been developed for middle school Physical, Life and Earth Science. In addition to simulations, students also utilize inquiry support tools, such as a hypothesizing tool and analyzing data tool. The tools help students conduct inquiry, keep track of their progress, and enable assessment by making their thinking explicit [29].

Inq-ITS is similar to other microworld/simulation-based discovery environments (e.g. [31], [32], [33], [34]) in that the computer-based activities structure students' exploration and share a goal in bootstrapping the acquisition of content knowledge. In particular, like [35], it also emphasizes performance-based assessment of inquiry skills. Also, like [36], the system aims to provide real-time feedback to students as they work. Inq-ITS specifically aims to assess and provide real-time feedback on skills identified by national and state frameworks [1] like hypothesizing, designing and conducting experiments, interpreting data, and communicating findings. Thus, the system aims to provide students with supports so they do not flounder or engage in unproductive, haphazard inquiry behaviors [10], [37]. This approach is commensurate with a conceptualization of science inquiry described by Kuhn et al. [38], p.497, "students investigate a set of phenomena – virtual or real – and draw conclusions about the phenomena."

Each Inq-ITS activity is a performance assessment; the actions students take within the environment and work products they create are the bases for assessment. Towards the goal of real-time formative assessment, the system provides automatic feedback both to students and educators as students engage in the inquiry activities. For educators, the system automatically generates formative metrics and summary reports on the development of these skills. Educators can pinpoint which students are having difficulty and on what specific inquiry skills. For students, a pedagogical agent gives immediate feedback on their work products and experimentation processes to support them in improving their inquiry skills. To concretize these notions, we describe inquiry activities for Phase Change, a physical science topic that is the focus of this study, and describe how the pedagogical agent provides feedback on students' data collection processes.

The Phase Change activities [29] seek to foster understanding about the melting and boiling properties of water. In these activities (like all Inq-ITS activities), students are first given an initial goal around which their inquiry is to be conducted. In phase change, a goal would be to determine if one of three factors (size of a container, amount of ice to melt, and amount of heat applied to the ice) affects various outcomes (e.g. melting or boiling point). They then engage in a semi-structured scientific inquiry to address the goal as follows. First, they articulate a hypothesis to be tested using a hypothesis widget (Figure 1). The widget supports students in forming a testable hypothesis. Next, students collect data to test their hypothesis (Figure 2) with the Phase Change simulation. Students change the simulation's variables, and then run, pause, and reset it to collect their data. A data table tool auto-populates and shows the data they collected thus far. Once finished, they analyze their data (Figure 3) by forming an argument and selecting trials to indicate whether their hypotheses were supported or refuted based on the data they collected. Thus, in each activity, students hypothesize, collect and interpret data, and warrant their claims to address the goal. More information about the learning environment and support tools is described in [29].

This work centers on assessing and scaffolding two skills associated with productive data collection, designing controlled experiments and collecting data to test one's stated hypotheses [28], [29]. Briefly, students design controlled experiments when they generate data that make it possible to infer how changeable factors affect outcomes. This skill relates to the application of the Control of Variables Strategy (CVS; cf. [15]), but unlike CVS, it takes into consideration *all* the experimental design setups run with the simulation, not just isolated, sequential pairs of trials [28], [39]. Students test their stated hypotheses when they collect data that can support or refute an explicitly stated hypothesis. These skills are separated for two reasons. First, each skill can be demonstrated separately as students collect data. Students may attempt to test their hypotheses with confounded designs, or may design controlled experiments for a hypothesis not explicitly stated. As will be described below, this also enables the system to provide different scaffolds based on whether students have difficulty with either (or both) skills. Second, skill at testing hypotheses may be indicative of students' successful planning and monitoring of their inquiry [40].

Since these are process skills, assessment is based on students' interactions with the simulation, like running trials and changing the simulation's variable values (Figure 2), while collecting data. Next, we describe how the system provides automated feedback to help students learn these skills.

