Title
Making Controlled Experimentation More Informative in Inquiry Investigations

Author
McElhaney, Kevin

Publication Date
2010

Making Controlled Experimentation More Informative in Inquiry Investigations

by

Kevin Wei Hong McElhaney

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Education

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Marcia C. Linn, Chair
Andrea A. diSessa
Michael J. Clancy

Spring 2010

Making Controlled Experimentation More Informative in Inquiry Investigations

by Kevin Wei Hong McElhaney

Abstract

Making Controlled Experimentation More Informative in Inquiry Investigations

by

Kevin Wei Hong McElhaney

Doctor of Philosophy in Education

University of California, Berkeley

Professor Marcia C. Linn, Chair

This dissertation incorporates three studies that examine how the design of inquiry based science instruction, dynamic visualizations, and guidance for experimentation contribute to physics students' understanding of science. I designed a week-long, technology-enhanced inquiry module on car collisions that logs students' interactions with a visualization. Students studied the module and responded to pretests, posttests, and embedded prompts that assessed students' understanding of motion graphs and collisions. In Study 1, students (N=148) made large, significant overall pretest to posttest gains. Regression models showed that the propensity for students to conduct controlled trials 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. Study 2 investigated the effect of limiting students' experimentation on their planning, strategies, and learning outcomes. Students (N=58) made large, significant overall pretest to posttest gains. Students constrained to twelve trials isolated variables in their experiments better than the unconstrained students. However, the constrained students significantly underperformed the unconstrained students on the module assessments, indicating that isolating variables during experimentation did not lead to improved learning outcomes. In Study 3, students (N=166) were assigned to conditions that prompted them either to isolate or compare variables. Both groups made moderate, significant pretest to posttest gains. Students in the *compare* treatment used more diverse experimentation strategies than students in the *isolate* treatment. *Compare* students made nuanced interpretations of collision events based on threshold values. Case studies illustrate how comparing rather than isolating variables helped students use wide-ranging strategies to reach complex insights. The findings illustrate how students can benefit from experimentation strategies that do not isolate variables. Spontaneous exploration can help students test new questions that arise from unexpected results, informing the design of controlled tests that better reveal subtle characteristics of the variables. Guidance that encourages students to compare rather than isolate variables may have important benefits, such as prompting students to search for distinctions among the variables. The findings have important implications for the design of inquiry-based science instruction.

To my parents, who have made all my accomplishments possible.

Contents

List of Figures

List of Tables

Acknowledgments

First and foremost I thank my advisor and mentor Marcia Linn. Her guidance and support throughout my time as graduate student, and particularly my time as a new father, have been invaluable.

Many thanks to my dissertation and orals committee members Andy diSessa, Michael Clancy, and Sophia Rabe-Hesketh for their guidance during the process.

I wish to thank my past and present colleagues in the Linn research group for their valuable discussions about my work, as well as their friendship: Kathy Benemann, Mandy Bliss Jonathan Breitbart, Janet Casperson, Jennifer King Chen, Jennie Chiu, Stephanie Corliss, Libby Gerard, Tara Higgins, Jeff Holmes, Freda Husic, Doug Kirkpatrick, Suparna Kudesia, Jacquie Madhok, Norma Ming, Kelly Ryoo, Elissa Sato, Beat Schwendimann, Vanessa Svihla, Erika Tate, Tammie Visintainer, Lindsay Wells, Helen Zhang, and Tim Zimmerman.

I also wish to thank members of the extended TELS research community for their valuable input into my work: Stephen Bannasch, Ken Bell, Jane Bowyer, Hsin-Yi Chang, Raj Chaudhury, Doug Clark, Chris Hoadley, Paul Horwitz, Yael Kali, Hee-Sun Lee, Ji Shen, Jim Slotta, Robert Tinker, Keisha Varma, and Michelle Williams.

I wish to thank members of the WISE/TELS development team for their hard work in making the tools I needed to conduct this research: Turadg Aleahmad, Matt Fishbach, Geoffrey Kwan, Tony Perritano, and Hiroki Terashima.

Thank you to all teachers and students who were instrumental in the success of my curriculum materials.

Special thanks go to David Crowell for helping me keep everything in order.

I thank the Spencer Foundation for their support during the dissertation year.

Finally, I thank my family for their love and support, without whom all my accomplishments would be meaningless: my parents Ronald and Therese, my sister Christine, my wife Christin, and my son Lucas.

Chapter 1: Introduction & Research Questions

Citizens are regularly misled by persuasive messages concerning everyday science. Advertisements and political messages about important socioscientific issues such as health care and energy policy can take advantage of laypersons' naïve views of science. A recent example concerns a study published in a British medical journal suggesting a significant relationship between the measles, mumps, and rubella (MMR) vaccine and the incidence of autism (Wakefield et al., 1998). Subsequent research failed to identify evidence for such a connection, and the original paper was eventually retracted from the journal based on flawed methodology and conflict of interest. Despite the lack of scientific evidence, high media publicity of the initial study contributed to wide-spread belief (particularly in the United Kingdom) that risks of the vaccine outweighed its benefits. This belief led to decreased vaccination rates (McIntyre & Leask, 2008) and sharp increases in the incidence of both measles and mumps in the UK (Asaria & MacMahon, 2006; Gupta, Best, & MacMahon, 2005). The MMR vaccine controversy points strongly to the importance of scientific literacy among the general population.

Research examining students' beliefs about the nature of science is consistent with the reluctance of many citizens to consider the nature of scientific knowledge (Carey & Smith, 1993). For example, Carey, Evans, Honda, Jay, and Unger (1989) conducted interviews with seventh grade students that elicited their understanding of the nature, purpose, and practice of science. The interviews revealed that the majority of students made no distinctions between scientific ideas and practices and identified the goals of science as discovering facts about the world and inventing things. A common belief among students was that a scientist "tries it to see if it works." (Carey, et al., 1989, p. 520)

Traditional science instruction does little to augment students' views about the nature of science. Lectures cast teachers as authorities on scientific knowledge rather than participants in a community-wide endeavor. Knowledge assessments often measure students' ability to recall facts and solve quantitative problems, characterizing the nature of scientific knowledge as definitive rather than tentative. Textbooks often highlight prominent discoveries while minimizing the evolution of the ideas and processes of argumentation that led to these discoveries. Laboratory activities frequently resemble recipes in their emphasis on procedures that lead to predetermined outcomes rather than modeling authentic inquiry. These practices reinforce the idea that science necessarily leads to the "right answer" and that the goal of science is to achieve these answers.

In efforts to reform science education, national standards (American Association for the Advancement of Science, 1993; National Research Council, 1996) call for increased emphasis on inquiry-based approaches to science. These standards aim to heighten students' understanding of the nature of science and scientific knowledge by engaging students in more authentic science investigations. These investigations allow students to engage in authentic scientific inquiry as they pose research questions, generate hypotheses, use experiments to test ideas, gather and evaluate evidence, analyze data, construct arguments, participate in debate, and critique the work of peers. These activities can help make classroom science more closely resemble professional science, possibly leading to more normative views of the nature of science as well as a stronger understanding of science content.

Science education research reveals authentic scientific inquiry to be much more difficult for students than for professional scientists. Studies identify difficulties students have with every stage of the inquiry process. Students struggle to identify research questions (White & Frederiksen, 1998), fail to generate a sufficient range of hypotheses (Klahr & Dunbar, 1988), conduct experiments only to confirm prior beliefs (Dunbar, 1993), ignore anomalous data (Chinn & Brewer, 1993), neglect to support their claims with evidence (Sandoval & Millwood, 2005), and lack sophisticated criteria for evaluating the ideas they encounter (Clark & Slotta, 2000). That the complexity of full-fledged scientific investigations tends to overwhelm science novices is not surprising, despite the potential richness of inquiry activities.

Computer-supported learning environments show promise in guiding students through the complexity of science investigations (Quintana et al., 2004). Software tools can provide students with the support they need to interpret data, document findings, and link claims to evidence. Environments can provide students with access to real-world, real-time data via the world-wide web and help elucidate patterns in these data. Communication tools can facilitate collaboration between students across the globe. Some environments support teacher customization so that investigations can be made more relevant to specific school communities. Current research efforts continue to seek effective ways to support students in these complex activities. Technology-based inquiry can also incorporate sophisticated visualizations of scientific processes. Dynamic visualizations can bring unobservable phenomena such as molecular interactions or planetary motion to life. Visualizations allow students to explore situations that would be impossible to explore with physical materials. Visualization also support student-initiated investigations using virtual experiments or model-building environments. Future research needs to generate design principles for combining complex visualizations and support for inquiry so that investigations are coherent.

My interest in furthering these strands of science education research extends from my experiences as a scientist and a classroom teacher. After leaving professional science and upon embarking on a career as a science teacher, I hoped to bring my experience as a scientist to the classroom to enrich traditional approaches to science instruction. As a science teacher, I found that most of the prepared curriculum materials that were readily available to me neither engaged students in authentic science nor addressed topics that were especially relevant to students' interests or experiences. Though I was occasionally able to develop my own curriculum materials that could accomplish these goals, the day to day demands of teaching precluded me from devoting significant time toward this end. This dissertation represents my continued effort toward bringing authentic and relevant inquiry experiences to science students as part of their regular classroom instruction.

My dissertation research examines how guidance for computer-based experimentation activities can improve students' learning outcomes from inquiry investigations and how students use dynamic visualization tools to conduct experiments. I have designed a week-long, computer-based inquiry module for use in high school physics classes, titled *Airbags: Too Fast, Too Furious?* (henceforth *Airbags*). *Airbags* helps physics students integrate their understanding of motion and graphs during an investigation of the safety of airbags in car collisions. The *Airbags* module and the dynamic visualizations have been iteratively refined based on classroom trials so that learners can use them effectively. *Airbags* is delivered by a computer-based learning environment that guides students' interactions with visualizations and provides opportunities for discussion and reflection. My research takes advantage of software that can log students' explanations, drawings, and interactions with the visualizations so that I can analyze detailed information about their inquiry activities. This dissertation aims to extend previous research on the design of computer-based inquiry instruction to incorporate visualizations and to extend laboratory studies on student learning from visualizations and experimentation to classroom settings.

Research questions

My dissertation research consists of three successive classroom studies using *Airbags*: a pilot study and two design-based comparison studies. This dissertation addresses three overarching research questions concerning the design of inquiry instruction and visualizations, the relationship between students' experimentation strategies and learning outcomes, and the design of guidance to promote informative experimentation.

First, I investigate the overall effectiveness of the *Airbags* module's design:

> *How can a week-long module that incorporates a dynamic experimentation environment be designed to support physics students' understanding of motion and graphs?*

Airbags makes use of empirical design patterns and principles that aim to help students integrate their scientific ideas, incorporates a virtual experimentation environment that allows students to test their own ideas. What is the overall impact of airbags on students' understanding of motion and graphs? How can instruction build on students' everyday understanding of science? How can instruction help students integrate their ideas by highlighting key distinctions between concepts? What design patterns support the use of dynamic visualizations for science learning?

Second, I investigate the relationship between how students conduct experiments and what they learn from their experimentation:

> *How does students' experimentation with a dynamic visualization featuring animated and graphical representations contribute to their ability to interpret and construct graphs of motion and use physics to generate explanations of real-world phenomena?*

Experimentation encourages students to use their domain knowledge to test their own ideas. This research examines students' experimentation strategies and links them to students' understanding of motion graphs and the investigation context of airbag safety. What is the relationship between students' experimentation strategies and their learning? How do students use their prior knowledge about physics and the investigation context to design and interpret experiments? What is the role of controlling variables in helping students generate narrative accounts of real-life science?

Third, I investigate the impact of different forms of guidance for experimentation on student learning:

> *How can curriculum materials guide students' experimentation with visualizations to improve their understanding of motion and graphs and the investigation context?*

Carefully designed guidance may be able guide students toward more informative experimentation strategies and deeper insights about the investigation context. This research examines the effect of different forms of guidance on students' scientific understanding. How can guidance for experiments highlight key concepts? What is the effect of constraining students’ experimentation on their strategies and learning? What can students learn by comparing rather than isolating variables?

Chapter 2: Theoretical and Empirical Foundations

The primary foci of this dissertation are instructional design principles, patterns, and guidance for science instruction and their relationship to students' science learning. This foundations chapter will therefore mainly address research on the design of environments and tools for science instruction. However, robust design frameworks must necessarily extend from theoretical perspectives on the nature of knowledge and learning. For this reason, I begin this chapter by briefly summarizing the constructivist views of learning that inform my design approach. I then proceed to outline in greater detail how empirical studies have extended theoretical perspectives to designing computer-based inquiry environments, visualization tools, and guidance for experimentation.

View of the learner

Ideas originate in distinct contexts and from experience

Learners possess a wide range of scientific ideas. Learners acquire their diverse array of ideas from many places such as everyday observations and classroom experiences. For example, Howe (1998) examined the ideas of students aged six to 15 about buoyancy. Interviews revealed that typical students held between five and 15 distinct ideas about buoyancy. These ideas covered a wide range of objects' properties such as shape, texture, temperature, substance, and direction of orientation. Clark and Linn (2003) found students to hold a similarly diverse array of ideas about heat and temperature during interviews spanning five years. Furthermore, many of these ideas contradict each other. A student called Cedar articulates in his initial interview that metal and wood objects in the same oven should be at the same temperature. When prompted to explain why objects at the same temperature feel different, he immediately revises his view and states that the objects are actually at different temperatures. Clark and Linn found Cedar's views to vary depending on the context of the discussion (e.g. inside a hot car or a cold ski cabin). Like many other students in the study, Cedar did not attempt to connect ideas unless prompted to do so.

Research on conceptual change (e.g. Strike & Posner, 1985) aimed to identify situations where students replaced old ideas with new ideas. In these situations, when students encountered new ideas, they recognized the limitations of their old ideas and abandoned them in favor of the alternative. Other research however indicates that ideas learners generate from repeated observations and personal experience are powerful and strongly resistant to change. Shipstone (1985) illustrates students' non-normative views on direct current circuits. Students explain lighting a light bulb with a battery as the "clashing" of two currents, or as the light bulb consuming current from the battery. These non-normative views reflect common everyday observations such the release of energy from collisions or dying batteries. Students held these views despite receiving instruction about electricity during the year they were assessed. This study and studies in other domains such as thermodynamics (e.g. Lewis, 1996) and force and motion (e.g. diSessa, Elby, & Hammer, 2002) illustrate the persistence of students' intuitive ideas in the face of formal instruction.

