Impacts of students’ experimentation using a dynamic visualization on their understanding of motion

Kevin W. McElhaney & Marcia C. Linn,
University of California at Berkeley, 4523 Tolman Hall MC1670, Berkeley, CA 94720, USA
Email: kevin777@berkeley.edu, mclinn@berkeley.edu

Abstract: This study examines how students’ experimentation with a dynamic visualization contributes to their understanding of science. We designed a week-long, technology-enhanced inquiry module on car collisions. The module uses new technologies that log student interactions with the visualization. Physics students (N=148) in six diverse high schools studied the module and responded to pretests, posttests, and embedded assessments. We scored students’ experimentation using three methods: total number of trials, how widely students changed variables between trials (variability), and how well students connected content knowledge to experimentation strategies (validity). Students made large, significant overall pretest to posttest gains. Regression models showed that validity was the strongest predictor of learning when controlling for prior knowledge and other experimentation measures. Successful learners employed a goal-directed experimentation approach that connected their experimentation strategy to content knowledge.

Introduction

Can computer-based visualizations aid students’ understanding of dynamic processes, particularly those that are too fast to be directly observed? We investigate this question by studying how students experiment with a dynamic visualization of motion and what they learn from this experience. We study a week-long inquiry module titled *Airbags: Too Fast, Too Furious* (henceforth *Airbags*) that features a visualization illustrating the motion of an airbag and the driver of a car during a collision. The module uses new technologies that log student interactions with the visualization to connect experimentation and learning.

Airbags addresses concerns that dynamic visualizations may overwhelm learners with their complexity (Tversky, Morrison, & Betrancourt, 2002) or disengage cognitive processes required for learning (Hegarty, Kriz, & Cate, 2003). The design of *Airbags* is based on the knowledge integration framework and design principles emerging from empirical studies (Kali, 2006). *Airbags* uses a technology-enhanced learning environment to guide students’ interactions with visualizations, connect visualizations to students’ everyday experiences, and provide opportunities for discussion and reflection. The module promotes active integration of visualizations with students’ prior knowledge rather than encouraging passive observation, so students in diverse school settings can achieve more coherent understandings of motion. The design of *Airbags* has been iteratively refined based on classroom trials.

This study documents the impact of the *Airbags* curriculum as a whole and connects the validity of students’ experiments (as logged by the learning environment) with the quality of learning. We address two research questions: (1) How does an inquiry-based investigation help physics students connect their understanding of motion, motion graphs, and the dynamics of car collisions? and (2) How do students’ experimentation knowledge and practice contribute to their scientific understanding?

Rationale

This research contributes to understanding of how best to scaffold inquiry investigations to enhance students’ learning of complex science (Quintana et al., 2004). Studies on modeling environments such as Model-It (e.g. Spitulnik, Krajcik, and Soloway, 1999) show that student-initiated investigations of complex science topics enhance learning because participants engage in inquiry activities such as identifying and relating system variables, conducting experiments, and building explanations. Learning environments such as ThinkerTools (White & Frederiksen, 1998), which features dynamic visualizations of force and motion, show the benefit of interactive simulations combined with opportunities for reflection. Other environments such as WorldWatcher (Edelson, Gordin, & Pea, 1999) and Kids as Global Scientists (Lee & Songer, 2003) make complex science topics accessible to students by building on their everyday experiences. The *Airbags* design extends these findings by demonstrating how student-initiated experiments with a dynamic visualization can help students integrate their ideas about motion.

This study also examines how students’ experimentation choices contribute to their understanding of science content. Research has focused on children’s ability to isolate variables during experimentation (Inhelder & Piaget, 1958; Tschirgi, 1980; Klahr & Nigam, 2004), incorporate domain-specific understanding into experimentation (Linn, Clement, & Pulos, 1983; Schauble, 1996), and benefit from experimentation within a context of authentic scientific inquiry (Lehrer, Schauble, & Petrosino, 2001). We extend this work by documenting students' experimentation practices and examining how these practices contribute to integrated understanding of the physics of motion in the real world context of airbag safety.