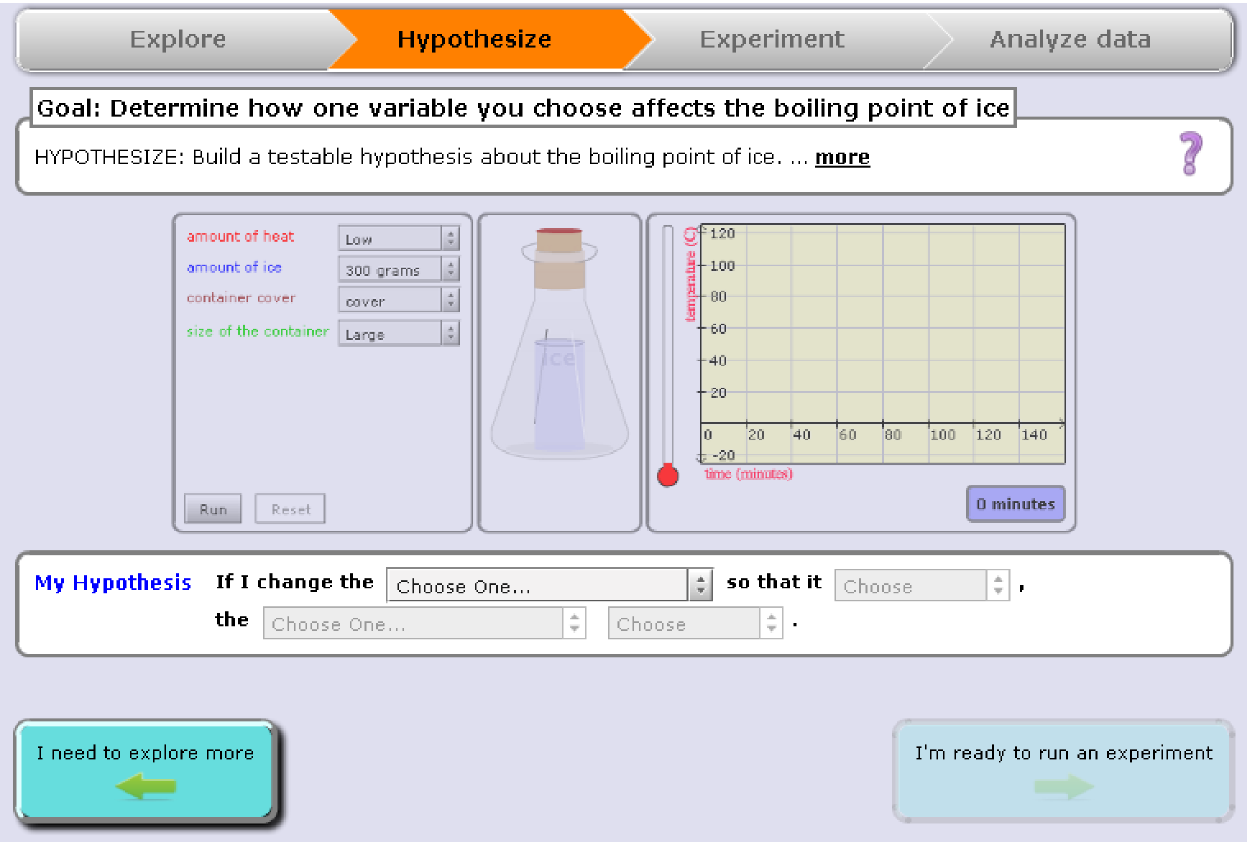

Figure 1. In a Phase Change inquiry activity, students first attempt to construct a hypothesis they can test using the hypothesis widget.

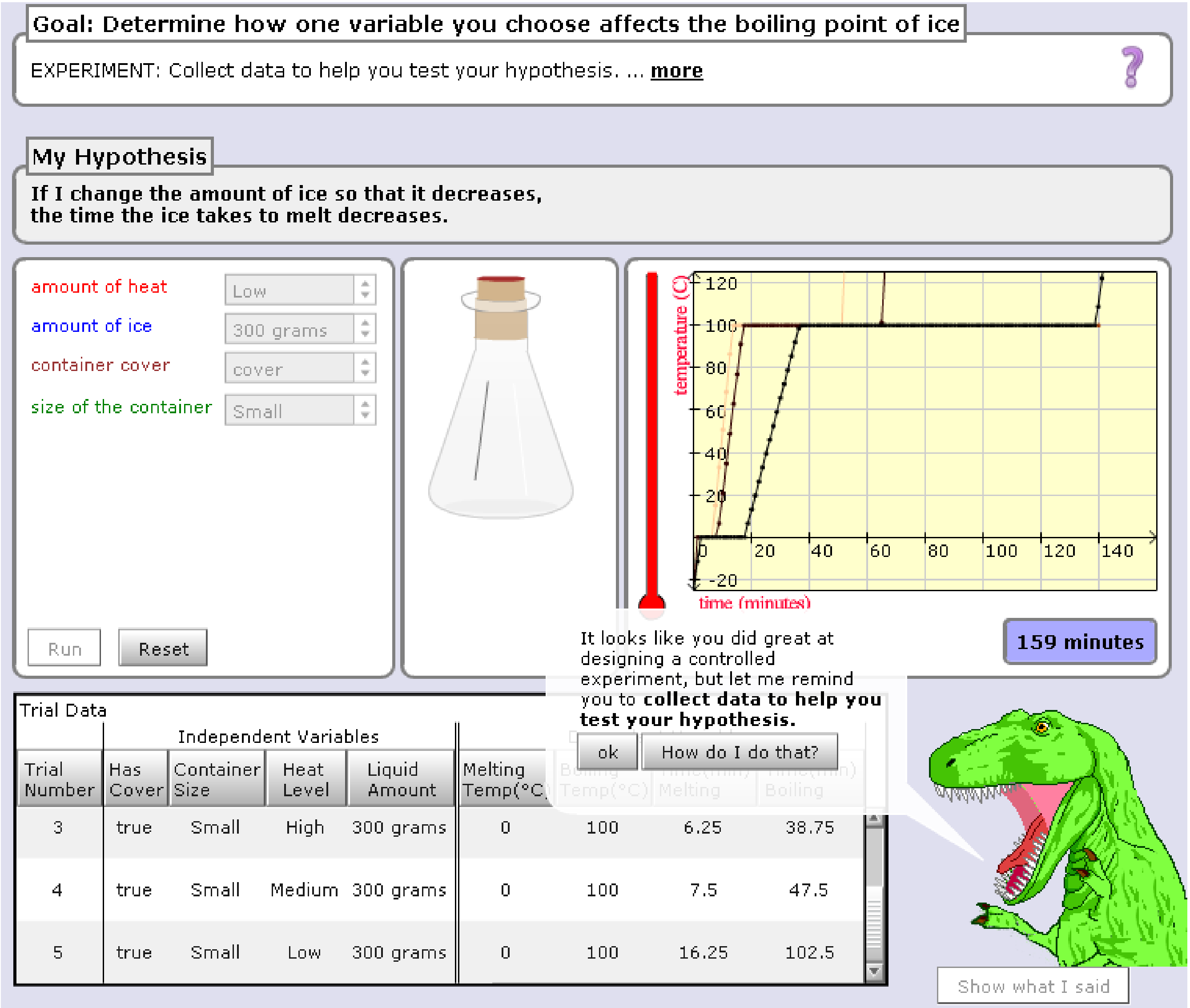

Figure 2. After hypothesizing, students collect data by designing and running experiments with the Phase Change simulation. Here, the pedagogical agent Rex responds to a student who appears to be designing controlled experiments, but is not testing their hypothesis. They can continue experimenting or ask Rex for more help, in this case by clicking “How do I do that?”