Furthermore, research shows that students hold contradictory ideas they acquire within school and outside of school simultaneously, and that students sequester these two views of science from each other (Gilbert & Boulter, 2000). Students may believe in science class that objects in motion remain in motion, but that in real life objects gradually slow to a stop. diSessa's well known account of a university student called "J" illustrates this phenomenon in striking fashion (diSessa, et al., 2002; diSessa & Sherin, 1998). In the interview, J struggles to explain how Newton's second law of motion (F = ma) describes a book being pushed across a table at constant velocity. J was unable to reconcile her intuitive idea that the motion of the book must result from unbalanced forces with her formal knowledge that unbalanced forces must produce an acceleration. Despite that she successfully applied the law F = ma to solving problems from her physics class, she chose her intuitive idea as the normative explanation and cast F = ma aside as one of "those darn equations" that "aren't applicable to every single thing." (diSessa & Sherin, 1998, p. 1184) J's explanation demonstrates the power that learners' intuitive ideas have in governing learners' understanding of science and the degree to which multiple learning contexts can lead learners to hold contradictory ideas of the same phenomenon.

Learners' knowledge is fragmented and incoherent

Another interview with J shows another instance in which J misapplies her intuitive ideas of unbalanced forces. Asked to describe the physics of tossing a ball into the air, J used normative accounts of force and energy (presumably as she has learned in physics class), clearly indicating the presence of only one force (gravitational) on the ball while it is in the air. However, asked to explain what happens at the peak of the toss, J invoked the naïve "impetus" theory (McCloskey, 1983), referring to a force imparted by the hand that gradually dies away until overcome by the gravitational force. J's selective application of the impetus theory suggests her ideas on the toss lack coordination and originate from multiple sources.

diSessa refers to the incoherent nature of naïve learners' ideas as *knowledge in pieces* (diSessa, 1988). diSessa coined the term *phenomenological primitive*, or p-prim, to describe conceptions that learners possess concerning force and other physical phenomena. P-prims reflect learners' experiences with the natural world. diSessa posits that learners have hundreds or thousands of p-prims, such as "force as mover" or "actions die away." These p-prims represent intuitive explanations of experiential phenomena, are grounded in specific contexts, and are loosely connected to other ideas. P-prims are not "wrong" in the sense that they accurately describe learner's experiences with the physical world, but they are often misapplied. For instance, J appears to invoke some combination of the "force as mover" and "actions die away" p-prims in her account of the first half of the ball toss—as the ball slows on its way up, the net force that moves the ball upward must be gradually dying away. Her views are in some sense correct, however, in that a net upward force was responsible for the ball's initial upward movement, and that the ball's momentum (rather than a force), could be said to be dying away as the ball approaches its peak.

Minstrell (2001), inspired by p-prims, identified what he called "facets" of students' thinking about physics. Facets are pieces of knowledge or reasoning that can be organized in "clusters" by their proximity to normative ideas. Minstrell illustrates facets with some examples concerning the nature of air pressure, such as "air is light" or an apparent relationship between the vacuum and weightlessness in outer space. In contrast to p-prims, facts are emergent ideas that may or may not originate from everyday experience. Similar to p-prims, learners misapply facets or generalize them to inappropriate contexts.

The *scaffolded knowledge integration perspective* (Linn & Eylon, 2006; Linn & Hsi, 2000) emphasizes the *repertoire of ideas* learners hold about scientific phenomena as well as their naïve views about the nature of science. These ideas may concern the ways scientific investigation methods contribute to scientific knowledge, the nature of scientific knowledge itself, and the role of scientists in society. Learners commonly hold beliefs about the nature of science that can interfere with their ability to make their ideas cohere. For instance, citizens often ignore mediating variables when examining the relationship between two correlating events, or see scientists as absolute authorities on knowledge rather than investigators who hold conflicting ideas. Instruction can help students connect their understanding of the domain to the nature of science, improving learning outcomes.

Learning as developing coherence of ideas

Traditional science instruction usually involves at most three steps (Stigler & Hiebert, 1999). In some cases instruction aims to *motivate* students to acquire new ideas. Next traditional instruction *adds normative ideas* to students' repertoire. Finally, teachers (and standardized assessments) *assess* how well students can recall these new ideas. This three-step approach fails to provide students with opportunities to distinguish new ideas from old ideas or reconcile contradictions between prior intuitions and new formalisms. As a result, learners isolate new, normative ideas from their intuitive understanding and prior experience, resulting in brittle understanding that students find difficult generalizing to new contexts.

Learners' tendency to retain old ideas even when instruction adds new ideas points to the need to give learners opportunities to distinguish, evaluate, prioritize, and sort out the full array of old and new ideas they encounter in science. Though traditional instruction rarely helps learners make their ideas cohere, well designed instruction can help learners achieve a durable and generalizable understanding of science. For example, sixth grade students who used the ThinkerTools curriculum (White, 1993) to engage in a model-based inquiry of force and motion concepts outperformed high school physics students taught using traditional methods. Clark and Linn (2003) describe how Cedar connected his diverse ideas and observations into a more cohesive account of heat flow, insulation and conduction, and thermal equilibrium over the course of a computer-supported inquiry module on thermodynamics.

The *knowledge integration* perspective describes learning as occurring when students articulate their everyday ideas and intuitions then add new, normative ideas about science to the mix. Instruction then prompts students to bump these ideas up against one another, giving students the opportunity to connect ideas and resolve conflicts. Other activities can help students monitor their own understanding so that they can identify and repair gaps in their knowledge. In this way, new knowledge is anchored to prior educational and personal experiences. The knowledge integration perspective leads to an instructional design framework that aims to take advantage of the variation in students' ideas in order to help learners achieve integrated understanding of science. I describe this framework in more detail in the next section.

Design of inquiry-based science instruction

Scaffolding inquiry

The knowledge integration perspective synthesized the constructivist view of the learner and research on the design of instruction into four metaprinciples that take advantage of the repertoire of ideas to help learners achieve coherent understanding of science (Clark & Linn, 2003; Linn & Hsi, 2000). Instruction should *make science accessible* by centering instruction around compelling and relevant topics in order to better leverage students' prior ideas and interests. Instruction should *make thinking visible* by encouraging written articulation of ideas, providing data management tools, and incorporating visual representations of phenomena. Instruction should *encourage collaboration* to make use of the collective array of ideas students have provide students with opportunities to exchange feedback with their peers. Finally, instruction should *promote autonomy* by helping students to monitor the quality of own learning. The knowledge integration design framework also identified pragmatic design principles that would support each of these metaprinciples. For instance, *encouraging students to investigate personally relevant problems* can make science more accessible to learners, *providing visual representations of phenomena* can make thinking visible, *designing social activities* can encourage collaboration, and *engaging students as critics of scientific information* can promote autonomy.

Quintana et al. (2004) synthesized several design approaches such as scaffolded knowledge integration, problem-based inquiry (Kolodner et al., 2003), principles of learner centered design (Quintana, Krajcik, & Soloway, 2001) as well as the cognitive load perspective (Chandler & Sweller, 1991) into a single design framework for computer supported inquiry. Though unlike the knowledge integration framework it does not extend any single specific view of the learner, it was successful in organizing design principles across many research studies according to three elements of inquiry investigation: sense-making, process management, and articulation and reflection. The framework developed by Quintana et al. identifies twenty scaffolding strategies used across the range of research on inquiry instruction that support these three inquiry components.

Knowledge integration processes

The remainder of this section focuses on the knowledge integration instructional design framework, which forms the basis for the design of the curriculum module I developed for this dissertation. Constructivist views of the learner and empirical research on the design of science instruction point to the benefits of engaging students in four interrelated knowledge integration processes to help them connect ideas in their repertoire (Linn & Eylon, 2006). First, instruction should *elicit ideas* that students have about the topic of study. Instruction can elicit students' ideas about the discipline and related personal experiences, as some ideas are likely to be grounded in these contexts. This step helps ensure that new ideas are not isolated from prior knowledge. Eliciting ideas takes advantage of the variety of ideas that learners have. Instruction can elicit ideas in many ways. Prompting students for predictions about phenomena before making observations can improve outcomes (Crouch, Fagen, Callan, & Mazur, 2004; Linn & Songer, 1991). Benchmark lessons (Minstrell, 2001) and reflective discourse (van Zee & Minstrell, 1997) elicit ideas from students as part of group discussions.

Next, instruction should *add normative ideas* to students' repertoires. Though traditional instruction adds ideas using lectures and texts, research illustrates ways that instruction can carefully add ideas in ways that students can more easily integrate into their understanding. Minstrell's benchmark lessons use demonstrations and class discussions to add relevant ideas. Bridging analogies (Clement, 1993) present students with analogs that are incrementally more distant from an anchoring situation that is easy to understand, and where students can make comparisons between successive situations. Pivotal cases (Linn, 2005) present students with comparative situations that exhibit key differences in order to highlight important concepts.

Traditional instruction often ends at this point, assessing learners' ability to recall new ideas. Often instructors use multiple choice assessments, which can be insensitive to the coherence of students' ideas (Linn, Lee, Tinker, Husic, & Chiu, 2006) and can give students the illusion of deep understanding. At this point, students may hold ideas that are contradictory or generalized to inappropriate contexts. Instruction must help students make their old and new ideas cohere.

Third, instruction should encourage learners to *distinguish their ideas*. Faced with a mix of old and new ideas, students must distinguish productive and relevant ideas from unproductive and irrelevant ones. Students must distinguish ideas from one context and fro other contexts. Making distinctions often requires students to evaluate their ideas. Pivotal cases (Linn, 2005) prompt students to make comparisons in order to distinguish one situation from another. Critique activities can encourage students to distinguish between normative and non-normative explanations or investigation methods (Zhang, 2010).

Finally, instruction must allow students to *reflect* on their understanding, so that they can identify inconsistencies and gaps in their own understanding. Generating explanations (Chi, De Leeuw, Chiu, & LaVancher, 1994) and self-assessment (Chiu & Linn, 2008; White & Frederiksen, 1998) can lead students to recognize these gaps and revisit their ideas in efforts to repair these gaps.

Instructional activities

Many activities can guide students through the four knowledge integration processes. Here I summarize four common approaches to inquiry-based instruction and outline the ways that they elicit and add ideas, highlight distinctions, and promote reflection.

Scientific modeling. Modeling environments such as Model-It (Spitulnik, Krajcik, & Soloway, 1999; Stratford, Krajcik, & Soloway, 1998), Virtual Solar System (Barab, Hay, Barnett, & Keating, 2000) and ThinkerTools (White & Frederiksen, 1998) ask students to build, test, and revise models that explain phenomena. Modeling elicits students' ideas by prompting students to identify system variables or asking students to make predictions about the roles of these variables. Testing models adds ideas about the effect variables have on the system. Students make distinctions by evaluating their models and comparing their observations to predictions. Students reflect on their ideas by refining their models to better describe observations and explaining the results of the modeling process.

Design tasks. Instruction can provide students with design tasks, to which students must apply domain knowledge in order to solve. Instructional frameworks that take advantage of design include Learning-for-Use (Edelson, 2001), Design Based Science (Fortus, Dershimer, Krajcik, Marx, & Mamlok-Naaman, 2004) and Learning by Design™ (Kolodner, et al., 2003). Design tasks elicit ideas through brainstorming, both individually and as design teams. Students add ideas through research and testing of prototypes. Students must distinguish ideas as they evaluate their designs and compare design approaches. Students reflect on their understanding as they refine and improve their designs based on feedback.

Community-based learning. Collaboration among students takes advantage of the breadth of ideas that exist in a community of learners, such as in Fostering Communities of Learners (FCL, Brown & Campione, 1996) or Computer Supported Intentional Learning Environments (CSILE, Scardamalia & Bereiter, 1994). Community-based learning efforts elicit and add ideas through mutual exchange of ideas and peer teaching. Students also add ideas by doing research or consulting experts in their chosen topic. Students must make distinctions by evaluating and weighing the contributions of individuals toward the common goal or with consequential tasks that require students to choose specific ideas to apply toward problem solutions. Students reflect on ideas by refining documents and other artifacts for publication or presentation to the rest of the community, and by incorporating feedback from other community members.

Inquiry investigations of socio-scientific issues or controversies. Inquiry environments such as BGuILE (Reiser et al., 2001), Kids as Global Scientists (Lee & Songer, 2003), and WISE (Linn, Davis, & Bell, 2004), scaffold the complex process of scientific investigation for students. Inquiry investigations elicit ideas by focusing on issues relevant to students' everyday lives and by eliciting hypotheses or predictions. Experimentation or other data collection methods add ideas to the repertoire. Experiments also help students make distinctions between variables or situations. Students reflect on their knowledge by constructing arguments based on the results of investigations or applying ideas to new situations.

Some environments synthesize approaches in ways that effectively combine successful features. ThinkerTools combines modeling activities within an inquiry cycle, allowing students to use experiments to test their models. Many inquiry environments incorporate aspects of community based learning to leverage the full range of ideas held by all students in a classroom. The combination of approaches to engage students in the four knowledge integration processes has led to the identification of instructional patterns (Linn & Eylon, 2006), which I discuss in the next section.

Instructional Design Patterns for Knowledge Integration

Linn and Eylon (2006) synthesized the research on the design of science instruction into *design patterns*. Patterns are instructional methods or activities that engage learners in the four knowledge integration processes. Each pattern focuses strongly on one or two processes, though some may promote all four, depending on how they are used. Instructional designers can combine multiple patterns in effective sequences to promote knowledge integration. The patterns can help designers create effective science instruction by taking advantage of design knowledge from the broader educational research and design communities. Patterns give designers a framework that ensures instruction provides students with opportunities to build on prior knowledge, distinguish between new and old ideas, intentionally refine their knowledge, and generalize their ideas to broader contexts.

In addition to helping students integrate ideas about the scientific domain, most patterns also engage learners in authentic science, provide learners with agency, and lend legitimacy to their views. Patterns provide an alternative to traditional science instruction where students serve only as recipients of knowledge that is comprised of a collection of facts. Patterns that reflect authentic methods of scientific inquiry emphasize the tentative nature of scientific knowledge and the authority of students as well as teachers in contributing to community knowledge. The patterns can thus help to illustrate normative views of scientific inquiry and epistemology.

Here I discuss four important patterns that research identifies as effective in science instruction and ways they can engage students in the knowledge integration processes: *orient and elicit*; *predict, observe, and explain*; *conduct and experiment*; *construct an argument*. During the five-year evolution of its design, *Airbags* has made use of each of these four patterns. I discuss specifically how *Airbags* incorporates these patterns in the following chapter on curriculum design.

Orient and elicit. The *orient and elicit* pattern helps learners familiarize themselves and articulate their initial ideas about the investigation context. Learners' initial ideas may reflect everyday experiences with the context, beliefs about a current issue, or content from previous coursework. The pattern sets the stage for subsequent instruction to connect new ideas to these prior ideas, allowing the investigation to extend these ideas to new situations or contrast new situations with current views.