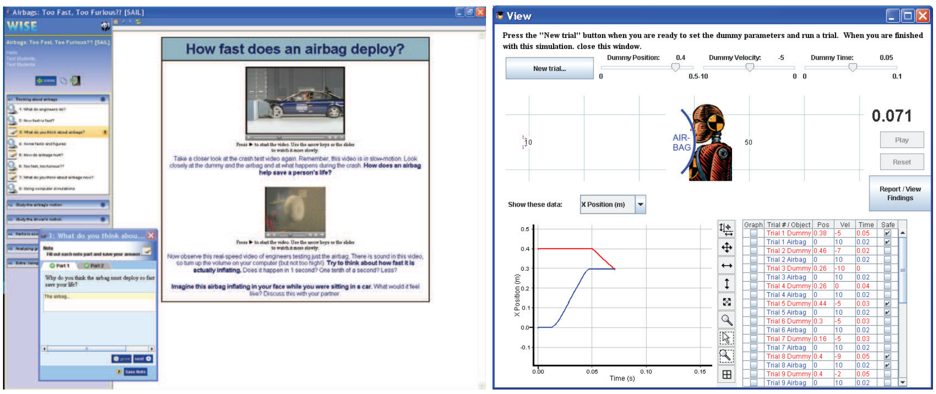
The *Airbags* module is designed using the knowledge integration perspective to promote coherent understanding (Kali, 2006; Linn & Eylon, 2006). The knowledge integration perspective describes learners as simultaneously holding multiple, sometimes conflicting, and often isolated ideas about science (Linn & Hsi, 2000). Research has identified instructional design patterns that help students develop normative understandings of science through four knowledge integration processes (Linn & Eylon, 2006). These processes include eliciting students' current ideas to ensure that new ideas connect to existing knowledge, introducing new ideas to help students distinguish among their alternatives, developing criteria for evaluating ideas and connections, and sorting out ideas to identify contradictions and revise understanding. The design of *Airbags* deploys these processes in varied patterns to meet the needs of diverse learners. In particular, experimenting with the visualization helps students add new ideas about motion and graphs, develop criteria for conducting valid experiments, and sort out ideas according to the evidence they produce with their experiments.

Airbags design

We designed *Airbags*, a one-week curriculum module for high school physics classes, using the Web-based Inquiry Science Environment (WISE) software (Linn, Davis, & Bell, 2004). This paper discusses the third design iteration of the *Airbags* module. The learning goals of the module are two-fold. First, students explore the relationship between the nature of one-dimensional motion and the characteristics of position and velocity graphs. Second, students investigate factors that lead to a high risk for injury to the driver from an airbag. In *Airbags*, students connect their observations of crash-test videos to their understanding of dynamic visualizations. They respond to embedded prompts that ask them to articulate their current ideas, interpret graphs, generate explanations, and reflect on previous work. They conduct experiments to investigate the role of motion variables on the safety of drivers in collisions. At the end, students recommend design improvements to cars and airbags to make them safer. Figure 1a shows a screenshot of *Airbags'* introductory activity.

In *Airbags* students interact with five dynamic visualizations. The first four visualizations help students link their understanding of the graphs to animated representations of the airbag's and driver's motion during a collision within a *predict-observe-compare-explain* design pattern. Students view an animation of the airbag or driver and *predict* the appearance of graphs representing their motion. Students then *observe* the animation alongside computer-generated graphs. They *compare* the observed graph with their predictions and develop criteria for distinguishing the two. Then they *explain* the relationship between graph characteristics and motion.

This paper focuses on students' experimentation with the fifth visualization (shown in Figure 1b). Students investigate three questions concerning airbag safety: (1) Why are shorter drivers at greater risk for injury than taller drivers? (2) Are drivers at greater risk for injury in high speed or low speed collisions? and (3) How does the car's ability to "crumple" during the collision affect the driver's risk for injury? These three questions map directly onto the three motion variables students can manipulate in the visualization: (1) the initial position of the driver, (2) the velocity of the driver toward the airbag after impact, and (3) the time that elapses after impact and before the driver's initial motion toward the airbag. Students conduct experiments to answer each investigation question. In order to conduct an experimental trial, students must first select an investigation question from a drop down menu (or indicate they are just exploring the visualization). Next, students specify the values of the position, velocity, and time variables. Finally, students run the crash simulation and judge whether the trial was "safe" or "unsafe." In previous activities, students determine that a driver must encounter an airbag after it has finished inflating in order to be "safe." Students conduct as many trials as they choose in order to answer the investigation questions.