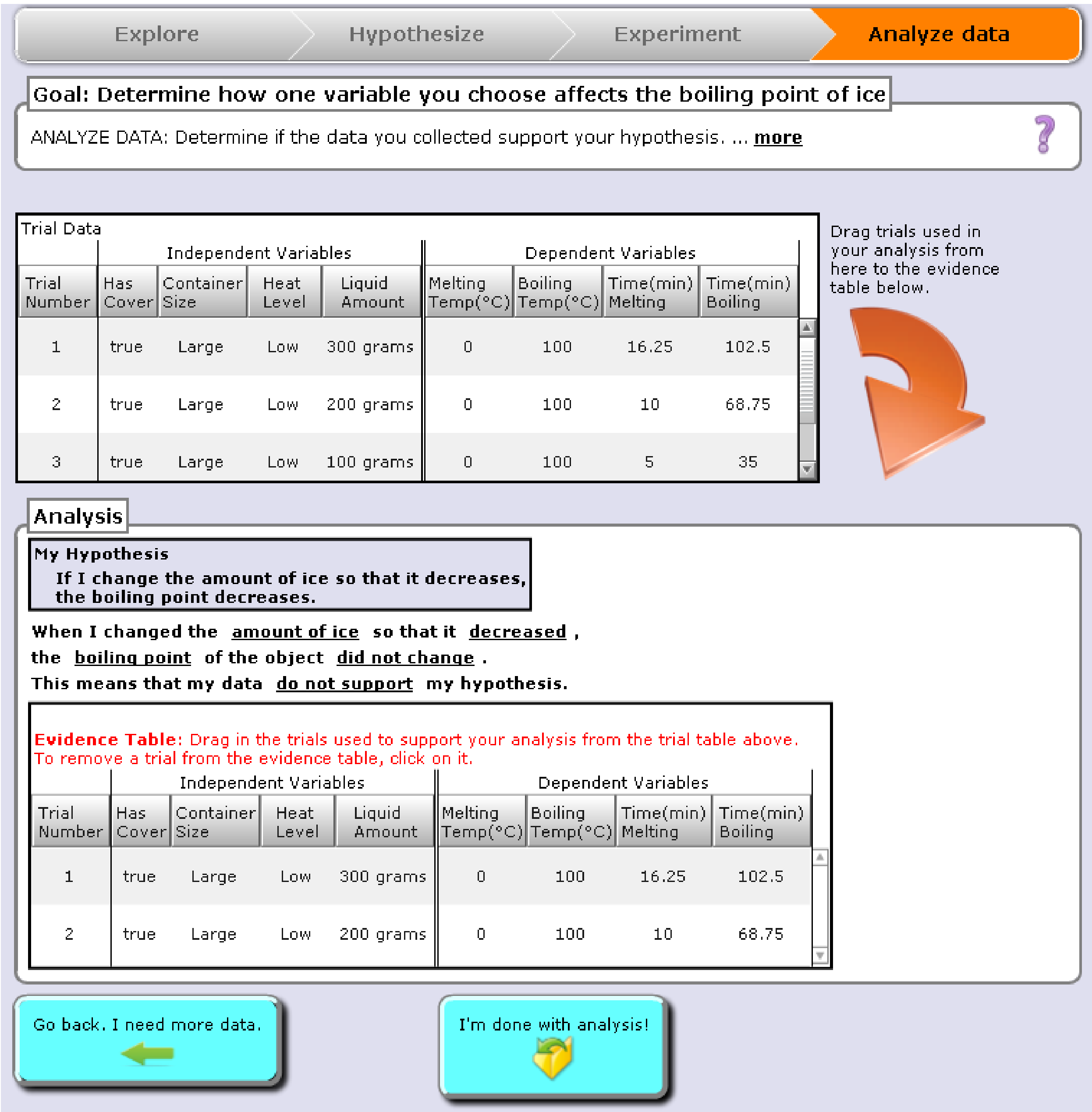

Figure 3. After collecting data, students then determine if their hypothesis was supported or not by analyzing their data. Here, students construct an argument using the analysis widget pulldown menus and select the trials they use as evidence to warrant their claim.

Real-Time Scaffolding and Evaluation of Data Collection Skills

Inq-ITS delivers scaffolds and hints to students via a pedagogical agent named Rex, a cartoon dinosaur (Figure 2). Rex provides feedback as students experiment when he detects they are off track. If a student continues to struggle, more targeted feedback is provided, similar to Cognitive Tutors (e.g. [41], [42], [43]). For example, if Rex detects that a student is designing controlled experiments but not collecting data to test their hypothesis, Rex will say "It looks like you did great at designing a controlled experiment, but let me remind you to collect data to help your test your hypotheses." If the student continues struggling, "bottom-out" feedback is given (cf. [42]): "Let me help some more. Just change the [IV] and run another trial. Don't change the other variables. Doing this lets you tell for sure if changing the [IV] causes changes to the [DV]". While collecting data, students may exhibit skill at testing their hypotheses, designing controlled experiments, both, or neither. To account for these possibilities, we implemented different scaffolding levels for each case in which the skills were not demonstrated. Finally, we also provided on-demand help [44] students can activate on their own to ask Rex for more clarification (e.g. clicking "How do I do that?" for the scaffold presented in Figure 2). The full hierarchy of scaffolds for data collection used in this study are provided in the Appendix.

The proactive, automated feedback approach was chosen for two reasons. First, prior work suggests that in general, students have difficulty engaging in product inquiry without support [11], [40], [10]. Second, students may lack the metacognitive help-seeking skills to recognize when they should ask for help [45], [44]. Reacting when students appear to be off-track, therefore, may be beneficial [46], [47]. A pedagogical agent was chosen specifically, because they have been shown to benefit learners (cf. [48], [49]), possibly by increasing students' engagement and motivation [50], [51], [52], [53], [54].