Many learning environments for science make use of the orient and elicit pattern near the beginning of a curricular unit to motivate students and create demand for learning. The "Jasper" series (Cognition and Technology Group at Vanderbilt, 1992) used videos as an *anchoring event* to provide context for a mathematics investigation. The BGuILE Struggle for Survival unit (Reiser, et al., 2001) begins with *staging activities* that involve brainstorming about beliefs and understanding of island ecosystems. The Virtual Solar System project (Barab, et al., 2000) uses *seed questions* to elicit ideas around which students can begin to generate solar system models. This way, models can extend students' everyday conception of moon phases or seasons. The Design-Based Science framework (Fortus, et al., 2004) begins each design cycle with contextualization, which makes design problems significant to students and provides a point of entry for students' initial designs.

Predict, observe, and explain. The *predict, observe, and explain* pattern takes advantage of the positive effect of predictions in helping learners understand what they observe and of explanations in helping learners synthesize their ideas. Crouch et al. (2004) found science demonstrations promoted better learning outcomes when preceded by predications and followed with peer discussions. The pattern works by eliciting learners' initial ideas with predictions, illustrating a phenomenon to add normative ideas that may conflict with expectations, and finally prompting learners for explanations that must reconcile the conflict.

Inquiry instruction makes use of predict, observe, and explain when learners add new ideas by observing phenomena or collecting data. For example, in the Create-a-World curriculum (Edelson, 2001) students are given blank maps of the world and instructed to use crayons to draw a "best guess" about world temperatures in July. Students later compare their guesses to real temperature data, calling attention to gaps in their knowledge about climate. Students then participate in group discussions about their observations of the data. In the Computer as Lab Partner thermodynamics curriculum (Linn & Songer, 1991), students make predictions about the cooling curve for a container of water, observe a real cooling curve, then record observations and use the data to make new predictions. ThinkerTools (White & Frederiksen, 1998) prompts students to predict the motion of an object, observe this motion, then generate a motion law that explains the motion.

The *predict, observe, and explain* pattern may vary or be combined with other patterns suitable for the investigation context. For instance, rather than observing a phenomenon, students may explore a simulation or conduct experiments (Jackson, Stratford, Krajcik, & Soloway, 1994). Explanations may also prompt students to construct arguments as part of a debate based on observed outcomes of investigations (Bell & Linn, 2000).

Conduct an experiment. Experimentation activities can guide students through the full range of knowledge integration processes. Experiments elicit ideas though hypotheses or predictions. Experiments add ideas by introducing system variables and their relationship to outcomes. Learners distinguish ideas by comparing multiple trials to each other, or by comparing observed outcomes to expectations. Finally, experiments help students revisit their ideas as they explain their findings or use results as evidence for arguments.

The experiment pattern is frequently and easily combined with most other patterns. Lehrer and Schauble (2004) provide an interesting example in a curriculum unit for fifth graders on plant growth. Students used experiments to examine the effect of sunlight and fertilizer on plant growth. In addition to eliciting ideas in the form of hypotheses and adding ideas about plant biology, the curriculum was centered on the generation of representations that could capture the distribution of plant geometry and on sampling methods. Experiments thus provided a means for young students to explore not only science domain knowledge but also the representation and analysis of data, ideas central to the nature of scientific methods. Combining experimentation with the creation of artifacts that represent the experimental outcomes and argumentation about these artifacts resulted in a rich inquiry curriculum.

Construct an argument. Arguments require learners to bring multiple sources of evidence together, such as domain knowledge, observations and outcomes from experiments, research from the world-wide web, ideas from peers, and everyday conceptions of science. Because of the range of ideas learners can use in constructing arguments, generating argument span all the knowledge integration processes. Arguments can elicit new ideas and build on learners' everyday conceptions of science. Arguments require learners to distinguish among multiple pieces of evidence and choose those that best support their views. Arguments can also highlight gaps in learners' knowledge when all the evidence needed to support a view is not available. Argumentation can be particularly effective following modeling or experimentation, requiring students to carefully consider the evidence they gather during investigations experimentation (Stratford, et al., 1998). Opportunities to reflect on or critique arguments can help strengthen the links between claims and evidence (Clark & Sampson, 2007).

Many students struggle to construct strong arguments (Driver, Newton, & Osborne, 2000). Well designed software tools can scaffold the process, helping students support views with appropriate evidence. In the BGuILE curriculum, an explanation constructor (Sandoval & Reiser, 2004) helps students incorporate data as evidence for their arguments. Clark and Sampson (2007) use principle-building software to form heterogeneous students discussion groups based on students' prior views of a heat flow, encouraging students to consider multiple perspectives when making arguments. SenseMaker (Bell & Linn, 2000) allows students to sort evidence according to their support for various theories (or their irrelevance to the investigation). Students use their arguments in a subsequent debate activity.

Other important knowledge integration patterns that *Airbags* does not prominently feature include *create an artifact*, *illustrate ideas,* and *critique*. These patterns are a particularly important for curricula focused on design. The Design-Based Science and Learning By Design™ approaches incorporate these patterns in brainstorming, prototyping, constructive feedback, and cycles of revision, though they are also useful for inquiry based investigations.

One main goal of this dissertation is to explore ways that instruction can modify and combine patterns to support the use of dynamic visualizations. The driving questions behind *Airbags* address a topic that would be virtually impossible to investigate using physical materials. *Airbags* uses visualizations to illustrate the motion of the airbag and driver during head-on collisions. How can the patterns be adapted to take advantage of the affordances of visualizations? What sequence of patterns can help students overcome learning difficulties associated with dynamic visualizations and integrate their ideas? The next section of this chapter discusses research on the design of dynamic visualizations in science instruction.

A second goal of this dissertation is to examine what patterns support experimentation within a broader inquiry investigation. In *Airbags* students must connect their experimental designs and the outcomes of their experiments to ideas about motion, graphs, and car collisions. How can experimentation be combined with other patterns to help students make these connections? The final section of this chapter discusses research on learning from experimentation and the design of instruction that can make experimentation more informative.

The next chapter of this dissertation goes into detail about the specific design of the *Airbags* curriculum and how I used the design patterns to incorporate a virtual experimentation environment within the *Airbags* investigation.

Using Dynamic Visualizations to Enhance Science Learning

Computer-based visualization tools show promise for science education with their ability to illustrate dynamic phenomena. Visualizations can represent the temporal evolution of events in ways that static media cannot. Visualization tools have the potential to help learners understand scientific representations by linking them to familiar representations or real-life events. Perhaps most importantly, visualizations can allow students to interact with learning environments and test their own ideas. Learners can take advantage of the interactivity of visualizations to initiate scientific investigations or create virtual artifacts.

Visualization tools have many advantages over physical materials. Visualizations can illustrate phenomena that occur too quickly, too slowly, are too large or too small to directly observe or to examine in instructional settings. Inquiry environments can take advantage of visualizations to enable learners to conduct virtual investigations about complex topics such as water quality, climate change, or home insulation. Virtual design tools can support rapid prototyping for design tasks, allowing learners to quickly test and refine their ideas. Simulations of real world phenomena can give learners more control over their testing environments, such as the ability to turn friction or gravity on or off to illustrate their effects on motion. Research suggests that physical and virtual materials are similarly effective for mediating scientific investigations (Klahr, Triona, & Williams, 2007; Triona & Klahr, 2003), further arguing for their potential benefits in science classrooms.

Though research rarely questions the importance of dynamic visualizations to professional science, research is divided on the use of visualization tools for science instruction. Like many other aspects of professional science, visualizations can confuse and overwhelm novice learners who lack the knowledge to benefit from complex visualizations. Research studies point to several reasons for the difficulties learners have with animated graphics compared to static graphics. First, learners with poor visual processing ability may struggle to understand dynamic visualizations (Mayer & Sims, 1994; Yang, Andre, & Greenbowe, 2003). Dynamic visualizations may thus cognitively overload students with both the amount and complexity of information in dynamic visualizations. Second, series of static images may require learners to actively integrate information and generate mental models, resulting in better learning outcomes (Hegarty, Kriz, & Cate, 2003) than with animations. Third, dynamic visualizations of complex processes may give learners the illusion of understanding, discouraging them from seeking to repair gaps in their knowledge (Chiu & Linn, 2008). In this way visualizations may be "deceptively clear".

The nature of learners' difficulties with visualizations suggests that they can be addressed with either additional instruction or improvements in design. The cognitive load perspective on learning (Chandler & Sweller, 1991) suggests certain design principles for visualizations stemming from three sources of cognitive load (Chandler, 2004). Extraneous cognitive load is determined by activities that are not directly related to learning. Intrinsic cognitive load reflects the complexity of the topic and the prior knowledge of the learner. Germane cognitive load is generated by activities associated with the construction and automation of knowledge in long term memory. Eliminating extraneous features is a design principle that is not specific to science learning or even visualizations—these principles are helpful for designers of any software program to achieve efficiency and usability. Germane features on the other hand reflect constructivist views of learning in science domains and are relevant to design principles for visualizations used for science instruction.

Design principles for integrating learners' ideas with visualizations

The knowledge integration perspective provides further insight on the design of visualizations to help learners achieve coherent understanding. How can the design of visualization support the four knowledge integration processes? From research on visualizations designed for instruction of science content, I identify three design principles that can engage students in the knowledge integration processes: *present multiple representations of phenomena*, *promote learner initiated investigations*, and *provide data recording tools*.

Present multiple representations of phenomena. Multiple representations provide multiple views of phenomena, giving learners a range of entry points toward understanding scientific concepts. Designers must be careful when providing students with multiple representations, as more representations generally leads to greater complexity. Designed correctly, multiple representations can contribute to learning by performing important cognitive functions, such as serving complementary roles when learners have multiple tasks to perform, or by use one representation to constrain learners' interpretation of another (Ainsworth, 1999).

Visualizations can use real-life or familiar representations to elicit learners' ideas about a phenomenon. Linking new representations to these familiar representations can then add ideas about the new representations and build on learners' prior knowledge. For instance, 4M:Chem (Kozma, 2003) combines macroscopic, graphical, symbolic, and atomic representations of chemical reactions. The macroscopic view helps learners build representational understanding from their everyday understanding of observable phenomena. ThinkerTools uses a familiar two-dimensional motion animation linked to a data-cross that adds ideas about the independence of motion in the x and y directions.

Multiple representations also offer many opportunities for students to distinguish their ideas. When presented with more than one representation of a phenomenon, students must distinguish between the ways the same information presented differently in each representation, augmenting their understanding. The Graphs and Tracks dynamics environment (McDermott, 1990) allows students to examine position, velocity, and acceleration graphs that accompany motion animations. Students come to understand the graphs partly by distinguishing between the characteristics of each type of graph in representing the same motion profile. Multiple representations can also require learners to choose which of several representations is most suitable for a task. For instance, EChem (Wu, Krajcik, & Soloway, 2001) allows students to view space-filling, wireframe, and symbolic representations of organic compounds. Each representation serves a different purpose during the course of their inquiry investigation on toxins.

Promote learner-initiated exploration. Visualizations can allow learners to initiate investigations that connect to the learning goals of instructions, rather than be passive observers of animations. Learners can use visualizations to test their own ideas, perform controlled comparisons, and gradually refine their understanding of the topic.

Learner-initiated investigations elicit ideas by providing opportunities for learners to pose questions and articulate predictions. Students can then explore the visualization to answer their questions or test their predictions, and compare outcomes to expectations. Virtual experiments are particularly effective for this pattern of instruction with visualizations. Game-like tasks can present learners with a goal to achieve, eliciting ideas on how to accomplish the task. Games add ideas as learners interact with the visualizations to test these ideas. ThinkerTools takes advantage of a game-like environment to allow students to explore the nature of motion. Giving students the opportunity to test their own ideas helps ensure that the visualization adds ideas at the right level at a time when students are ready for new ideas.

Learner-initiated investigations require learners to distinguish between multiple situations in order to make insights. In a motion environment such as ThinkerTools, students must distinguish between the effects of impulses in different directions. Experimentation environments may require learners to distinguish between a pair of controlled trials to see the effect of a particular variable on the outcome (Varma, 2010).

Finally, unexpected results from investigations can raise new questions for learners, encouraging them to revisit their ideas and refine their understanding. Clark and Jorde (2004) examine the way students interact with a tactile model of heat flow, where they can feel different materials at different temperatures. Upon discovering whether these different objects feel hot or cold, students have the chance to revise their conceptions about the difference between heat and temperature.

Provide record keeping tools. Records of interactions help learners monitor their progress with visualizations. These tools can take the form of notebooks or journals, or they can automate the routine task of recording what learners have done. Records obviate the need for learners to keep previous interactions in memory, allowing them to focus on the learning goals. In addition to providing metacognitive support, record-keeping tools can help learners integrate their ideas. Visual records of interactions can make patterns in data more explicit, adding ideas about the effect variables have outcomes. Records place the outcomes of multiple inquiries in the same visual space, allowing learners to make distinctions more easily between multiple situations. Visual records can also facilitate the revisiting of previous work and to see the full range of activity with the visualization, promoting reflection. An example of record keeping tools is the World Watcher environment (Edelson, Gordin, & Pea, 1999), which makes use of a notebook where students store snapshot of data visualizations, add annotations, and include links to the original data. The notebook combines the results of students' investigations and their thoughts into one location students can use to identify patterns, determine relationships between temperature and geography, and distinguish multiple situations. Hyperlinks facilitate returning to the original data to revise ideas.

Using design patterns to support visualizations

Despite their promise, visualizations themselves are limited in their ability to support complex learning. Engaging students in complex science investigations requires learning environments that go beyond the capabilities of visualization tools. Though well-designed visualizations can help students integrate ideas, learning environments that incorporate visualizations can guide students more effectively through the knowledge integration processes. Inquiry environments can elicit and add ideas that help students interpret visualization tools, or provide opportunities for reflection after using visualizations to illustrate phenomena.

A recent meta-analysis on the effectiveness of dynamic visualizations (Chang, Chiu, McElhaney, & Linn, in preparation) shows that pretest-posttest studies measuring knowledge gains from instruction using dynamic visualizations had a large average effect size (1.32). All but one of these 68 effect sizes reflect student learning in classroom settings. This finding suggests benefits for the combination of visualizations and instructional support from other software tools, peers, and teachers.

Research on the design of instructions points to the need to incorporate visualizations within well-studied instructional patterns to help students integrate their understanding of visualizations with their everyday ideas, prior knowledge about the domain, and the investigation context. Many design patterns are well-suited to incorporating visualizations. The *predict, observe, and explain* pattern can elicit predictions about a phenomenon, then allow students to use a visualization to observe the phenomenon. The *experiment* pattern can take advantage of virtual experimentation environments. Visualizations can provide students with evidence on which to *construct an argument* about the investigation. Visualizations also provide material for *critique* of investigation methods or the design of the visualizations themselves.