(a) (b)

Figure 1. (a) The first activity of *Airbags*. (b) The experimentation visualization.

Airbags provides moderate scaffolding for students as they conduct investigations. Unlike model-building curricula using software such as STELLA (e.g. Mandinach & Cline, 1994) or Model-It (e.g. Spitulnik, Krajcik, & Soloway, 1999), in *Airbags* students investigate questions and variables identified by the authors of the module. Students have more choices than in other interventions designed to help students master the control-of-variables strategy (e.g. Klahr & Nigam, 2004). For instance, students must map investigation questions onto the experimentation variables and plan their own experimentation strategies in discussions with a partner or teacher. Thus, *Airbags* seeks to optimize the instructional benefits of both structure and individual initiative.

Methods

We used a pretest-posttest design combined with embedded assessments (explanation prompts and experimentation data logs) within *Airbags*. We measured progress in developing coherent understanding of the physics of motion and airbag safety.

Participants

There were six implementations of *Airbags* with 148 high school physics students in diverse settings (see Table 1). Three of the teachers were experienced and had taught previous versions of *Airbags*. All teachers participated in targeted professional development (Varma, Husic, & Linn, in press). Most students worked in dyads on the activities. Unpaired students worked on their own computers and engaged in discussions with a student dyad at another computer. In all six implementations, every student taking physics at the school participated in this study; some schools had low enrollment in physics.

Table 1: Summary of *Airbags* classroom implementations.

School	# students	# classes	Classroom/school description
1	38	2	Honors ability, wide geographical area, gifted science students
2	15	1	Mixed ability, suburban, 18% reduced lunch, ethnically diverse
3	28	1	Mixed ability, urban, 67% reduced lunch, ethnically diverse
4	12	1	Mixed ability, suburban, 31% reduced lunch, 61% African-American
5	9	1	Mixed ability, urban, 54% reduced lunch, 95% African-American
6	46	3	Mixed ability, suburban, 52% reduced lunch, ethnically diverse

Assessments and scoring

Pretest/posttest assessments. Pretests and posttests were administered the day before the start of implementation and the day after completion. Posttests addressed similar issues as the pretest items but were changed slightly to reduce possible gains due to retesting, except in School 3 where the tests were identical. Due to absences, some students did not take either the pretest or the posttest. Pretests and posttests consisted of seven items (six GRAPHING items and the AIRBAGS item). The GRAPHING items were scored from zero to four and measured students' ability to interpret and construct position and velocity graphs. GRAPHING items did not address motion specifically in the context of airbag safety. The AIRBAGS item was scored from zero to five and measured students' understanding of the risks for injury to drivers from airbags. A high score on the AIRBAGS item required students to connect conceptually three separate ideas in a valid explanation and thus captured multiple aspects of students' understanding of the airbag deployment. Students in School 3 did not receive the AIRBAGS item. We scored pretest and posttest items using a knowledge integration rubric (Linn, Lee, Tinker, Husic, & Chiu, 2006) that rewards valid scientific connections between concepts. The total pretest and posttest scores were the sum of the scores from the individual items.

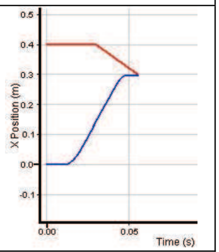
Embedded prompts. We scored responses to 12 embedded prompts as outcome measures of student learning. Six items (INVESTIGATIONS) asked students immediately before (predictions) and after (explanations) conducting their experiments to answer the three investigation questions. These items measured whether students' understanding of the investigation questions improved from conducting experiments. Six items (INTERPRETATIONS) asked students to describe motion that could produce graphs like the ones in the visualization. These items measured students' ability to interpret the graphs as they used the visualization and occurred just after the experimentation activity. Responses to the embedded prompts were scored using a knowledge integration rubric. Table 2 provides examples of pretest, posttest, and embedded assessment items.