The core of this approach hinges on the system's ability to *evaluate* students' experimentation patterns to determine when a student demonstrates good data collection skills, and subsequently intervene with automated scaffolding when they are not. The assessment challenge is that these process skills are ill-defined; students' data collection patterns can vary widely and there are many ways to successfully demonstrate (or not demonstrate) them [55]. In our approach, we built and validated data-mined detectors[2] to evaluate students' data collection [28], [30], [56], [57]. This approach was chosen for two reasons. First, the approach attempts to overcome limitations of other models that either under- or over-estimate students' skill [55]. Data mining attempts to account for "corner" cases when students do not conduct their inquiry in lock-step, unlike other approaches that make this assumption (e.g. [27], [33]). Again, this is particularly important since there is variability and ways in which students interleave behaviors and strategies when designing experiments [55]. Second, the data mining approach enables easier validation of how well the model performs for new student interactions and new activities by testing it against data not used to build it, addressing issues of reliability and scalability in performance-based assessments (see [55], [57] for a discussion).

The detectors aim to replicate a human expert's ability to look at a student's log file and determine whether or not they designed controlled experiments and tested their stated hypotheses. Thus, to train and test the detectors, labels of students' log files were generated using text replay tagging [58], [28]. In this process, a human expert looks at "pretty-printed" versions of log files and labels whether the student designed controlled experiments and/or tested their stated hypotheses. From there, a feature set—indicators of whether or not students demonstrate skills computed over the log files—was derived. Example features considered

[2] In our prior work, we identified a situation in which the detector for designing controlled experiments incorrectly identifies skill demonstration. For this situation only, we authored a rule to evaluate experimentation [30].

from the full set outlined in [59], [30] include: number of trials run, number of hypotheses stated, count of pairwise controlled trials, time spent running experiments, and number of simulation pauses. The detector, originally built for Phase Change, was further refined by choosing features that increased the theoretical construct validity of the detector, and by iteratively refining it to find an optimal feature set [30], [57]. Full details about the detector construction process and its features can be found in [30], [57].

We also conducted extensive validation tests to show the applicability of these detectors across our physical science activities to identify the skills and determine when students are off-track. In the context of Phase Change, for example, they have been shown to adequately identify skill when students complete their experimentation [30]. They also could be used, as is, to detect when students are off-track and thus can be used to drive scaffolding *before students complete their data collection* [30]. Finally, they could be applied to new students [56] and to detect skill demonstration in other physical science topics [56], [57], and a life science topic with a complex systems simulation [60].

In this work, we aim to determine whether the data collection scaffolds are effective at helping students acquire the two data collection skills in the context of Phase Change, a physical science Inq-ITS activity set. The detectors are leveraged both to determine who should receive scaffolding and to evaluate who acquired the skills.

Method

Participants

Participants were 299 eighth grade students from three schools in Central Massachusetts. Some had prior experience conducting inquiry within Inq-ITS, and for others, this was their first experience.

Procedure

Five inquiry-based activities were developed for Phase Change. Three targeted specific Phase Change concepts and two had students test their own hypotheses, subject to the factors they could vary with the simulation. For the first four activities, students practiced inquiry in one of two learning conditions, randomly chosen by the system:

- Data collection scaffolding (DCS) condition: Scaffolds for writing a testable hypothesis (Figure 1) and for collecting data (Figure 2) were given to students if the system detected they were off-track. Scaffolds for analyzing data were not provided.
- No data collection (NoDCS) scaffolding condition: Only scaffolds for writing a testable hypothesis were present.

We highlight that students in the "No Data Collection (NoDCS) scaffolding condition" received scaffolds on constructing a testable hypothesis with the hypothesis widget (Figure 1), but not how to collect data to test that hypothesis. Furthermore, neither condition received scaffolds on analyzing data. This experimental design was chosen for two reasons. First, the design guarantees that students formulate a syntactically correct, testable hypothesis when they enter the experiment phase. Second, the design ensures that the efficacy of the data collection scaffolds is tested in isolation. For example, if the system provides feedback during data analysis to collect more data (e.g. all the student's trials are confounded, preventing a good data analysis), the feedback could affect their performance at data collection. This would prevent us from disentangling the impacts of data collection scaffolds from analyzing data scaffolds.

Finally, both groups completed an "immediate acquisition test", a fifth Phase Change activity with no scaffolding for hypothesizing, data collecting, or analyzing data. This enabled measuring the impacts of scaffolding on skill acquisition when the scaffolds were removed.

Results

We aim to determine the efficacy of our automated scaffolding approach for helping students acquire two data collection inquiry skills, designing controlled experiments and testing one's stated hypotheses. Efficacy is determined by evaluating if students who received scaffolding (DCS condition) were more likely to demonstrate the skills in the final, completely unscaffolded (fifth) Phase Change activity, than those who did not receive scaffolding (NoDCS). As mentioned, the detectors evaluate whether or not students demonstrated the two skills.

In this analysis, we consider *only students who did not demonstrate either skill in their first data collection* opportunity for two reasons. First, this approach accounts for students in either condition who may already know both skills. Second, it accounts for students in the DCD condition who may never have received scaffolding because they already knew the skills. This enables a more rigorous test of the efficacy of the scaffolding approach. From the original set of 268 students, 123 students who did not design controlled experiments in their first data collection, and 95 who students did not test their stated hypotheses were used in the analyses.

As shown in Table 1, the scaffolding approach appeared to help students who initially did not know the skills acquire those skills. More specifically, 92.9% of students who did not initially design controlled experiments in the DCS condition did so in the unscaffolded Phase Change activity compared 58.5% of the students in the NoDCS condition, $\chi^2(1) = 20.79$, $p < .001$. In addition, 91.7% students in the DCS condition tested their stated hypotheses compared to 53.2% of the students in the NoDCS condition, $\chi^2(1) = 17.69$, $p < .001$. The implications of these findings are discussed next.

Table 1. Crosstabulations of practice condition, and whether students demonstrated skill in the immediate transfer test, a completely unscaffolded Phase Change inquiry activity, $n = 123$ for designing controlled experiments, and $n = 95$ for testing stated hypotheses. The detectors were used here to evaluate whether students designed controlled experiments and/or tested their stated hypotheses. Students considered in this analysis originally did not demonstrate skill in their first attempt at conducting inquiry.