Research must refine the use of design patterns in ways that best take advantage of the affordances of visualizations. Laboratory studies on visualizations can examine with great detail the ways learners use visualizations and their ability to enhance domain knowledge. However, these studies generally fail to capture how visualizations contribute to students' scientific understanding within the context of authentic investigations. Classroom studies that compare versions of instruction with and without visualizations may lack the detail necessary to identify what aspects of the instructional design features best support students' use visualizations and what types of interactions best promote coherent understanding. Research on instruction using visualizations must examine the role of specific types of support, how they influence the way learners interact with visualizations, and the resulting insights learners make.

In this dissertation I present the results of design experiments (Brown, 1992; Collins, Joseph, & Bielaczyc, 2004) that compare different types of support for virtual experimentation and examine how these supports affect learners' experimentation choices and insights. I take advantage of software that logs students' interactions with visualizations in the context of their own classrooms so that I can analyze in detail the way students conducted their investigations. I also analyze embedded assessments, which provide students with opportunities to connect the results of their experiments with the goals of the broader investigation. In this next section of this chapter, I review research on how learners conduct and learn from experiments and ways that instruction can guide learners toward informative experimentation.

Experimentation

Experimentation is a critical aspect of professional scientific inquiry (T. Kuhn, 1970; Latour & Woolgar, 1986; Thagard, 1992). Instruction that incorporates experimentation provides students with opportunities to engage in many other related inquiry activities such as posing questions, generating hypotheses, exploring relationships between system variables, analyzing data, critiquing methods, generating explanations for results, and constructing arguments using evidence.

Beginning with Inhelder and Piaget (1958), researchers have found that students struggle with various aspects of experimentation. For example, during hypothesis generation, learners frequently fail to specify variables of interest and the relationships among them (Njoo & Jong, 1993) and avoid precise hypotheses, which are less likely to be rejected (Klahr, Fay, & Dunbar, 1993; Klayman & Ha, 1987). While conducting experiments, learners fail to conduct controlled tests when appropriate (Schauble, 1996; Tschirgi, 1980), conduct experiments that aim only to confirm hypotheses (Klahr & Dunbar, 1988), use experiments to produce specific outcomes rather than test hypotheses (Schauble, 1990), ignore unfamiliar variables (Linn & Swiney, 1981), and overemphasize the effect of one causal variable while ignoring the others (D. Kuhn, Black, Keselman, & Kaplan, 2000). When interpreting experimental evidence, learners struggle to understand multivariable causality (Keselman, 2003), distort evidence to support a prior belief (Dunbar, 1993), and ignore, reject, or misinterpret data that do not fit into their existing theories (Chinn & Brewer, 1993; D. Kuhn et al., 1995). These studies point to the need for research on improving learners' ability to conduct and interpret valid experiments.

Experimentation in knowledge-lean and knowledge-rich contexts

Early research on scientific reasoning addressed children's ability to isolate variables in knowledge-lean experimentation contexts. For instance, Inhelder and Piaget (1958) designed a task [later adapted by Kuhn and Phelps (1982) and others] that asked subjects to determine what combination of colorless fluids would yield a specific reaction outcome. Siegler and Liebert (1975) examined the ways subjects determined how an electric train runs on the basis of four binary switches (though in actuality, a researcher operated the train using a secret switch to ensure that subjects would test all 16 combinations). These studies examined experimentation as domain-general logical inference, as subjects had no information on which to base testable hypotheses. In these situations, subjects could make valid inferences only by isolating variables to logically eliminate possibilities.

Over time, research has increasingly examined knowledge-rich contexts and revealed the important role of context-specific knowledge in conducting experiments. Experimentation studies in realistic contexts illustrate how designing and interpreting experiments involves a much more complex and nuanced set of factors than simply the ability to logically confirm or disconfirm hypotheses using controlled experiments. For example, studies show that children are more likely to test plausible rather than implausible hypotheses (Klahr, et al., 1993; Tschirgi, 1980), focus on variables they believe to be causal (Kanari & Millar, 2004), and use experiments to achieve specific outcomes rather than test hypotheses (Schauble, 1996). Though learners' ideas about the investigation context may lead them toward invalid experimental designs or inferences, students may also use ideas productively, such as by narrowing the range of testable values or eliminating implausible explanations. Tschirgi (1980) argued that children's tendency to use "invalid" strategies when determining the ingredients needed to bake a good cake is reasonable, given real-life goals of reproducing positive results (good cakes) and eliminating negative ones (bad cakes). Koslowski (1996) also argued that using prior knowledge to generate and interpret evidence is a good strategy, particularly when understanding mechanisms informs the interpretation of outcomes. These studies indicate that learners' alternative strategies sometimes stem from efforts to refine their understanding of the situation, such as by narrowing the set of investigation questions or exploring the nature of the variables.

Other research shows the extent to which learners' prior understanding of the domain may affect their learning outcomes. Linn, Clement, and Pulos (1983) compared the students' reasoning in laboratory tasks and naturalistic tasks involving the effects of system variables on an outcome. The study found that part of the variance in performance on these tasks was associated with task content knowledge. Schauble (1996) examined experimentation by children and adults in two science domains. The study revealed that subjects who conducted valid experiments often reached invalid conclusions informed by their prior knowledge of the system, and that subjects' knowledge sometimes informed their experimentation strategies. These findings suggest that the contribution of domain knowledge to scientific reasoning is more important than studies on experimentation strategies may acknowledge, and that knowledge of the domain and strategies may exhibit a "bootstrapping" effect on learning from experimentation. These studies thus go beyond examining learners' combinatorial reasoning and incorporate learners' rational decision making about the experimentation context.

These research studies point to experimentation as an important way to extend learners' understanding of a domain as well as appreciation for methods for advancing knowledge. Experimentation thus constitutes a key design pattern for knowledge integration. Incorporating experimentation activities within inquiry investigations provides learners with opportunities to test their own ideas about the domain and use the outcomes of experimentation to generalize knowledge to new contexts.

Designing guidance for experimentation

Experimenting in realistic contexts requires learners to consider a wide range of ideas to design informative experiments. Learners need to integrate everyday ideas they have about the topic, formal knowledge about the science domain, and knowledge about strategies for experimentation in order to investigate complex questions. This multitude of ideas suggests that procedural guidance alone (such as domain general instruction of controlling variables) may be insufficient to promote informative experimentation. How can guidance help learners focus on the ideas that will help them use experiments to extend their knowledge of inquiry investigations?

Here I discuss two important studies that shed light on how experimentation guidance can help learners integrate ideas about the domain with ideas about experimentation strategies. Klahr and Dunbar (1988) presented subjects with a task to identify the function of the repeat command in algorithms that controlled the motion of a robot called BigTrak. They compared the performance of subjects who were asked to determine how the repeat command worked without any guidance to subjects who were first asked to generate several hypotheses in advance. The study found that subjects who elicited multiple hypotheses needed fewer experimental trials to solve a problem and more frequently conducted experiments that allowed them to distinguish between hypotheses than subjects who did not elicit hypotheses.

Klahr and Dunbar interpret their results using their dual search model. They suggest that subjects who tested just one hypothesis at a time more frequently searched the experiment space in efforts to confirm their hypothesis, while subjects who generated multiple hypotheses searched the hypothesis space and thus aimed to conduct experiments that discriminated between the hypotheses. From the knowledge integration view, generating hypotheses served to bring multiple ideas about the repeat command to the forefront. Determining the function of the repeat command depended on subjects' ability to distinguish correct and incorrect explanations, which they could only do with a small subset of algorithms. Klahr and Dunbar's study illustrates the key role of distinguishing ideas in interpreting the outcomes of experimental trials.

In a second study, Vollmeyer, Burns, and Holyoak (1996) allowed subjects to interact with a multivariable system during a learning phase and a problem-solving phase. In the specific goal condition, subjects were informed of the system state they would be asked to achieve before the learning phase. In the non-specific goal condition, subjects were not informed of this system state until after the learning phase. The study found that subjects with non-specific goals outperformed those with specific goals on a transfer task and that strategies used by subjects with specific goals were effective for achieving the goal, but ineffective for discovering the underlying structure of the system. Though Vollmeyer et al. interpret the results using rule search and instance search, the knowledge integration perspective also highlights the importance of distinguishing ideas. During the learning phase, specific goals prompted subjects to distinguish between system states in order to revise their experimental approach, but not to distinguish carefully between the variables' effects on the outcomes. Subjects with non-specific goals, on the other hand, had to distinguish between the nature of the system variables and their roles in determining the outcomes.

Emphasis on the Control-of-Variables Strategy (CVS)

The majority of studies on experimentation examine learners' ability to design and interpret controlled experiments. This research tradition not only early research on children's logical thinking skills but addresses a crucial aspect of real-life scientific inquiry. School science standards also capture the equating of experimentation with controlled comparisons, such as one of the eighth grade scientific inquiry standards from the American Association for the Advancement of Science (1993): "If more than one variable changes at the same time in an experiment, the outcome of the experiment may not be clearly attributable to any one variable" (p. 12). Standards like this make the control-of-variables strategy (CVS) a prominent goal for classroom science instruction.

School science's strong emphasis on CVS has led to research efforts that aim specifically to promote the CVS in classroom settings. For instance, Klahr and colleagues (Chen & Klahr, 1999; Klahr & Nigam, 2004) have generally found direct instruction of the control of variables strategy (CVS) to be more effective than indirect or self-directed methods in helping students design controlled experiments and improve their domain knowledge. Klahr and Nigam also found that a substantial number of students were able to master CVS by using discovery methods and that these students performed just as well on a subsequent transfer task as students who mastered CVS by direct instruction. Dean and Kuhn (2007) found that direct instruction was neither necessary nor sufficient for students to retain mastery of CVS on a delayed transfer task, and that guided practice, rather than direct instruction, was the determining factor in long term mastery of CVS.

One limitation of these studies by Klahr, Kuhn, and their colleagues on promoting use of CVS is that they focus almost entirely on procedural aspects of experimentation and do not consider how learners make use of prior knowledge about the domain in designing their experiments. The implications of these studies for promoting valid experimentation in contexts where domain knowledge plays an important role is therefore unclear.

Furthermore, in certain situations, experimentation strategies other than CVS may also yield important insights. Learners may use diverse strategies to explore the nature of the variables and their connections to the investigation context. Unstructured or spontaneous exploration of the variables may help learners identify consequential experiments involving critical values. Characterizing experiments as either controlled or uncontrolled may not capture learners' tendencies to investigate new questions that arise during the course of the investigation. This dissertation research extends studies that focus on CVS by examining an experimentation context where the inferences students need to make require more than just controlling variables. In this way I make a distinction between *controlled* experiments and *informative* experiments.

Dissertation goals

This dissertation aims to combine three strands of science education research: the design of inquiry instruction, learning from dynamic visualizations, and how learners design and interpret experiments. *Airbags* unifies these three research strands by using design patterns to incorporate the visualizations and guide students' experimentation. This dissertation aims to extend previous research along these three strands in several ways. First, I extend research on the design of inquiry instruction to the inclusion of a virtual experimentation environment. The relevant everyday context of airbag safety provides an anchor for interpreting the visualization and gives meaning to the experimentation activity. *Airbags* makes the topic of motion graphs consequential by connecting it to the driving question and incorporating graphs within the experimentation activity. Students use experiments to extend their domain knowledge to the relevant context of airbags, rather than merely to demonstrate their capacity for logical reasoning.

Second, I extend research on how students learn from experimentation to a realistic investigation context and to classroom settings by examining the relationships among students' prior understanding, their experimentation strategies, and their insights. *Airbags* presents students with a complex investigation task requiring students to combine multiple sources of evidence and their prior knowledge about graphs and the investigation context. Students' everyday understanding of motion, their physics domain knowledge about motion graphs, and evidence from the World-Wide Web all have the potential to contribute to the way students design and interpret their experiments.

Third, I extend studies that focus on students' use of CVS by examining an experimentation context where the inferences students need to make require more than just isolating variables. In *Airbags*, strategies that isolate variables may not provide all the information required to understand the situation. I also explore how students come to understand *thresholds* (values of a particular variable above or below which the outcome is independent of the other variables). The presence of thresholds requires students to distinguish between covariation and thresholds to achieve a complete understanding of *Airbags*. Though students may readily observe the effects of thresholds by isolating variables, other strategies may inform students' understanding of thresholds as well.

This dissertation documents the impact of the *Airbags* curriculum as a whole and connects students' experimentation strategies (as logged by the learning environment) to the quality of their learning. I investigate whether imposing a constraint, designed to encourage planning, has an effect on students' experimentation strategies and the inferences they make from the outcomes of their experiments. I also investigate what students can learn by using their experiments to compare rather than isolate variables.

Chapter 3: Design of Curriculum Materials

Airbags is a one-week curriculum module for high school physics classes (typically eleventh and twelfth grade physics). I designed *Airbags* using the Web-based Inquiry Science Environment (WISE, Linn & Hsi, 2000). WISE allows designers of instruction to build inquiry modules using steps that promote scientific understanding. For instance, students can view evidence, compose reflection notes, engage in online discussions or debates, interact with visualizations, or illustrate their ideas with drawing tools.

Learning goals

Airbags has two primary learning goals. First, students examine the relationship between the nature of one-dimensional motion and the characteristics of position-time and velocity-time graphs. *Airbags* addresses students' difficulties with connecting graphs and physics (McDermott, Rosenquist, & Zee, 1987), differentiating between the height and slope of a graph (Leinhardt, Zaslavsky, & Stein, 1990), and distinguishing position, velocity, and acceleration (Trowbridge & McDermott, 1980, 1981). These topics constitute part of national science and mathematics standards (American Association for the Advancement of Science, 1993; National Council of Teachers of Mathematics, 2000). The design of *Airbags* assumes students' classroom curricula have already introduced the motion concepts of position, velocity and acceleration in one dimension. I designed *Airbags* to serve as a capstone unit that strengthens students' prior ideas about motion and graphing, and not to introduce these concepts to students for the first time.