Experimentation data logs. Pedagogica software (Buckley, Gobert, & Horwitz, 2006) logged the investigation question and variable values students chose for each trial. We used the reports of students' trials to score each student groups' experimentation strategy in three ways:

- *Total trials.* We measured the number of times students used the visualization by computing the total number of trials each group conducted. Because some students occasionally conducted identical trials multiple times, we also computed the number of unique trials each group conducted. Unique trials correlated highly (r = .95) with total trials, so we used total trials in the analysis.

- *Trial variability*. To measure how widely students changed the variables between trials, we computed a variability score. We computed for each of the three investigation variables (position, velocity, and time) (1) the number of *unique values* tested as a fraction of the maximum number tested by any student, (2) the *range of values* tested as a fraction of the total allowable range for that variable, and (3) the number of *boundary values* (minimum or maximum allowable values) tested as a fraction of the total possible number of boundary values. We computed the mean of the unique value, range, and boundary value fractions to generate a subscore for each investigation variable, then computed the mean of these three subscores to generate the overall variability score scaled from zero to 100. The three subscores exhibited an internal consistency (Cronbach's α) correlation of .91, suggesting that the mean of the subscores provides a reliable overall measure of the variability of students' experimentation.
- *Experimentation validity*. We measured the extent to which students conducted valid experiments by employing a control-of-variables experimentation strategy that was consistent with the investigation question they chose for each trial. Only trials where students selected one of the three investigation questions were used for this score. We scored the experimentation sequences three times (once for each investigation question) on a scale of zero to five using a knowledge integration rubric (Table 3) that rewards consistency between the investigation question and the variable choices. The overall validity score was the mean of the subscores for the three investigation questions. The three subscores exhibited an α correlation of .71, demonstrating that the validity of students' experiments was fairly uniform across the three investigation questions.

Table 2: Examples of pretest, posttest, and embedded assessment measures.

Name (type)	Example
GRAPHING (pretest/posttest)	A car starts at point A and speeds up at a constant rate until it reaches point B in 4 seconds, where it suddenly stops. The car waits at point B for 2 seconds. It then travels at constant speed in the opposite direction, reaching point A again in another 3 seconds. Sketch a **POSITION-TIME graph** and a **VELOCITY-TIME graph** of the motion during these 9 seconds.
AIRBAGS (pretest/posttest)	Two identical cars are traveling 10 mph in a parking lot and collide head-on. Airbags in both cars deploy. The driver of the car on the left is a 5'3", 165 lb. adult male. The driver of the car on the right is a 5'11", 120 lb. adult female. Which driver do you think is **more likely to be injured** by an airbag deploying? Explain your answer.
INVESTIGATIONS (embedded)	[Asked as predictions and explanations.] Why are shorter drivers at greater risk for injury from an airbag than taller drivers?
INTERPRETATIONS (embedded)	[Refers to the graph.] Describe what happened between the driver and airbag in this crash. Was the driver injured by the airbag? Explain based on the graph. 

Analysis

In School 1, a subset of students' experimentation records failed to upload to the servers. Students at this school reported how many experimentation trials they conducted in an in-class survey. Six student workgroups (12 students) at this school whose self-reports differed obviously from the incomplete uploaded information were removed from analysis. Eleven student workgroups (19 students) at all the schools who failed to respond to at least 75% of the modules' prompts due to class absences were also removed.

We used two-tailed, paired t-tests to measure learning gains from pretest to posttest and from INVESTIGATIONS predictions to explanations. Because the school samples were small and students worked in dyads, the data grossly violate the assumptions of equal standard deviations and normality for an analysis of covariance. We therefore could not use school as a covariate. We pooled the students from all schools and employed multiple linear regression models to relate learning outcomes to experimentation measures, controlling for prior knowledge using either pretest scores or responses to embedded prompts that occurred before experimentation.

Table 3: Knowledge integration (KI) rubric for scoring experimentation validity.