	Designed Controlled Experiments?		Tested Stated Hypotheses?	
	No	Yes	No	Yes
No DC Scaff.	22	31	22	25
DC Scaff.	5	65	4	44
	$\chi^2(1) = 20.79$***		$\chi^2(1) = 17.69$***	

Discussion and Conclusions

In this study, we extended our inquiry environment, Inq-ITS, a computer-based environment that can automatically assess students' scientific inquiry skills [29], to incorporate automated, real-time scaffolding. We explored whether scaffolding would help students acquire and transfer two data collection skills, designing controlled experiments and testing stated hypotheses, in the context of one set of inquiry activities on Phase Change, a middle school physical science topic. The real-time scaffolding was driven, in part, by data-mined detectors of these skills that could determine when students were haphazard in their data collection [30]. Overall, we found that our scaffolding approach was effective in helping students acquire these skills by comparing two groups of students in a randomized controlled study, those who received scaffolding and those who did not.

This work makes three contributions to the literature on inquiry learning and on providing interventions using data-mined detectors. First, these findings are particularly promising as an approach to simultaneously assess and support inquiry skill development in a scalable way because the approach is entirely computer-based. Thus, this system has the potential to be implemented readily in a classroom setting, or as virtual homework, and can provide individualized support to students who need it. Second, though other successful interventions have been developed that leverage data mining-based detectors (e.g. [61], [62]), this is the first system to our knowledge that evaluates students' skills, particularly inquiry processes (the actions they take while collect data), and uses that information to provide immediate feedback. Third, we showed that a "middle ground" between direct instruction and discovery learning [13], [14] has the potential to enable acquisition of these skills. The approach employs Vygotsky's original notion of scaffolding [24], prescribing that the scaffold be removed once the skill has been internalized by the student [63].

There are some limitations to this study. First, we did not fully address whether this environment could be used to teach the data collection skills to students who do not know these skills (cf. [64]). In general, we envision our learning environment to be an assessment platform that provides students just-in-time help, not as a pure instructional tool. In other words, we expect this tool to be used as an environment to hone inquiry skills that provides scaffolds as needed during practice (cf. [41], [65], [66]), after students are exposed to these inquiry topics in their regular curriculum [29]. Second, the evidence of acquisition and transfer is rooted primarily in procedural demonstration of the two data collection skills, and does not tap changes in conceptual / metastrategic knowledge of when and why one should apply them (e.g. [3], [67], [17]). One possible way to address this is to have students explain why they chose to design the experiments they did and code the open responses for evidence of such understanding (e.g. [22]). This approach, however, would be difficult to scale. Another possibility is to triangulate students' performance in the "analyze data" task in which students make inference about the data they collected [29] with their performance in the "experiment" task. If students are able to successfully warrant their interpretations relative to their hypotheses by identifying which data enabled them to make inferences, this would be evidence of conceptual understanding of the data collection skills.

Finally, though our results are encouraging, we recognize that the acquisition test was in the same science topic as that in which students practiced their inquiry [68]. In related work, we have shown that this scaffolding approach also helped students transfer these skills to a second physical science topic, also with a similar structure [59], [55], [69]. Our line of research will continue conducting similar studies using activities from dissimilar domains with different activity structures, like Life and Earth Science [29], to tease apart these possible effects and determine if scaffolding enables broad transfer. In addition, we will develop scaffolds for other inquiry skills (e.g. analyzing data skills) and determine how these scaffolds can promote skill acquisition and transfer in a scalable manner using our web-based system.

References

1. National Research Council: A Framework for K-12 Science Education. National Academies Press, Washington, D.C. (2011)
2. NGSS Lead States: Next Generation Science Standards: For States, By States. The National Academies Press, Washington, DC (2013)
3. Kuhn, D.: Education for thinking. Harvard University Press, Cambridge, MA (2005)
4. Bereiter, C.: Education and Mind in the Knowledge Age. Lawrence Erlbaum Associates, Mahwah, NJ (2002)
5. Clarke-Midura, J., Dede, C., Norton, J.: The Road Ahead for State Assessments., Policy Analysis for California Education and Rennie Center for Educational Research & Policy, Cambridge, MA (2011)
6. de Jong, T., van Joolingen, W.: Scientific Discovery Learning with Computer Simulations of Conceptual Domains. Review of Educational Research 68, 179-201 (1998)
7. van Joolingen, W. R., de Jong, T.: Supporing Hypothesis Generation by Learners Exploring an Interactive Computer Simulation. Instructional Science 20(5-6), 389-404 (1991)
8. Schauble, L., Glaser, R., Raghavan, K., Reiner, M.: Causal Models and Experimentation Strategies in Scientific Reasoning. Journal of the Learning Sciences. 1(2)., 201-238 (1991)
9. Glaser, R., Schauble, L., Raghavan, K., Zeitz, C.: Scientific Reasoning Across Different Domains. In DeCorte, E., Linn, M., Mandl, H., Verschaffel, L., eds. : Computer-based Learning Environments and Problem-Solving. Springer-Verlag, Heidelberg, Germany (1991) 345-371
10. Buckley, B. C., Gobert, J., Horwitz, P.: Using Log Files to Track Students' Model-Based Inquiry. In : Proceedings of the 7th International Conference on Learning Sciences, Bloomington, IN, pp.57-63 (2006)
11. Schauble, L.: Belief revision in children: The role of prior knowledge and strategies for generating evidence. Journal of Experimental Child Psychology, 49, 31-57 (1990)
12. Brown, A. L., Campione, J. C.: Guided discovery in a community of learners. In : Classroom lessons: integrating cognitive theory and classroom practice. MIT Press, Cambridge, Massachusetts (1994)
13. Kirschner, P., Sweller, J., Clark, R.: Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching. Educational Psychologist, 41(2), 75-86 (2006)
14. Hmelo-Silver, C., Duncan, R., Chinn, C.: Scaffolding and Achievement in Problem-Based and Inquiry Learning: A Response to Krischner, Sweller, and Clark (2006). Educational Psychologist, 42(2), 99-107 (2007)
15. Chen, Z., Klahr, D.: All Other Things Being Equal: Acquisition and Transfer of the Control of Variables Strategy. Child Development 70(5), 1098-1120 (1999)
16. Klahr, D., Nigam, M.: The equivalence of learning paths in early science instruction: effects of direct instruction and discovery learning. Psychological Science, 15(10), 661-667 (2004)