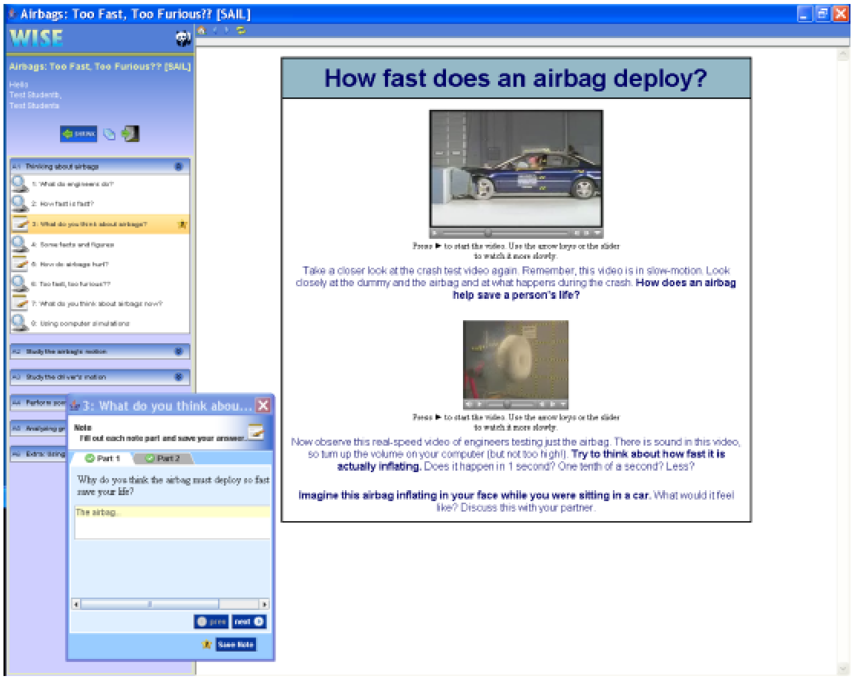
Second, *Airbags* aims to help students understand the dynamics of airbag deployment and the risks for injury from an airbag in a head-on collision. In *Airbags*, students investigate factors that lead to a high risk for injury to the driver from an airbag. I discuss the specific variables students investigate in more detail later in this chapter. The design of *Airbags* aims to integrate these two learning goals by prompting students to use graphs to further their understanding of the investigation context.

Curriculum sequence and design patterns

As described in the previous chapter, I designed *Airbags* using the knowledge integration perspective to promote coherent understanding. *Airbags* uses the knowledge integration design patterns that aim specifically to support experimentation. *Airbags* underwent frequent review by a wide range of experts including education researchers, technology developers, physics experts, and classroom teachers. The module underwent yearly design revisions in response to these reviews as well as student feedback and students' responses to embedded prompts. Revisions addressed the content of evidence pages, design of the visualizations, phrasing and conceptual focus of the prompts and embedded assessments, and the sequence of steps used to support inquiry activities. Here I describe the curriculum sequence of the final version of the module, used for Study 3. Previous versions of *Airbags* used for Study 1 and Study 2 used the same patterns, with slightly different steps, prompts, and embedded assessments. The *Airbags* module takes advantage of a series of dynamic visualizations, created using Dynamica software and developed by the Concord Consortium (www.concord.org).

Activity 1: Orient and elicit

The first activity uses the *Orient and elicit* pattern to introduce students to the investigation context and elicit their ideas about how airbags work and why they might present dangers in certain circumstances. A screenshot of the first activity of *Airbags* appears in Figure 1. The activity presents students with different types of evidence, such as a slow motion video of a head-on crash test, a full-speed video of an airbag deploying, and fatality statistics from accidents involving airbags. Students articulate their initial ideas in response to prompts concerning how airbags are designed to work, why they must deploy at such high speed, and the conditions in which they might be dangerous. The activity encourages students to view the crash test video multiple times to familiarize them with the sequence of events that occur during a head-on collision. The prompts guide students toward developing the main criterion for determining whether the driver was injured by the airbag—encountering an airbag that has not finished inflating. The subsequent activities build on these ideas by introducing the motion characteristics as variables and the safety of the driver as the outcome in experimental trials.



How fast does an airbag deploy?

Press ► to start the video. Use the arrow keys or the slider to watch it more slowly.

Take a closer look at the crash test video again. Remember, this video is in slow-motion. Look closely at the dummy and the airbag and at what happens during the crash. **How does an airbag help save a person's life?**

Press ► to start the video. Use the arrow keys or the slider to watch it more slowly.

Now observe this real-speed video of engineers testing just the airbag. There is sound in this video, so turn up the volume on your computer (but not too high!). **Try to think about how fast it is actually inflating.** Does it happen in 1 second? One tenth of a second? Less?

Imagine this airbag inflating in your face while you were sitting in a car. What would it feel like? Discuss this with your partner.

Note
Fill out each note part and save your answer...

Why do you think the airbag must deploy so fast save your life?

The airbag...

Figure 1. Screenshot of an evidence step and an embedded note in the first activity of *Airbags*.

This pattern takes advantage of the benefits of relevant problems (Linn & Hsi, 2000) and driving questions (Krajcik, Blumenfeld, Marx, & Soloway, 1999) for inquiry instruction. I designed *Airbags* for students at or near legal driving age, making automobile safety a particularly interesting topic for many students. Furthermore, the motion and forces that students experience as either drivers or passengers in cars from day to day provide students with a kinesthetic understanding that can be extended to this investigation.

Activities 2 and 3: Predict, observe, explain, and compare.

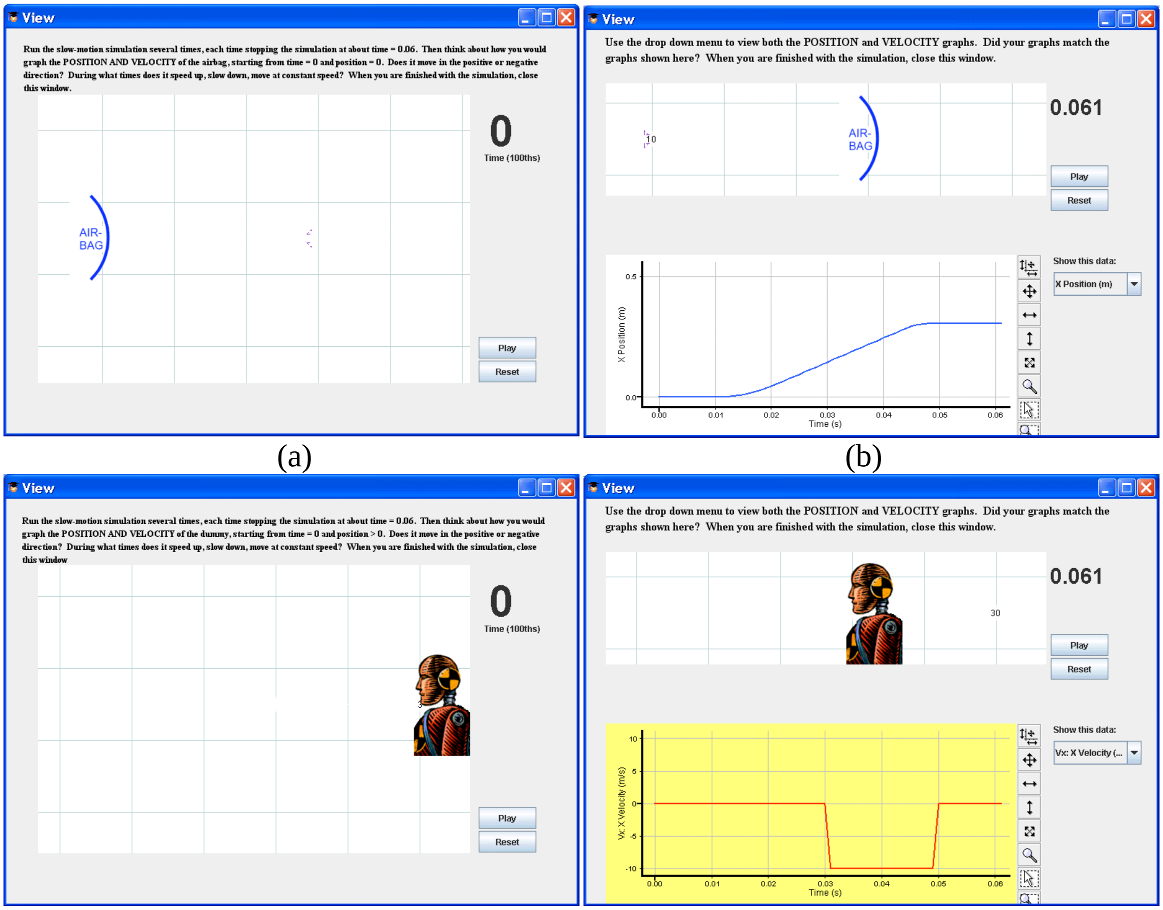
The second and third activities use a modified version of the common *predict-observe-explain* pattern to help students understand the nuances of airbag's and driver's motion during the collision and how graphs represent this motion. Activity 2 prompts students to watch the crash test video, focusing on the motion of the airbags, then presents students with a simple animation of this motion (Figure 2a). Students view this motion as many times as they need to, pausing the animation if necessary to characterize details about the motion profile. Activity 2 then prompts students identify the time intervals where different types of motion occur (such as speeding up, slowing down, traveling at constant speed, and not moving). Students then use a drawing tool to sketch position and velocity graphs of the motion they observe in the animation. Most students revisit the animation several times and engage in an extended discussion with their working partner in order to sketch a reasonably accurate graph. After receiving approval from me or their teacher, they observe the animation concurrently with dynamically generated position and velocity graphs (Figure 2b). Prompts ask students to compare the computer-generated graph to their own graph, compare the characteristics of different parts of the graph, and discuss and how different sections of the graph represent different types of motion. These prompts call students' attention to the difference between their initial ideas and new normative ideas about graphing, as well as highlighting distinctions between constant velocity, positive acceleration, and negative acceleration.

Activity 3 focuses on the motion profile of the driver (Figure 2c and d) and is nearly identical in structure to Activity 2. In addition to the step sequence of Activity 2, Activity 3 prompts students to compare the motion of the driver to the motion of the airbag. This comparison highlights the distinction between graphs that represent motion in opposite directions. In Activities 2 and 3, the predict-observe-explain pattern helps students distinguish between their prior conceptions of motion and graphs from new, normative ideas.

Activity 4: Conduct an experiment

In Activity 4 students conduct virtual experiments to investigating the effect of three motion variables on the driver's risk for injury from an airbag. The activity starts by reminding students about the criterion for incurring injury from an airbag, introducing the three collision factors, and prompting students to conjecture about the role of each of the factors on the driver's safety. Next, students plan experiments that answer the investigation questions using three motion variables. Students use a dynamic visualization (Figure 3) to conduct their experiments, and finally report their findings. I will describe in more detail the ways students interact with the visualization in the following section, as it comprises the core inquiry activity of *Airbags*.

The experimentation activity aims to promote a sophisticated understanding of the collision situation by allowing students to engage in key inquiry processes. The activity builds on students' prior ideas by eliciting conjectures and a plan for experimentation. The visualization adds normative ideas by showing graphical representations of unique collision outcomes in response to students' variable choices. Students are required to make distinctions by comparing the effects of the three variables on the outcome, and comparing the outcomes of controlled trials. The experiments set the stage for the following activity where students must use the evidence they generate to interpret new collision situations.



Run the slow-motion simulation several times, each time stopping the simulation at about time = 0.06. Then think about how you would graph the POSITION AND VELOCITY of the airbag, starting from time = 0 and position = 0. Does it move in the positive or negative direction? During what times does it speed up, slow down, move at constant speed? When you are finished with the simulation, close this window.

Use the drop down menu to view both the POSITION and VELOCITY graphs. Did your graphs match the graphs shown here? When you are finished with the simulation, close this window.

Run the slow-motion simulation several times, each time stopping the simulation at about time = 0.06. Then think about how you would graph the POSITION AND VELOCITY of the dummy, starting from time = 0 and position > 0. Does it move in the positive or negative direction? During what times does it speed up, slow down, move at constant speed? When you are finished with the simulation, close this window

Use the drop down menu to view both the POSITION and VELOCITY graphs. Did your graphs match the graphs shown here? When you are finished with the simulation, close this window.

(a) (b)

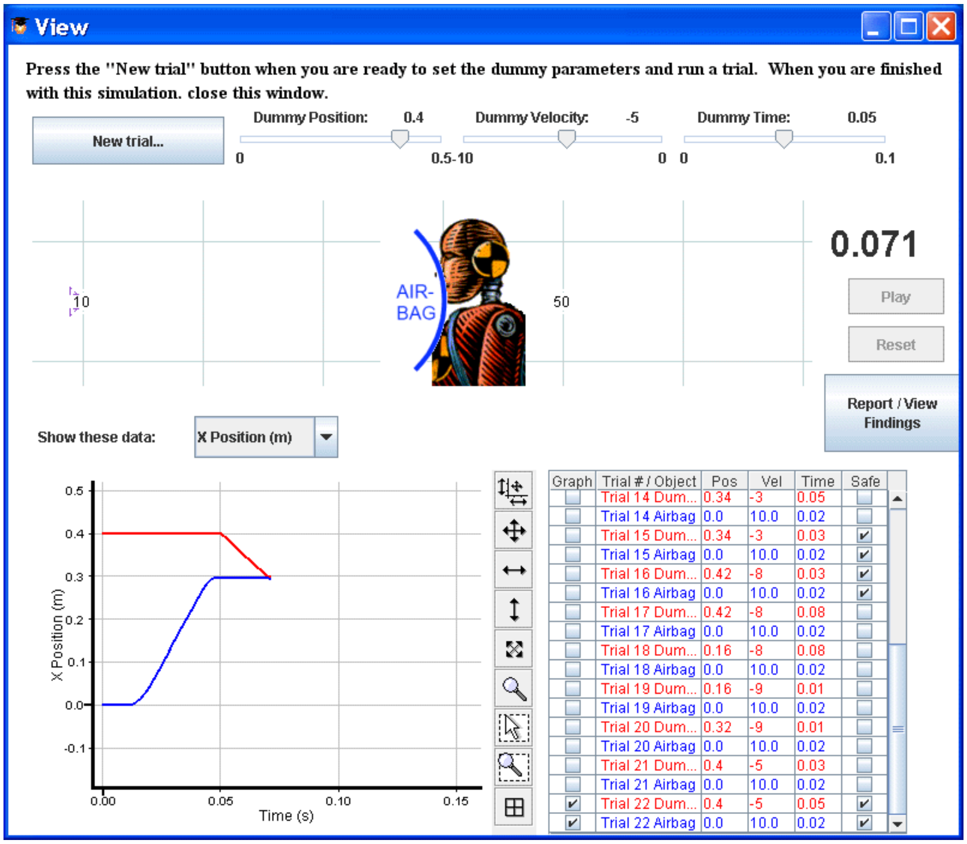
(c) (d)

Figure 2. Visualizations in the *Airbags* module. Students observe an animation of motion [(a) and (c)], predict the appearance of graphs, then observe computer-generated graphs simultaneously with the motion [(b) and (d)].