KI level	Score	Description
blank	0	students do not conduct any trials
none	1	students conduct exactly one trial
invalid/ isolated	2	students change all three variables between trials OR hold the investigation variable constant
partial	3	students change exactly two variables between trials, including the investigation variable
basic	4	students change only the investigation variable between trials that produce the same outcome (either safe or unsafe)
complex	5	students change only the investigation variable between trials that produce opposite outcomes (safe/unsafe) OR students conduct two separate sets of controlled trials

Note: Rubric is applied to each group's experimentation sequence three times, once for each investigation variable.

Results and discussion

Teachers implemented *Airbags* as intended. Students found the module engaging and responded to the visualizations as intended.

Overall learning gains

Students who took both the GRAPHING and AIRBAGS subtests made large, significant pretest to posttest gains (M = 16.51, SD = 5.68 pre; M = 20.12, SD = 4.75 post), $t(89) = 7.77$, $p < .001$, $d = .69$. Students' gains were significant on both the GRAPHING subtest (M = 13.64, SD = 5.93 pre; M = 16.62, SD = 4.24 post), $t(108) = 6.37$, $p < .001$, $d = .58$ and the AIRBAGS subtest (M = 2.06, SD = .96 pre; M = 3.76, SD = 1.17 post), $t(89) = 11.78$, $p < .001$, $d = 1.6$. Considering that *Airbags* typically requires just 4-5 hours of class time, the positive learning gains attest to the success of the module in helping students understand motion graphs and the dynamics of airbag deployment. Gains were significant ($p < .05$) for students at each school except for School 5 (which had just 9 students take both pretest and posttest and lacked statistical power), illustrating the success of *Airbags* in promoting gains in understanding across diverse and authentic instructional settings. Results indicate that even students with very high levels of prior content knowledge gained insights about the applications of physics to a relevant socio-scientific issue such as airbag safety.

In addition, students made significant improvements in INVESTIGATIONS from predictions to explanations (M = 2.44, SD = .54 predictions; M = 2.77, SD = .76 explanations), $t(119) = 5.24$, $p < .001$, $d = .50$. These gains reflect students' improvements in explaining how physical characteristics of the driver and the car put a driver at risk of being injured by an airbag. Gains are attributable mainly to students' experimentation with the visualization, as INVESTIGATIONS occurred immediately before and after experimentation. Table 4 includes examples of the progressions three student groups made. Group A progresses from an unrealistic conception of relative distances within a car to a complex understanding of the relationship among a driver's height, sitting position, and risk for injury. Group B begins with the belief that the risk for injury from an airbag is measured relative to the risk from other factors, but acquires a more normative (though partial) understanding of the specific relationship between speed and risk. Group C progresses from having vague understanding of how car crumpling affords time for the driver to providing a specific mechanism for the relationship between crumpling and risk for injury.

Distinguishing experimentation variability and validity

Student groups made very different experimentation choices as reflected in the variability and validity scores. We compare the experimentation sequences of a low variability, high validity group (Group D) and a high variability, low validity group (Group E) by plotting the variable values each group chose for each trial they conducted (Figure 2). This graphical representation makes trial variability apparent by illustrating the range of values students explored, how often the students changed the values between trials, and whether students tested boundary values. Experimentation validity is reflected in whether students hold two variables constant while varying the appropriate investigation variable between two trials (shown in bold).

Group D's below average variability score (39.6) reflects the relatively small number of unique values they chose, the small range of values they tested for the position variable, and their failure to test many boundary values. This group did successfully conduct two controlled tests of an investigation variable in Trials 8 through 11. These students might have intended to conduct a controlled test for the time variable between Trials 1 and 2 or the position variable between Trials 5 and 6, but they did not adequately communicate their intention to do so by selecting the appropriate investigation question for each trial. Group D's high validity score (4.00) reflects an ability to conduct controlled experiments consistent with their investigation goals. Group D's strategy illustrates how students could conduct valid experiments without widely varying the variables between trials.

Group E's above average variability score (67.8) reflects the wide range of values (including four boundary values) they tested. Though their choices vary more widely than those of Group D, the plot shows that Group E either changed every variable or held every variable constant between consecutive trials and did not conduct any controlled trials. Group E's low validity score (1.66) reflects their failure to investigate the individual variables in their trials. Group E's strategy illustrates how students could vary the variables widely without conducting valid experiments.