17. Zohar, A., David, A.: Explicit Teaching of Meta-Strategic Knowledge in Authentic Classroom Situations. Metacognition Learning 3, 59-82 (2008)
18. Strand-Cary, M., Klahr, D.: Developing elementary science skills; Instructional effectiveness and path independence. Cognitive Development 23(4), 488-511 (2008)
19. Sao Pedro, M., Gobert, J., Heffernan, N., Beck, J.: Comparing Pedagogical Approaches for Teaching the Control of Variables Strategy. In : N.A. Taatgen & H. vanRijn (Eds.), Proceedings of the 31st Annual Meeting of the Cognitive Science Society, Amsterdam, Netherlands, pp.1294-1299 (2009)
20. Sao Pedro, M., Gobert, J., Raziuddin, J.: Comparing Pedagogical Approaches for the Acquisition and Long-Term Robustness of the Control of Variables Strategy. In Gomez, K., Lyons, L., Radinsky, J., eds. : Learning in the Disciplines: Proceedings of the 9th International Conference of the Learning Sciences, ICLS 2010, Volume 1, Full Papers, Chicago, IL, pp.1024-1031 (2010)
21. Kuhn, D., Schauble, L., M., G.-M.: Cross-Domain Development of Scientific Reasoning. Cognition and Instruction 9, 285-327 (1992)
22. Kuhn, D., Pease, M.: What Needs to Develop in the Development of Inquiry Skills? Cognition and Instruction 26(4), 512-559 (2008)
23. Dean Jr., D., Kuhn, D.: Direct Instruction vs. Discovery: The Long View. Science Education, 384-397 (2006)
24. Vygotsky, L. S.: Mind in Society: The Development of Higher Psychological Processes. Harvard University Press, Cambridge, MA (1978)
25. Wood, H. A., Wood, D. J.: Help Seeking, Learning and Contingent Tutoring. Computers and Education 33, 153-169 (1999)
26. White, B., Frederiksen, J.: Inquiry, Modeling and Metacognition: Making Science Accesible to All Students. Cognition and Instruction, 16(1), 3-118 (1998)
27. Koedinger, K., Suthers, D., Forbus, K.: Component-Based Construction of a Science Learning Space. International Journal of Artificial Intelligence in Education (IJAIED) 10, 292-313 (1999)
28. Sao Pedro, M. A., Baker, R. S. J. d., Gobert, J. D., Montalvo, O., Nakama, A.: 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 (2013)
29. Gobert, J., Sao Pedro, M., Baker, R., Toto, E., Montalvo, O.: Leveraging educational data mining for real time performance assessment of scientific inquiry skills within microworlds. Journal of Educational Data Mining 4(1), 111-143 (2012)
30. Sao Pedro, M., Baker, R., Gobert, J.: Improving Construct Validity Yields Better Models of Systematic Inquiry, Even with Less Information. In Masthoff, J., Mobasher, B., Desmarais, M., Nkambou, R., eds. : Proceedings of the 20th Conference on User Modeling, Adaptation, and Personalization (UMAP 2012), Montreal, QC, Canada, pp.249-260 (2012)
31. White, B. Y.: ThinkerTools: Causal Models, Conceptual Change, and Science Education. Cognition and Instruction 10(1), 1-100 (1993)
32. Levy, S. T., Wilensky, U.: Emerging Knowledge through an Emergent Perspective: High-school Students' Inquiry, Exploration and Learning in the Connected Chemistry Curriculum. In : Presented at the annual meeting of the American Educational Research Association, San Francisco, CA (2006, April 11)
33. McElhaney, K., Linn, M.: Helping Students Make Controlled Experiments More

Informative. In Gomez, K., Lyons, L., Radinsky, J., eds. : Learning in the Disciplines: Proceedings of the 9th International Conference of the Learning Sciences (ICLS 2010) - Volume 1, Full Papers, Chicago, IL, pp.786-793 (2010)

34. de Jong, T., van Joolingen, W., Giemza, A., Girault, I., Hoppe, U., Kindermann, J., Kluge, A., Lazonder, A., Vold, V., al., e.: Learning by creating and exchanging objects: The SCY experience. British Journal of Educational Technology 41(6), 909-921 (2010)
35. Quellmalz, E. S., Timms, M. J., Silberglitt, M. D., Buckley, B. C.: Science Assessments for All: Integrating Science Simulations into Balanced State Science Assessment Systems. Journal of Research in Science Teaching 49(3), 363-393 (2012)
36. Joyner, D. A., Majerich, D. M.., Goel, A. K.: Facilitating Authentic Reasoning about Complex Systems in Middle School Science Education. In : Proceedings of the 11th Conference on Systems Engineering Research, Atlanta, GA, vol. 16, pp.1043–1052 (2013)
37. Gobert, J., Schunn, C.: Supporting Inquiry Learning: A Comparative Look at What Matters. In : A symposium presented at the Annual Meeting of the American Educational Research Association. Chicago, IL, April 9-13 (2007)
38. Kuhn, D., Black, J., Keselman, A., Kaplan, D.: The Development of Cognitive Skills to Support Inquiry Learning. Cognition and Instruction 18(4), 495-523 (2000)
39. Gobert, J., Sao Pedro, M., Baker, R., Toto, E., Montalvo, O.: Leveraging educational data mining for real time performance assessment of scientific inquiry skills within microworlds. Journal of Educational Data Mining 4(1), 111-143 (2012)
40. de Jong, T.: Computer Simulations - Technological advances in inquiry learning. Science 312(5773), 532-533 (2006)
41. Anderson, J. R., Corbett, A. T., Koedinger, K. R., Pelletier, R.: Cognitive Tutors: Lessons Learned. The Journal of the Learning Sciences 4(2), 167-207 (1995)
42. Corbett, A. T., Anderson, J. R.: Knowledge-Tracing: Modeling the Acquisition of Procedural Knowledge. User Modeling and User-Adapted Interaction 4, 253-278 (1995)
43. Koedinger, K. R., Corbett, A. T.: Cognitive Tutors: Technology Bringing Learning Sciences to the Classroom. In : The Cambridge Handbook of the Learning Sciences. Cambridge University Press, New York (2006) 61-77
44. Aleven, V., Stahl, E., Schworm, S., Fischer, F., Wallace, R.: Help Seeking and Help Design in Interactive Learning Environments. Review of Educational Research 73(3), 277-320 (2003)
45. Aleven, V., Koedinger, K. R.: Limitations of Student Control: Do Students Know When They Need Help? In Gauthier, G., Frasson, C., VanLehn, K., eds. : Proceedings of the 5th International Conference on Intelligent Tutoring Systems, ITS 2000, Berlin, pp.292-303 (2000)
46. Mayer, R. E.: Should There Be a Three Strikes Rule Against Pure Discovery? The Case for Guided Methods of Instruction. American Psychologist 59(1), 14-19 (2004)
47. Rieber, L. P., Tzeng, S., Tribble, K.: Discovery Learning, Representation, and Explanation within a Computer-Based Simulation: Finding the Right Mix. Learning and Instruction 14, 307-323 (2004)
48. Rickel, J., Johnson, W. L.: Animated Agents for Procedural Training in Virtual reality: Perception, Cognition, and Motor Control. Applied Artificial Intelligence 13, 343-382 (1999)
49. Moreno, R.: Multimedia Learning with Animated Pedagogical Agents. In : The