The experimentation activity aims to reinforce the learning goals about motion and graphs by promoting use of the graphical representations to interpret and compare the outcomes of their trials. Changes to each variable correspond to unique changes in the appearance of the graph of the driver (the red line in Figure 3). For instance, changes to the driver's initial position translate the driver's graph vertically on the axes, while changes to the driver's velocity change the slope of the line representing the driver's approach to the airbag. In this way, controlled comparisons highlight the relationship between the nature of motion and the characteristics of graphs.

Activities 5 & 6: Construct an argument

Activities 5 and 6 employ the *Construct an Argument* pattern to encourage students to assess the quality of their own understanding and bring together multiple sources of evidence from the module. The first part of Activity 5 is framed as an opportunity to examine hypothetical data from a "black box" device that captures the motion of the driver and airbag during a collision. The activity provides students with examples of collision graphs from the visualization, and students must explain whether it represents a safe or unsafe outcome, and which variable was most responsible for this outcome. The second part of Activity 5 requires students to construct graphs to support their views. Students must use the graphs to distinguish between two collision scenarios, such as those involving a tall or a short driver.



Press the "New trial" button when you are ready to set the dummy parameters and run a trial. When you are finished with this simulation. close this window.

Dummy Position: 0.4 (0 – 0.5)
Dummy Velocity: -5 (-10 – 0)
Dummy Time: 0.05 (0 – 0.1)

0.071

Show these data: X Position (m)

Graph	Trial # / Object	Pos	Vel	Time	Safe
☐	Trial 14 Dum...	0.34	-3	0.05	☐
☐	Trial 14 Airbag	0.0	10.0	0.02	☐
☐	Trial 15 Dum...	0.34	-3	0.03	☑
☐	Trial 15 Airbag	0.0	10.0	0.02	☑
☐	Trial 16 Dum...	0.42	-8	0.03	☑
☐	Trial 16 Airbag	0.0	10.0	0.02	☑
☐	Trial 17 Dum...	0.42	-8	0.08	☐
☐	Trial 17 Airbag	0.0	10.0	0.02	☐
☐	Trial 18 Dum...	0.16	-8	0.08	☐
☐	Trial 18 Airbag	0.0	10.0	0.02	☐
☐	Trial 19 Dum...	0.16	-9	0.01	☐
☐	Trial 19 Airbag	0.0	10.0	0.02	☐
☐	Trial 20 Dum...	0.32	-9	0.01	☐
☐	Trial 20 Airbag	0.0	10.0	0.02	☐
☐	Trial 21 Dum...	0.4	-5	0.03	☐
☐	Trial 21 Airbag	0.0	10.0	0.02	☐
☑	Trial 22 Dum...	0.4	-5	0.05	☑
☑	Trial 22 Airbag	0.0	10.0	0.02	☑

Figure 3. Visualization used for the experimentation activity. At the top, students select and investigation question, specify variable values, and observe the animation for each trial. In the lower left, students can view a position or velocity graph of the airbag's and driver's motion. In the lower right, students can see their trial history, which they can use to sort trial outcomes and compare the graphs of multiple trials.

In constructing arguments in support of the view they choose, students must determine whether their current state of understanding is sufficient to distinguish among multiple scenarios. If they are unable to make the distinctions, students are encouraged to revisit previous evidence to refine their understanding. In this way, Activity 5 encourages self-monitoring and serves as a checkpoint for their understanding of the graphs and investigation variables.

The goal of Activity 6 is to help students bring together evidence from the module, experiments, and the World Wide Web and apply it to the first step of a design task: to recommend design improvements to cars and airbags to make them safer. Because some classes ran out of time to finish Activity 6, different teachers implemented it in different ways. Some teachers did it class as designed, others assigned it as a homework assignment, and a few omitted the activity altogether. One teacher who had professional experience with patents asked his students to generate a patent application for a new airbag feature. The activity aims to strengthen students' understanding of the collision dynamics by applying their understanding of factors that lead to injuries from airbags.

Students' interaction with the experimentation visualization

My dissertations studies focus on experimentation choices and learning outcomes using the virtual experimentation activity. The visualization (Figure 3) presents a two-dimensional representation of the motion of the airbag and driver, starting from the moment the car collides with a barrier and using the steering wheel as a point of reference. Students use the visualization to investigate three questions about the role of the driver's height, the collision speed, and the amount that the car can crumple in determining the driver's risk for injury. Each of these questions maps onto one of three motion variables students can manipulate in the visualization (the initial position of the driver, the velocity of the driver toward the airbag after impact, and the time between impact and driver's initial motion relative to the steering wheel).

Students interact with the visualization by conducting a series of trials to test their conjectures. To conduct a trial, students first click the New Trial button (Figure 3, upper left) and select an investigation question from a drop down menu (or indicate that they are "just exploring"). Next, students use the sliders at the top to specify values of the driver's position, velocity, and time variables. Students then play the crash simulation, at which time they can judge whether the trial was "safe" and record this outcome in the experimentation history using checkboxes. While students conduct their experiments, the software logs the investigation question and variable values students select for each trial, providing me with information on students' experimentation choices and their intentions.

In the visualization, two types of relationships govern the risk for injury to the driver from an inflating airbag. First, over a particular range of values, each of the three variables covaries with the time that elapses before the driver and airbag collide. Tall drivers, low speed collisions, and a large crumple zone therefore make a driver more likely to encounter a fully inflated airbag than short drivers, high speed collisions, and a small crumple zone. Second, two threshold values (for position and time) determine situations where the likelihood of injury is invariant: (1) short drivers who sit within an airbag's zone of deployment will *never* encounter a fully inflated airbag, and (2) for sufficiently tall drivers, if the duration of the crumple zone exceeds the deployment time for the airbag, drivers will *always* encounter a fully inflated airbag.

The combination of covariation-based and threshold-based relationships between variables and outcomes produces piecewise, rather than simple linear, functions that describe conditions that produce safe outcomes. These complex relationships force students to combine knowledge of the collision events, motion parameters, and graph interpretation with their knowledge of experimentation strategies in order to achieve a sophisticated understanding of the situation. Students' ability to design informative experiments therefore depends on more than their propensity to control variables according to predetermined patterns.

Visualization design considerations

In the previous chapter I described three design principles for visualizations to promote integrated understanding of science: *present multiple representations of phenomena*, *promote learner initiated investigations*, and *provide data recording tools*. Here I discuss how I implemented each of these design principles in the *Airbags* visualizations to help students achieve the learning goals.

Present multiple, simultaneous representations of motion

Several instructional interventions use multiple, simultaneous representations to illustrate motion concepts (McDermott, 1990; Parnafes & diSessa, 2004; Tao, 1997; White & Frederiksen, 1998). Graphs are effective for representing a motion profile in part because they leave a time-persistent record of motion that learners can view at a single glance (Ainsworth, 1999; Parnafes & diSessa, 2004). This characteristic facilitates comparing the events of multiple experimental trials at once. Animations and graphs can be mutually informative, as certain aspects of motion may be easier for students to observe using one representation than the other. For instance, the difference between zero and non-zero acceleration may be clearer in a graph than an animation, but distinguishing the direction of motion may be easier for some students using an animation.

Airbags includes both position and velocity graphs in the visualizations. I based my decision to allow students to toggle between the two graphs (rather than presenting them both simultaneously) on several factors. First, presenting one graph at a time keeps the appearance of the visualization from becoming too busy and overwhelming for students. Second, limited screen space would have required two graphs to be much smaller and possibly more difficult to use. Third, during the experimentation activity, toggling between the two graphs allows students to focus their attention on the graph that they find most helpful for interpreting the situation, without the distraction of the more difficult one. To prevent students from confusing the two graphs, I added a yellow background to the velocity graphs (Figure 2d).

Promote and facilitate experimentation

I discussed in the previous chapter the benefits of experimentation as an inquiry activity and curriculum design pattern for promoting coherent understanding. Several design features of the visualization aim to facilitate experimentation for students. Requiring students to choose an investigation question focuses students on the inquiry goals with every trial. (This feature also provides me as the researcher with more information about students' intentions, helping me interpret their experimentation choices.) Using sliders to input variable values allowed me to specify what values for each variable would be testable. Limiting the number of testable values for each variable made repeating tests using the same values easier for students, facilitating comparisons. Sliders also made interaction with the visualization less cumbersome than typing numerical values.

Experimentation history

The experimentation history includes several features to help students monitor their experimentation and facilitate analysis. First, it automates the recording of trials, leaving students with more cognitive resources to focus on analysis and inferences (Quintana, et al., 2004). Second, student can use checkboxes in the right hand column to categorize outcomes as either safe or unsafe. This feature encourages students to make a judgment on the outcome of each trial and can help students see patterns in the data. Third, students can use checkboxes in the left hand column to view the graphs of previous trials. This feature allows students to review their previous findings and conduct comparison between multiple trials. These comparisons are particularly useful for examining the effect that changes to a single variable have on the graph or on the outcome. In order to maintain simplicity of interaction, I did not allow students to sort the data in the history table. I discuss the implications of this particular design decision in the discussion chapter.

Chapter 4: General Methodological Issues

This dissertation research consists of three classroom studies conducted during three successive school years. Each of the next three chapters presents one of these studies, including methodological details specific to that study. This chapter presents general methodological issues that that pertain to all three studies. These issues include details about school settings, professional development, classroom enactment, the design of assessments, and development of rubrics for scoring the assessments.

Schools

The three studies in this dissertation occurred across seven public high schools in three states of the USA. The schools were part of school districts that were partners in the Technology-Enhanced Learning in Science (TELS) Center, a National Science Foundation Center for Learning and Teaching that funded this research. One of the primary aims of TELS was to examine the effectiveness of science inquiry learning in culturally diverse settings. The schools that took part in my studies on *Airbags* therefore represent a wide range of ethnic and socioeconomic student populations. Table 1 describes these populations. School 6, the school for which student population data are not reported, is a regional magnet program for gifted science and mathematics students drawing from multiple high schools across a large metropolitan area.

I used *Airbags* in junior and senior level physics classes at these seven schools. At many of these schools, enrollment in physics was very low (sometimes fewer than 10 students). At each school, the students who studied *Airbags* constituted all the students at that school who were enrolled in physics at that time.

Teachers and Professional Development

Teachers in these studies ranged in science teaching experience from zero to more than 30 years. Teachers who used *Airbags* in their classrooms did so by their own choice and were paid a modest stipend for their participation if they also took part in TELS large-scale yearly benchmark assessments. Teachers were also encouraged to attend TELS professional development workshops held during the school year (Varma, Husic, & Linn, 2008). Topics for professional development included fostering teaching strategies to facilitate learning from inquiry projects and dynamic visualizations, learning how to use teacher tools for assessing student work, using examples of student work to better understand students' scientific ideas, and module customization (teachers did not customize *Airbags* for the studies in this dissertation). Some teachers attended TELS summer research retreats, where they had the opportunity to learn about new technology developments in WISE, and provide feedback on WISE teacher tools and the content of individual modules.

Table 1. Schools participating in studies with *Airbags*

School	Ethnic diversity	% eligible for reduced price lunch
1	57% White 26% Hispanic/Latino 11% Asian 4% African-American	20%
2	53% Hispanic/Latino 18% White 13% African-American 12% Asian	59%
3	54% Hispanic/Latino 26% White 11% Asian 6% African-American	48%
4	65% African-American 31% White	27%
5	96% African American 4% White	52%
6	NR	NR
7	41% African-American 23% Hispanic/Latino 22% Asian 12% White	63%

Note: NR = not reported

Classroom enactment

Pre-enactment

A few weeks prior to a module run, I communicated with technology staff at the school to ensure that the school's network and computers were compatible with WISE. I tested Airbags on the computers that would be used for enactment, which were either the school's computers or laptop computers belonging to TELS.

During this time I also communicated with the teacher to discuss the details of enactment, usually in a face to face meeting. For teachers using *Airbags* for the first time, the teacher and I would go through each step of Airbags so that the teacher was familiar with the module's structure, pace, technology tools, and learning goals. I discussed the expectations for student responses and interactions with the visualizations, as well as the relationship between *Airbags'* design and my research questions. I also discussed the ideal role of the teacher as a facilitator for discussions among students. For teachers who had used *Airbags* in previous years, I discussed changes to the module since the previous enactment and new research questions.

The teachers and I worked together to divide students into working dyads that were homogeneous with respect to science and mathematics ability. My pilot studies on Airbags indicated students were most successful when paired with someone of similar ability level. When students had widely diverging ability, the more capable student often usurped responsibility for responding to prompts and making insights. Students of similar ability tended to work more collaboratively and engage in more productive discussions about the content of *Airbags*.

The school day before students began working on *Airbags*, the teacher administered the pretests. Pretests usually required between 20 and 30 minutes of class time. In some cases, students created accounts with WISE immediately before starting Airbags. In other cases, particularly when class time for *Airbags* was limited, I created accounts for students in advance and gave them login information at the beginning of the module.

Classroom enactment

Airbags usually required about four to five hours of class time, depending on students' ability level, the amount of time the teacher allotted for the module, and occasional technology troubleshooting. Students who were absent during the module were encouraged to spend time at the beginning of the following class to review the progress made by their partner while they were absent.

Students worked at their own pace on *Airbags* during the course of the week. The teacher and I would circulate throughout the class, inspecting students' work and responding to students' questions. Where we noticed students struggling to reach insights, we would initiate guidance to elicit or add appropriate ideas. At key points during the module I would facilitate brief (five minute) whole class discussions to facilitate the exchange of ideas across student groups. For instance, during Activity 1 I usually prompted students to share their interpretation of the fatality statistics, so that students could consider a wide range of alternatives to their initial interpretations. At the end of Activity 2 I engaged the class in a discussion about the relationship between the airbag's deployment and the graph of its motion. At times the teacher would prompt students to connect the *Airbags* activity to aspects of their previous instruction or clarify expectations for responses to embedded prompts. I also used a projector to demonstrate the basic features of the module's dynamic visualizations.

Conversations and discussions aimed to support the design patterns by promoting the four knowledge integration processes. Class discussions would elicit the range of students' ideas and help students add new ideas to each others' repertoires. After bringing alternative interpretations of visualizations or data out into the open, pointed questions could help students distinguish between their alternatives or recognize where their understanding was incomplete. The teachers and I could accomplish similar goals during conversations with students by asking questions, recommending students to revisit previous evidence, or prompting students to argue their viewpoint.

Post-enactment

On the last day of each module run, I held a brief discussion with students asking for their opinions of the module. Specifically I would ask students what they liked and did not like about

Airbags, as well as what feedback they had on the module's content and the WISE software. Within one or two days of all students finishing *Airbags*, the teacher administered the posttest.