Table 4: Students' responses to INVESTIGATIONS (predictions and explanations).

Group	Investigation question	Prediction (KI score)	Explanation (KI score)
A	Why are shorter drivers at greater risk for injury from an airbag than taller drivers?	"Shorter people might not reach the airbag, maybe fall lower than the airbag, or even suffocate." (2)	"The average short person has shorter arms and must move closer to the steering wheel. Thus, they can be able to hit a steering wheel faster than an airbag has time to deploy." (4)
B	Is a driver more likely to be injured by an airbag in a high speed or low speed collision? Explain.	"The driver is more likely to be harmed in a low speed collision when the airbag deploys because the force of the airbag deploying on to them is greater than the force of the actual crash." (2)	"If you are in a high speed crash you are more at risk of hitting the airbag before it completely inflates. In a low speed collision you can be hit by an airbag if u are too close" (3)
C	How do you think a car's ability to crumple affects a driver's risk for injury from an airbag?	"We think a car's ability to crumple affects a driver's risk for harm from an airbag is that it gives the driver more time for the airbag to inflate." (3)	"The longer the car took to crumple the more safe the driver is because it gave the airbag more time and space to inflate before the driver started to move." (4)

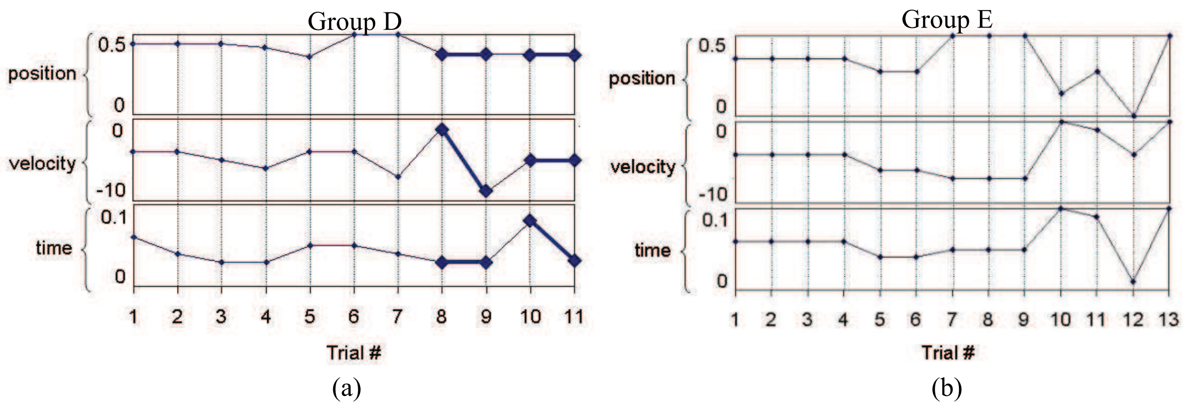


Figure 2. Experimentation sequences for two student groups: (a) low variability, high validity and (b) high variability, low validity. Trials where students conducted controlled trials appropriate for their chosen investigation question are shown in bold.

Impact of experimentation on learning outcomes

We investigated the relationships between each of the three experimentation scores and the GRAPHING posttest score, controlling for GRAPHING pretest score. Multiple linear regression models revealed marginally significant positive relationships for total trials ($\beta = .14$, $p = .077$) and validity ($\beta = .18$, $p = .054$) and a significant positive relationship for variability ($\beta = .20$, $p = .014$). In all three models, pretest scores were a much stronger predictor of posttest scores than the experimentation score. The rather weak relationships between experimentation and the GRAPHING posttest scores are not surprising, as students spent a small fraction of their total time on *Airbags* using the visualization (usually about 30 minutes of experimentation out of 4-5 hours using the module). Students had many opportunities to improve their knowledge of graphing other than the experimentation activity, such as the *predict-observe-compare-explain* activities, graph interpretations, and reflection prompts that promote connections between physics and real-world events.