Cambridge Handbook of Multimedia Learning. Cambridge University Press, New York (2005) 507-524

50. Lester, J. C., Towns, S. G., Callaway, C. B.: Cosmo: A Life-like Animated Pedagogical Agent with Deictic Believability. In : Working Notes of the IJCAI Workshop on Animated Interface Agents: Making Them Intelligent, Nagoya, Japan, pp.61-69 (1997)
51. Lester, J. C., Towns, S. G., Fitzgerald, P. J.: Achieving Affective Impact: Visual Emotive Communication in Lifelike Pedagogical Agents. International Journal of Artificial Intelligence in Education 10(3-4), 278-291 (1999)
52. Walker, J. H., Sproull, L., Subramani, R.: Using a Human Face in an Interface. In : CHI '94 Human Factors in Computing Systems, Boston, MA, pp.85-91 (1994)
53. Dehn, D., van Mulken, S.: The Impact of Animated Interface Agents: A Review of Empirical Research. International Journal of Human-Computer Studies 52(1), 1-22 (2000)
54. Mitrovic, A., Suraweera, P.: Evaluating an Animated Pedagogical Agent. Intelligent Tutoring Systems 1839, 73-28 (2000)
55. Sao Pedro, M.: Real-time Assessment, Prediction, and Scaffolding of Middle School Students' Data Collection Skills within Physical Science Simulations. Ph.D. Dissertation etd-042513-062949, Worcester Polytechnic Institution, Worcester, MA (2013)
56. Sao Pedro, M. A., Baker, R. S. J. d., Gobert, J. D.: What Different Kinds of Stratification Can Reveal about the Generalizability of Data-Mined Skill Assessment Models. In : Proceedings of the 3rd Conference on Learning Analytics and Knowledge, Leuven, Belgium, pp.190-194 (2013)
57. Gobert, J., Sao Pedro, M., Raziuddin, J., Baker, R.: From Log Files to Assessment Metrics for Science Inquiry using Educational Data Mining. Journal of the Learning Sciences 22(4), 521-563 (2013)
58. Baker, R. S. J. d., Corbett, A. T., Wagner, A. Z.: Human Classification of Low-Fidelity Replays of Student Actions. In : Proceedings of the Educational Data Mining Workshop held at the 8th International Conference on Intelligent Tutoring Systems, ITS 2006, Jhongli, Taiwan, pp.29-36 (2006)
59. Sao Pedro, M., Baker, R., Gobert, J.: Incorporating Scaffolding and Tutor Context into Bayesian Knowledge Tracing to Predict Inquiry Skill Acquisition. In D'Mello, S. K., Calvo, R. A., Olney, A., eds. : Proceedings of the 6th International Conference on Educational Data Mining, Memphis, TN, pp.185-192 (2013)
60. Sao Pedro, M., Gobert, J., Betts, C.: Towards Scalable Assessment of Performance-Based Skills: Generalizing a Detector of Systematic Science Inquiry to a Simulation with a Complex Structure. In : Proceedings of the 12th International Conference on Intelligent Tutoring Systems, Honolulu, HI (to appear)
61. Baker, R. S. J. d., Corbett, A. T., Koedinger, K. R., Evenson, E., Roll, I., Wagner, A. Z., Naim, M., Raspat, J., Baker, D. J., Beck, J.: Adapting to When Students Game an Intelligent Tutoring System. In Ikeda, M., Ashlay, K., Chan, T.-W., eds. : Proceedings of the 8th International Conference on Intelligent Tutoring Systems, ITS 2006, Jhongli, Taiwan, vol. LNCS 4053, pp.392-401 (2006)
62. Woolf, B. P., Arroyo, I., Cooper, D., Burleson, W., Muldner, K.: Affective Tutors: Automatic Detection of and Response to Student Emotion. In : Advances in Intelligent Tutoring Systems. Springer-Verlag, Berlin Heidelberg (2010) 207-227
63. Pea, R. D.: The Social and Technological Dimensions of Scaffolding and Related Theoretical Concepts for Learning, Education, and Human Activity. Journal of the

Learning Sciences 13(3), 423-451 (2004)