Assessment design

Design of pretest, posttest, and embedded assessments followed the knowledge integration assessment framework (Linn, et al., 2006), which focuses on how well students articulate valid scientific connections between ideas. Assessments for *Airbags* assess conceptual links between the nature of motion and characteristics of graphs.

Pretests and posttests

Most teachers were willing to devote no more than five to seven school days to using Airbags, including time for assessments. I therefore designed pretests and posttests so that students could complete them in 20 to 30 minutes. Posttests addressed the same concepts as pretests but differed slightly to reduce gains due to retesting. Pretests and posttests were administered to students individually.

To measure how well students could generalize their knowledge about motion to new contexts, graphing items assessed students' ability to interpret and construct motion graphs in real life motion contexts other than the *Airbags* context. Items assessed students' ability to interpret and construct position versus time and velocity versus time graphs for motion in one dimension. Items required students to distinguish between the stationary state, motion with constant velocity, and motion with constant acceleration. Graph interpretation items presented students with a motion scenario, using a graph to represent the motion. These items used a multiple choice plus constructed explanation format. Graph construction items presented students with motion scenario describing the motion in a few sentences, asking students to construct a graph that represents the described motion.

I used two other types of pretest/posttest items. Study 1 used one item to capture students' understanding of the role of the driver's height in determining their risk for injury from an airbag. Students demonstrated virtually no knowledge on this item on the pretest, so it was later incorporated into the embedded assessments. Study 2 used three items to capture students' ability to design and interpret controlled experiments. These items were removed for Study 3 because they no longer informed the study aims.

I refined pretest and posttest assessments during two years of pilot testing to improve clarity and their ability to elicit relevant scientific ideas. New and revised items were reviewed by research colleagues and experts in science assessment. Pretests and posttests used for the three studies appear in the appendices. I discuss examples of specific pretest and posttest assessment items and how they address with the research goals of the individual studies in the next three chapters.

Embedded assessments

I designed embedded assessments to capture students' understanding of graphs in the context of the *Airbags* investigation and their understanding of the relationship between the investigation variables and collision outcomes. Students responded to embedded prompts in their working dyads. Each of the three studies used different sets of items to capture knowledge relevant to the goals of that study. These items are described in detail in the following three chapters. Embedded assessments, like pretests and posttests, were reviewed by research colleagues and science assessment experts. As with the pretest and posttests, I discuss examples of specific embedded prompts items and how they address the research goals of the individual studies in the next three chapters.

Scoring rubrics

I designed scoring rubrics for pretests, posttests, and embedded assessments using the knowledge integration assessment framework (Linn, et al., 2006). Knowledge integration rubrics capture the number and quality of normative conceptual links between scientific ideas. Depending on the complexity of the item, I coded students' responses on a scale from zero to four (for simpler items) or zero to five (for more complex items). Table 2 illustrates the coding levels used throughout the three studies. Rubrics were reviewed along with their corresponding assessment items by research colleagues and science assessment experts. Rubrics for pretest/posttest and embedded assessments appear in the appendices. I discuss the specific features of rubrics for individual items in the next three chapters.

Table 2. General form of the knowledge integration (KI) scoring rubric for pretests, posttests, and embedded assessments

KI score	KI level	Description
0	None	No response
1	Irrelevant/ Off task	No relevant scientific ideas
2	Invalid	Relevant but invalid scientific ideas without conceptual links
3	Partial	Relevant conceptual link between two ideas that is not fully elaborated
4	Full	One relevant, fully elaborated conceptual link between two ideas
5	Complex	At least two relevant, fully elaborated conceptual links among three or more ideas

Overview of the three studies

This dissertation consists of a sequence of three studies where the designs of the last two build on the findings of the study that precedes it. Study 1 does not use a comparison condition—it could be considered a pilot study. Its main goal is to characterize the relationship between students' experimentation choices and their learning outcomes. Study 2 and Study 3 are design experiments (Brown, 1992; Collins, et al., 2004) that aim to examine the impact of subtle forms of guidance on students' experimentation choices and insights. These studies compare alternative designs of the instruction.

I chose to conduct design experiments for two main reasons. The first reason concerns the ethics of classroom research. Within the context of students' regular classroom curriculum, I believe researchers have an obligation to provide students and teachers with the best possible curriculum materials. Asking a subset of students to use curriculum materials believed by the designers to be inferior to other versions takes advantage of the trust that students place in their teachers and that teachers place in researchers. Design experiments allow researchers to compare versions of curriculum materials where neither version provides a clear advantage to the learner, or whose design aims to benefit students in different ways.

The second reason for design experiments concerns the nature of my research goals. This research aims to extend the current body of research on the design of inquiry instruction. Studies on benefits from inquiry investigations using pretest-posttest designs on a single version of the curriculum are limited in their ability to point to the benefits of specific aspects of the design. Instructional benefits in these kinds of studies can be attributed to the overall design of the intervention but are difficult to attribute to individual design features or guidance methods. Design experiments permit researchers to compare versions of instruction and examine the effects of subtle changes to the instructional design, potentially leading to more detailed design knowledge that can be more readily utilized by the education research community.

Design experiments also offer advantages over many experimental designs involving control conditions (such as traditional instruction). First, the instructional quality of the control condition can be called into question, while design experiments effectively serve as their own control. Second, research designs that compare conditions that are too different (such as animated and static graphics, or instruction with and without dynamic visualizations) may not adequately examine the benefits of specific design features of instruction, and may not account for the effects of poor design. My studies with *Airbags* aim to change the instructional conditions just enough to generate differences in learning outcomes that are observable only by using carefully designed assessment instruments or by examining specific subsets of the student population. These results point more clearly to the ways instructional design decisions affect students learning and for which students the effects occur.

The three studies form a coherent sequence by examining different ways that instructional design and students' prior knowledge, experimentation choices, and learning outcomes intersect. Study 1 related students' experimentation choices to their learning about motion and graphs. This study revealed that students who use a moderate number of trials to conduct controlled comparisons had the best learning outcomes. This finding did not however indicate whether the relationship was causal.

Based on the outcome of Study 1, I designed two conditions for Study 2 by imposing a constraint that aimed to promote experimentation conditions that were successful for students in Study 1. The study found that though the constraint promoted controlled trials, students' focus on the logistics of experimentation appeared to distract them from the learning goals. The findings suggest that students benefited from strategies other than isolating variables.

Study 3 built on Study 2 by eliminating the constraint and comparing the initial set of experimentation goals with a new set of goals aimed to promote alternative strategies. Study 3 also includes analysis of three case studies that illustrate how students made their experimentation choices and used the results of their experiments to reach insights about the *Airbags* situation.

Chapter 5: Study 1 - Students' Experimentation Strategies and Learning

I designed Study 1 primarily to address my first two research questions. Specifically, I sought to examine (1) the overall impact of *Airbags* on students' understanding of motion and graphs both within and outside the context of airbag safety, (2) what experimentation strategies students used without specific guidance on experimentation, and (3) the relationship between students' experimentation strategies and their learning. I used the results from this study to refine the instruction for Study 2, which isolates the effects of specific guidance on what students learned from their experiments.

Methods

To investigate the impact of *Airbags* I used a pretest-posttest design combined with embedded assessments (explanation prompts and logs of students' experimentation strategies). The assessments measured progress in developing coherent understanding of motion, graphs, and airbag safety.

Participants and Implementation

Six high school physics teachers used *Airbags* in the classrooms, encompassing 148 students. These schools included all the schools listed in Table 1 except for School 3. Three of the teachers were experienced and had taught previous versions of *Airbags*. All teachers participated in targeted professional development (Varma, et al., 2008). Most students worked in dyads on the activities. Unpaired students worked on their own computers while engaging in discussions with another group. At all six schools, every student taking physics at the school participated in this study (some schools had low enrollment in physics).

At all schools except for School 7, I was present in the classroom as a co-teacher alongside the students' regular classroom teacher. I played an active role in the implementation of *Airbags*, engaging students in individual and whole class discussion, responding to students' questions about the curriculum and content, and asking students for verbal explanations for their inquiry choices, and supporting the teacher in using the technology. I recorded observations of students' activities throughout the week-long implementation of *Airbags*.

Assessments and Scoring

Study 1 used three types of assessments of student understanding: pretest/posttest assessments, embedded assessments, and logs of students' experimentation sequences with the visualization.

Pretest/posttest assessments. I used pretests and posttests to measure students' broad understanding of the *Airbags* investigation and how well they could generalize their understanding of motion and graphs from *Airbags* to new motion contexts. Pretests and posttests were administered to individual students the day before the start of implementation and the day after completion. Posttests covered the same issues as the pretests but were changed slightly to reduce possible gains due to retesting, except in School 7 where the pretest was mistakenly administered as the posttest. Due to absences, some students did not take either the pretest or the posttest. Pretests and posttests consisted of 11 items on motion and graphs and were scored from either zero to four or zero to five. Students in School 7 received just 10 of the 11 items because of classroom time constraints. I scored pretest and posttest items using knowledge integration rubrics (Linn, et al., 2006) that reward valid scientific connections between concepts. The total pretest and posttest scores were the sum of the scores of the individual items. Examples of pretest and posttest assessments appear in Table 3.

Embedded assessments. I used embedded assessments to measure students' understanding of the experimentation visualization. Six embedded prompts (INTERPRETATIONS) asked students to describe motion that could produce graphs like the ones in the visualization. These items occurred just after the experimentation activity and measured students' understanding of graphs that illustrate the *Airbags* situation. Students responded to these prompts in their working groups. I scored these items on a scale from zero to four using knowledge integration rubrics.

- *Experimentation data logs.* Pedagogica software (Buckley, Gobert, & Horwitz, 2006) logged the investigation question and variable values students chose for each trial. I used the reports of students' trials to score each student groups' experimentation strategy in three ways:
    - *Total trials*. I computed the total number of trials each group conducted. Because some students occasionally conducted identical trials multiple times, I also computed the number of unique trials each group conducted. Unique trials correlated highly (r = .95) with total trials, so I used total trials in the analysis.
    - *Trial variability*. To measure how widely students changed the variable values, I computed a variability score. I computed for each of the three investigation variables (1) the number of unique values, (2) the range of values tested, and (3) the number of boundary values tested by each student group. I expressed these values as a fraction of the maximum possible number, computed the mean of these three 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.
    - *Use of the Control of Variables Strategy (CVS)*. I measured the degree to which students isolated variables in a way that was consistent with the investigation question they chose. Only trials where students selected one of the three investigation questions were used for this score. I scored the experimentation sequences once for each investigation question on a scale of zero to five using a knowledge integration rubric. The overall CVS score was the mean of the subscores for the three investigation questions. The three subscores exhibited an α correlation of .71, demonstrating that the degree to which students conducted controlled trials was fairly uniform across the three investigation questions.

Table 3. Examples of pretest/posttest and embedded assessment items

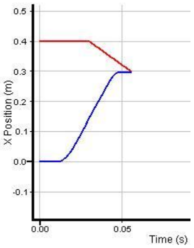
Name (type)	Example
MOTION (pretest/posttest)	A car starts at point A and speeds up at a constant rate until it reaches point B in 4 seconds, where it suddenly stops. The car waits at point B for 2 seconds. It then travels at constant speed in the opposite direction, reaching point A again in another 3 seconds. Sketch a POSITION-TIME graph and a VELOCITY-TIME graph of the motion during these 9 seconds.
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. 

Table 4. Knowledge integration rubric for the experimentation CVS score

KI level	Score	Description
blank	0	Conduct zero trials
none	1	Conduct exactly one trial
invalid/ isolated	2	Change all three variables between trials or hold the investigation variable constant
partial	3	Change exactly two variables between trials, including the investigation variable
basic	4	Change only the investigation variable between trials that produce the same outcome
complex	5	Change only the investigation variable between trials that produce opposite outcomes or conduct two separate sets of controlled trials

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

Analysis

In School 6, a subset of students' experimentation records failed to upload to the WISE servers. In order to confirm the completeness of the logs, students at this school reported how many experimentation trials they conducted on 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. I used two-tailed, paired t-tests to measure learning gains from pretest to posttest.

To examine school effects on student learning from *Airbags*, I conducted a one-way analysis of variance (ANOVA) on the 10 pretest and posttest items common to all students, using school as the between-groups factor. Though these analyses showed significant school effects on both pretest [F(5, 108) = 17.35, $p < .001$] and posttest [F(5, 107) = 6.38, $p < .001$], when School 6 (the only school with selective admission criteria) was removed from the analysis the effect for school on the posttest was not significant [F(4,86) = 1.39, p = .25]. This result suggests that the students' experience with *Airbags* and what they learned was similar across most school settings. Because of the similarity in students' achievement across school settings and the very small number of participating students at some schools, I pooled the students from all schools and used multiple linear regression models to relate learning outcomes to experimentation measures controlling for prior knowledge using scores on the common pretest items. Taken together, the students from these six schools provide a fairly representative sample of physics students from across the United States, diverse in their levels of prior knowledge and classroom context.

Results and Discussion

My classroom observations show that teachers implemented *Airbags* as intended and as documented in classroom observations. Students found the module engaging and interacted with the visualizations and their peers as intended.

Overall Impact of the Module

Students made large, significant pretest to posttest gains on the ten items administered to all students [M = 22.61, SD = 10.57 (pre); M = 28.08, SD = 7.39 (post), $t(108) = 6.40$, $p < .001$, $d = .60$]. 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 positive at all schools and significant ($p < .05$) for every school except for School 6 (which had very high pretest scores) and School 5 (which had just nine students). These gains illustrate the success of *Airbags* in promoting understanding across diverse settings. Positive gains for School 6, whose students were concurrently enrolled in calculus, demonstrate that even students with very high levels of prior content knowledge gained insights about the applications of physics to airbag safety. Table 5 shows the pretest scores, posttest scores, and effect sizes for students at each school.

Because of the pretest-posttest design, the pretest might have contributed to student improvement. However, I believe this is not a significant factor for several reasons. First, the items were altered from pretest to posttest, making them at least superficially distinct. Second, the test relies on generation items, and the more sophisticated responses on the posttest draw on experience with the unit. Third, previous studies show no significant gains on generation items scored using knowledge integration rubrics for students who did not receive instruction between pretests and posttests (Linn, et al., 2006). These three factors suggest most of the learning gains are attributable to students' experience with the *Airbags* module.