We next investigated the relationships between the three experimentation scores and learning outcomes related to the airbags context (INVESTIGATIONS explanations, INTERPRETATIONS, and the AIRBAGS posttest). We generated a multiple linear regression model for each learning outcome using prior knowledge, total trials, trial variability, and experimentation validity as predictors. We used INVESTIGATIONS predictions as a predictor for INVESTIGATIONS explanations, the GRAPHING pretest as a predictor for INTERPRETATIONS, and the AIRBAGS pretest as a predictor for the AIRBAGS posttest.

Table 5 lists the linear regression coefficients. The regression models reveal that experimentation validity was the strongest predictor of all three context-specific learning outcomes. Experimentation validity was a significant positive predictor for all three outcomes, controlling for the other experimentation scores and prior knowledge. Variability was a significant positive predictor only for the INVESTIGATIONS explanations, while total trials was a marginally significant *negative* predictor for the INVESTIGATIONS explanations and INTERPRETATIONS. The standardized coefficients (β) indicate that experimentation validity was a much stronger predictor of context-specific understanding than students' prior knowledge, even when controlling for the other experimentation scores.

Table 5: Summary of regression analysis for predicting learning outcomes related to the airbags context.

Learning outcome	Predictor	B	$SE\ B$	β	R^2
INVESTIGATIONS (explanations) (N = 114)	INVESTIGATIONS (predictions)	.38	.12	.27**	.43
	total trials	-.01	.007	-.19	
	trial variability	.007	.003	.26*	
	experimentation validity	**.22**	**.05**	**.43*** **	
INTERPRETATIONS (N = 114)	GRAPHING (pretest)	.12	.06	.17*	.42
	total trials	-.01	.007	-.20	
	trial variability	.001	.003	.05	
	experimentation validity	**.30**	**.05**	**.63*** **	
AIRBAGS (posttest) (N = 90)	AIRBAGS (pretest)	.15	.12	.12	.20
	total trials	.003	.02	.03	
	trial variability	.003	.006	.08	
	experimentation validity	**.26**	**.10**	**.34** **	

* $p < .05$; ** $p < .01$; *** $p < .001$

Results indicate that students benefited from the abilities to connect their experimentation strategy to the *Airbags* investigation context and to employ a goal-directed experimentation approach. A high validity score reflects several dimensions of students' knowledge other than just being able to control variables. First, students must map the inquiry questions onto the appropriate experimentation variables. Students cannot interpret controlled comparisons if they do not understand what the variables correspond to in a real car crash. Second, students must correctly interpret the outcomes of their trials as safe or unsafe. Otherwise, students cannot identify when a controlled comparison produces an effect on the outcome of the trial. Third, because students must articulate an investigation goal prior to conducting each trial, the validity score measures students' ability to plan investigations in advance, as post-hoc comparisons between trials do not produce a high validity score. Though the findings do not imply a causal relationship between experimentation choices and learning, they do highlight the importance of students' abilities to connect experimentation strategies to investigation goals and to plan in advance.

The weak relationship between trial variability and learning outcomes and the negative relationship between total trials and learning outcomes suggest that unplanned or haphazard experimentation approaches are ineffective for learning. Even though post-hoc analysis of a large number of haphazard trials could provide enough evidence for drawing valid conclusions, students who conduct valid experiments using fewer trials and communicate their investigation goals in advance learn more from *Airbags*. The findings suggest ways to scaffold experimentation for students who struggle to understand the visualization. Asking students to plan investigations in advance may strengthen connections between experimentation strategies and investigation goals and help students understand why controlling variables is necessary to reach valid conclusions. Just-in-time hints for students whose early trials appear haphazard could encourage students to reflect on what they can conclude from their early trials and consider how to design subsequent trials to better address the investigation questions.

Conclusions and Implications

Students' work with *Airbags* resulted in overall gains in understanding across diverse instructional settings. This finding demonstrates the effectiveness of scaffolding students' use of dynamic visualizations within an authentic inquiry investigation to improve scientific understanding. Logging of student interactions with the visualization provides insights into how students conduct experiments and the ways they interpret the evidence. This study shows that the number and variability of trials students conduct are less important than advance planning and connections between experimentation strategies and investigation goals. Though we cannot establish a causal relationship between experimentation choices and learning, this study shows that students who conduct valid experiments also learn more from the module.