64. Siler, S., Klahr, D., Magaro, C., Willows, K., Mowery, D.: Predictors of Transfer of Experimental Design Skills in Elementary and Middle School Children. In Aleven, V., Kay, J., Mostow, J., eds. : Proceedings of the Tenth International Conference on Intelligent Tutoring Systems, ITS 2010, Pittsburgh, PA, vol. Part II, LNCS 6095, pp.198-208 (2010)
65. Heffernan, N. T., Turner, T. E., Lourenco, A. L. N., Macasek, M. A., Nuzzo-Jones, G., Koedinger, K. R.: The ASSISTment builder: Towards an analysis of cost effectiveness of ITS creation. In : Proceedings of the 19th International FLAIRS Conference, Melbourne Beach, Florida, USA, pp.515-520 (2006)
66. VanLehn, K., Lynch, C., Schulze, K., Shapiro, J., Shelby, R., Taylor, L., Treacy, D., Weinstein, A., Wintersgill, M.: The Andes physics tutoring system: Lessons Learned. International Journal of Artificial Intelligence and Education, 15(3), 1-47 (2005)
67. Kuhn, D.: What needs to be mastered in mastery of scientific method? Psychological Science, 16(11), 873-874 (2005)
68. Thorndike, E. L., Woodworth, R. S.: The Influence of Improvement in One Mental Function Upon the Efficacy of Other Functions. Psychological Review 8, 247-261 (1901)
69. Sao Pedro, M., Gobert, J., Baker, R.: Automated, Real-Time Scaffolding Helps Students Acquire and Transfer Data Collection Inquiry Skills. (in prep)
70. Sao Pedro, M. A., Baker, R. S. J. d., Gobert, J. D.: What Different Kinds of Stratification Can Reveal about the Generalizability of Data-Mined Skill Assessment Models. In : Proceedings of the 3rd Conference on Learning Analytics and Knowledge, Leuven, Belgium (2013)

Appendix

Pedagogical agent Rex's messages given when the system detects that students are not designing controlled experiments, testing their stated hypotheses, or engaging in haphazard inquiry.

<table>
<tr><th>Constraint</th><th>Triggered Scaffold Message</th><th>Help Button</th><th>Help Response</th></tr>
<tr><td rowspan="6">Not designing controlled experiments and not testing stated hypotheses</td><td rowspan="4">Level 1:
I think the data you're collecting won't help you test your hypothesis because you aren't designing a controlled experiment.</td><td>How do I do that?</td><td>Design a controlled experiment by changing only the variable you are testing while keeping all the other variables the same.</td></tr>
<tr><td>Which variable am I trying to test?</td><td>It's in your hypothesis. It says you want to test if changing the [IV] affects the [DV].</td></tr>
<tr><td>I need more help</td><td>Run trials where you:
(1) Change only the [IV], and
(2) Keep all the other variables the same.</td></tr>
<tr><td>Why do this?</td><td>Changing only the [IV] while keeping everything else the same lets you tell for sure if the [IV] affects the [DV].</td></tr>
<tr><td>Level 2:
Let me help you some more.

You said you wanted to test if changing the [IV] affects the [DV] in your hypothesis.

To do this, run pairs of trials where you:
(1) Change only the [IV], and
(2) Keep all the other variables the same.</td><td>Why do this?</td><td>Changing only the [IV] and keeping everything else the same lets you tell for sure if the [IV] affects the [DV].</td></tr>
<tr><td>Let me help you some more.

Just change the [IV] and run another trial. Don't change the other variables.

Doing this lets you tell for sure if changing the [IV] affects the [DV].</td><td></td><td></td></tr>
</table>

<table>
<tr><th>Constraint</th><th>Triggered Scaffold Message</th><th>Help Button</th><th>Help Response</th></tr>
<tr><td rowspan="5">Designing controlled experiments, but not testing stated hypotheses</td><td rowspan="3">Level 1:
It looks like you did great at designing a controlled experiment, but let me remind you to collect data to help you test your hypothesis.</td><td>How do I do that?</td><td>Keep designing a controlled experiment, but make sure to try different values of variable you're trying to test.</td></tr>
<tr><td>Which variable is that?</td><td>Your hypothesis says you wanted to test if changing the [IV] affects the [DV].</td></tr>
<tr><td>Why do this?</td><td>Changing the [IV] while keeping everything else the same lets you see how changing the [IV] affects the [DV].</td></tr>
<tr><td>Level 2:
Let me help again.

You said you wanted to test if changing the [IV] affects the [DV] in your hypothesis.

Keep designing controlled experiments, and collect data for different values of the IV.</td><td>Why do this?</td><td>Changing the [IV] while keeping everything else the same lets you see how changing the [IV] affects the [DV].</td></tr>
<tr><td>Level 3:
Let me help some more.

Just change the [IV] and run another trial. Don't change the other variables.

Doing this lets you tell for sure if changing the [IV] causes changes to the [DV].</td><td></td><td></td></tr>
</table>

Constraint	Triggered Scaffold Message	Help Button	Help Response
Testing stated hypotheses, but not designing controlled experiments	Level 1: I see you're collecting data about the [IV], but you can't test your hypothesis because you **aren't designing a controlled experiment**.	How do I do that?	Design a controlled experiment by **changing only the variable you are testing** while **keeping all the other variables the same**.
		Which variable am I trying to test?	It's in your hypothesis. It says you want to test if changing the **[IV]** affects the **[DV]**.
		I need more help	Run trials where you: (1) Change only the **[IV]**, and (2) Keep all the other variables the same.
		Why do this?	Changing only the [IV] while keeping everything else the same lets you tell for sure if the [IV] affects the [DV].
	Level 2: Let me help again. You said you wanted to test if changing the **[IV]** affects the **[DV]** in your hypothesis. To do this, run pairs of trials and **change only the [IV]**. **Keep all the other variables the same** for the second trial.	Why do this?	Changing only the [IV] and keeping everything else the same lets you tell for sure if the [IV] affects the [DV].
	Level 3: Let me help you some more. Just change the **[IV]** and run another trial. Don't change the other variables. Doing this lets you tell for sure if changing the [IV] affects the [DV].		