These results raise several issues. First, knowing how to control variables is necessary but not sufficient for exploring the visualization and reaching valid conclusions. In *Airbags*, students must also incorporate disciplinary knowledge (such as interpreting graphs and understanding the nature of motion inside the car) into designing and interpreting their experiments in order to be successful. Second, students' ability to connect the visualizations to real-life situations and how this understanding influences students' use of visualizations merits further study. Finally, these results suggest sophisticated ways to use new logging technologies to understand students' science reasoning. Questions for future study include how software can use data on students' interactions with visualizations to provide prompts or hints for students who need guidance, or how teachers could take advantage of logging data to guide whole classes or specific individuals. These questions extend research on how students use visualizations from the laboratory into authentic classroom environments.

References

Buckley, B., Gobert, J. & Horwitz, P. (2006). Using Log Files To Track Students' Model-based Inquiry. In S. A., Barab, K. E Hay & D. T. Hickey (Eds.) *Making a difference: Proceedings of the Seventh International Conference of the Learning Sciences* (pp. 57-63). Mahwah, NJ: Erlbaum.

Edelson, D., Gordin, D. & Pea, R. (1999). Addressing the Challenges of Inquiry-Based Learning Through Technology and Curriculum Design. *Journal of the Learning Sciences, 8* (3&4), 391-450.

Hegarty, M., Kriz, S., & Cate, C. (2003). The role of mental animations and external animations in understanding mechanical systems. *Cognition and Instruction, 21*, 325-360.

Kali, Y., (2006). Collaborative knowledge-building using the Design Principles Database. *International Journal of Computer Support for Collaborative Learning, 1*(2), 187-201.

Klahr, D. & Nigam, M. (2004). The equivalence of learning paths in early science instruction: Effects of direct instruction and discovery learning. *Psychological Science, 15*(10), 661-667.

Lee, H.-S., & Songer, N. B. (2003). Making authentic science accessible to students. *International Journal of Science Education, 25*(8), 923-948.

Lehrer, R., Schauble, L., & Petrosino, A.J. (2001). Reconsidering the role of experiment in science education. In K. Crowley, C. Schunn, & T. Okada (Eds.), *Designing for science: Implications from everyday, classroom, and professional settings* (pp. 251-277). Mahwah, NJ: Erlbaum.

Linn, M. C., Clement, C., & Pulos, S. (1983). Is it formal if it's not physics? *Journal of Research in Science Teaching, 20*, 755-770.

Linn, M. C., Davis, E. A., & Bell, P. (Eds.). (2004). *Internet Environments for Science Education*. Mahwah, NJ: Erlbaum.

Linn, M.C. & Eylon, B.-S. (2006). Science Education: Integrating Views of Learning and Instruction. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of Educational Psychology, 2nd edition* (pp. 511-544). Mahwah, NJ: Erlbaum.

Linn, M. C., & Hsi, S. (2000). *Computers, Teachers, Peers: Science Learning Partners*. Mahwah, NJ: Erlbaum.

Linn M. C., Lee, H.-S., Tinker R., Husic F., Chiu J. L. (2006). Teaching and assessing knowledge integration in science. *Science. 313*, 1049–1050.

Mandinach, E. B., & Cline, H. F. (1994). *Classroom dynamics: Implementing a technology-based learning environment*. Hillsdale, New Jersey: Erlbaum.

Schauble, L. (1996). The development of scientific reasoning in knowledge-rich contexts. *Developmental Psychology. 32*(1), 102-119.

Spitulnik, M. W., Krajcik, J., and Soloway, E. (1999). Construction of models to promote scientific understanding. In Feurzeig, W., and Roberts, N. (Eds.), *Modeling and Simulation in Science and Mathematics Education*, Springer-Verlag, New York, pp. 70–94.

Tversky, B., Morrison, J. B., & Betrancourt M. (2002). Animation: Can it facilitate? *International Journal of Human-Computer Studies, 57*, 247-262.

Varma, K., Husic, F., & Linn, M. (in press). Targeted support for using technology-enhanced science inquiry modules. *Journal of Science Education and Technology.*

White, B.Y., & Frederiksen, J. R. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition & Instruction, 16*(1), 3–118.