Table 5. Pretest and posttest means, standard deviations, and effect sizes for Study 1 schools

	Pretest		Posttest		
School	M	SD	M	SD	Effect Size
1	27.67	8.17	33.75	9.40	.69*
2	17.95	8.79	27.97	6.77	1.28***
4	25.78	8.32	30.11	7.32	.55*
5	25.44	11.26	29.89	6.56	.48
6	37.73	3.01	39.23	4.85	.37
7†	18.16	9.55	27.78	7.27	1.13***

†The scores of School 3 include just 10 of the 11 total items administered to the rest of the schools.
*$p < .05$ ***$p < .001$

Students' Experimentation Strategies

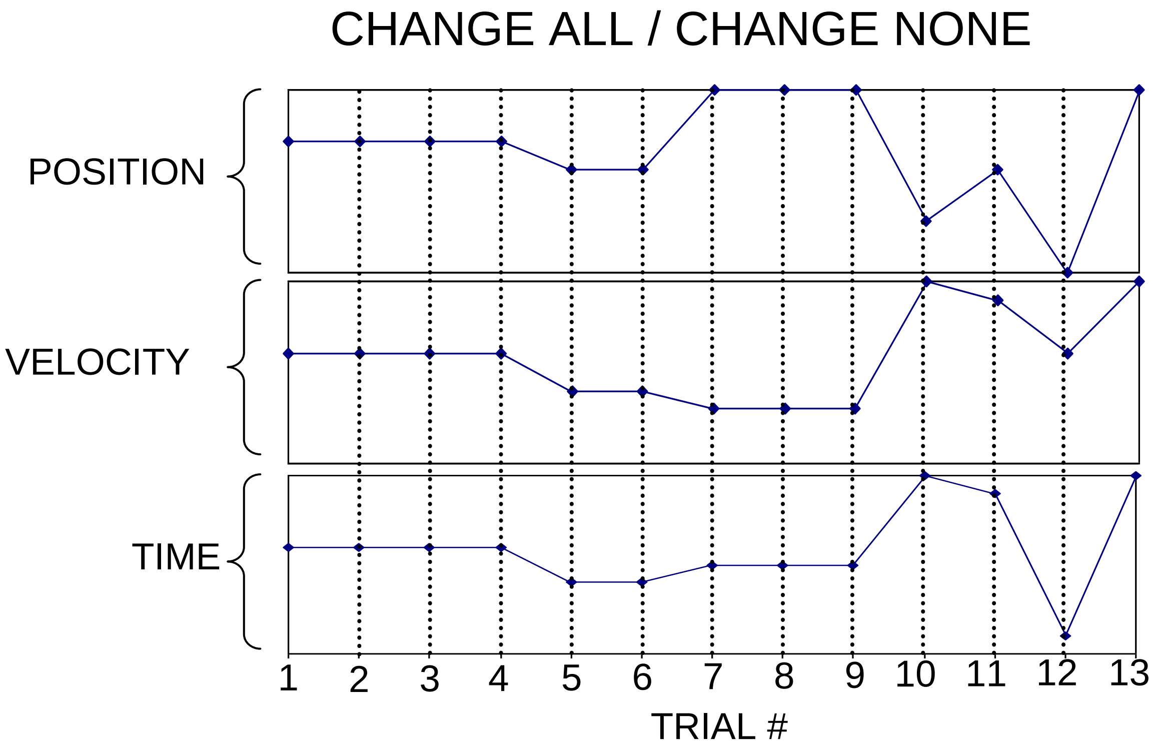
Students exhibited wide variation in their experimentation strategies. To illustrate this variation, I present the experimentation sequences of three dyads whose strategies represent the wide range of variability and CVS scores students achieved. Figure 4 shows the variable values each of these dyads chose for each trial. 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. I illustrate students' use of CVS when they have held two variables constant while varying the appropriate investigation variable between two consecutive trials (shaded in Figure 4).

Figure 4a shows a dyad that employed a "change-all/change-none" strategy, without isolating any variables. Their low CVS score (1.66) reflects their failure to investigate the individual variables in their trials. These students' above-average variability score (67.8) reflects the wide range of values (including four boundary values) they tested. Their strategy illustrates how students could vary the variables widely without isolating variables and suggests an experimentation approach that is *uninformed* by the investigation questions. Students who are uninformed may also employ *haphazard* approaches.

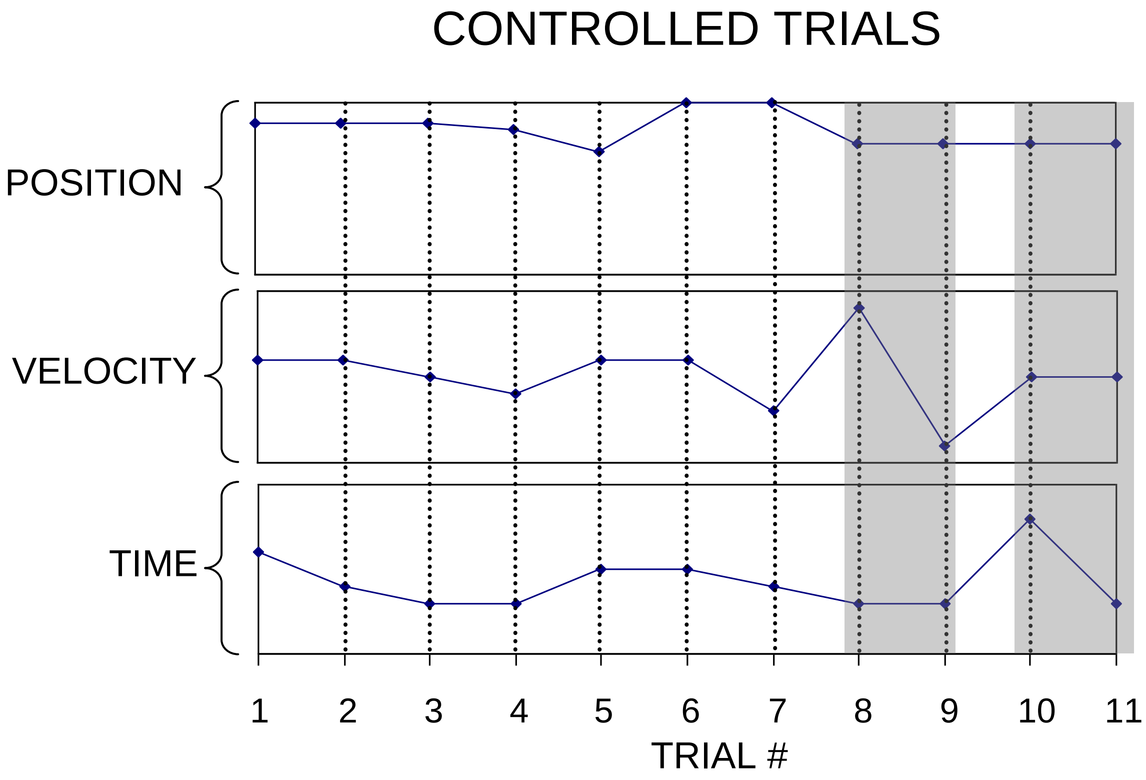
Figure 4b shows a dyad that successfully isolated two of the three variables, in Trials 8 through 11. Though these students might have intended to isolate the time variable between Trials 1 and 2 or the position variable between Trials 5 and 6, they did not communicate their intention to do so by selecting the appropriate investigation question for each trial. Nevertheless, their high CVS score (4.00) suggests they were aware that isolating appropriate variables would inform their understanding of the situation. Their strategy also exhibited a below-average variability score (39.6) that reflects the narrow range of values they tested for the position variable, and their reluctance to test boundary values. Their strategy illustrates how students can isolate variables without varying them widely and suggests an *efficient* experimentation approach, where the students purposefully tested each investigation question using a pair of controlled trials.

Figure 4c shows a dyad that conducted an *exhaustive* sampling of the system variables. These students isolated each variable appropriately at some point during the process, but they also conducted many tests that did not isolate a variable. Further illustrating the haphazard process were that the students changed the investigation question 18 times during their 47 trials, and that 13 of these trials duplicated a previous trial. Their high variability (86.1) and CVS (5.00) scores likely reflect the large number of trials they conducted, rather than purposeful investigation of the questions.

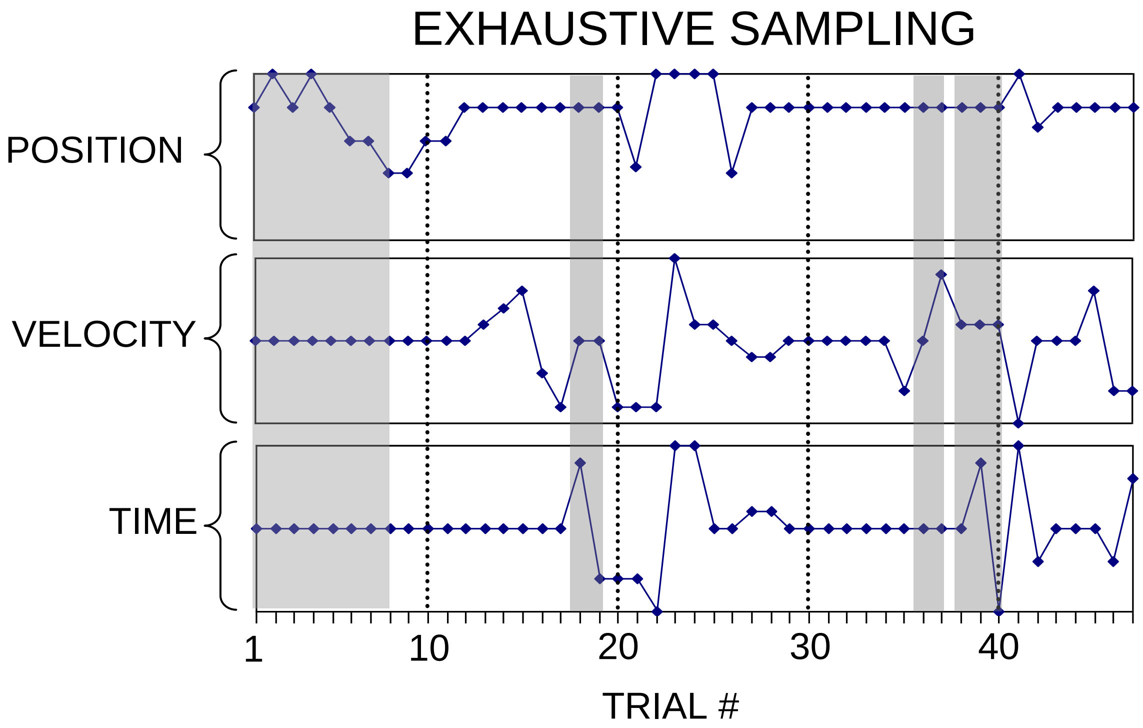
The students whose *uninformed*, *exhaustive*, and *efficient* strategies appear in Figure 4, in addition to those who were too *unengaged* or *confused* to conduct more than a few trials, represent four broad categories of experimenters I observed in *Airbags*. Though these categories are not completely distinct, they help to characterize how students approached the experimentation task. I suspect that these categories extend to other experimentation learning contexts as well. The subsequent analysis reveals how learning outcomes for these four types of experimenters compare with one another.



(a)


CONTROLLED TRIALS
POSITION
VELOCITY
TIME
1
2
3
4
5
6
7
8
9
10
11
TRIAL #

(b)



(c)

Figure 4. Experimentation sequences for three workgroups illustrating wide ranging strategies: (a) change-all/change-none (b) controlled trials (c) exhaustive sampling. Consecutive trials where students isolated the variable corresponding to the investigation question they chose are shaded.

Relationship Between Experimentation and Learning Outcomes

I investigated the relationships between the three experimentation scores and students' posttest scores, controlling for pretest scores. A multiple linear regression model revealed that none of the relationships were significant and that pretest scores were a much stronger predictor of posttest scores than any of the experimentation scores. The weak relationships between experimentation and posttest scores were expected, 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 numerous graph interpretation and construction activities and reflection prompts.

I next investigated the relationships between the three experimentation scores and the students' responses to the embedded INTERPRETATIONS prompts, which measure students' understanding of the content of the experimentation activity. I generated a regression model for each learning outcome using pretest scores, total trials, trial variability, and CVS score as predictors. Table 6 lists the linear regression coefficients. The regression models show that CVS score was a significant positive predictor controlling for the other experimentation scores and pretest scores. The standardized coefficients (β) indicate that CVS score was an even stronger predictor of learning than students' prior content knowledge.

Interestingly, total trials was a marginally significant negative predictor of learning. I conjectured that in addition to the unengaged or confused students who conducted very few trials, students who used exhaustive strategies might also have demonstrated poor learning outcomes. To test this conjecture, I divided students into two groups according to the number of trials they conducted using the median split of 12.5 trials. Separate linear regression analyses on each group showed a strong and significant *positive* relationship between total trials and embedded assessment performance for the low-trials group, but a strong and significant *negative* relationship for the high-trials group. This result supports our conjecture and reaffirms the idea that exhaustive strategies, though they may isolate variables, may also reflect students' failure to link their strategies to the inquiry investigation. Table 6 presents the results of all three regression models relating experimentation and learning.

Table 6. Summary of regression analysis for predicting outcomes on the embedded INTERPRETATIONS assessment

Predictor	*B*	*SE B*	*β*
All Students (N = 114)			
Pretest	0.02	0.01	0.17*
total trials	-0.01	0.01	-0.20
trial variability	0.14	0.29	0.05
CVS score	0.30	0.05	0.63***
Trials < 12.5 (N = 58)			
Pretest	0.01	0.01	0.06
total trials	0.14	0.04	0.51***
trial variability	-1.56	0.47	-0.45***
CVS score	0.26	0.07	0.46***
Trials > 12.5 (N = 56)			
Pretest	0.03	0.01	0.31*
total trials	-0.02	0.01	-0.56***
trial variability	0.43	0.36	0.18
CVS score	0.15	0.06	0.32*

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

The analysis also indicates that students whose strategies exhibited high CVS scores while using a moderate number of trials were the most successful on the embedded assessments. A high CVS score reflects several dimensions of students' knowledge other than just being able to isolate variables. First, students must map the investigation questions onto the appropriate variables. Second, students must correctly interpret the outcomes of their trials (as safe or unsafe). Third, because students must articulate their investigation goal before conducting each trial, the CVS score measures students' ability to plan investigations in advance. The findings thus highlight the several aspects of informative experimentation. In addition to isolating variables, connecting experimentation strategies to investigation goals, building on one's everyday knowledge of the situation, and planning in advance can all contribute to valid inferences.

Though efficient experimenters generally demonstrated the best understanding of the airbags situation, it is uncertain whether the relationship between implementing specific strategies and student learning is causal in nature. Did students have better learning outcomes because they used appropriate strategies, or does students' ability to experiment efficiently reflect their understanding of the airbags situation? Both explanations seem plausible. On one hand, conducting controlled tests could help students gain insights about how individual variables affect the experimental outcomes and the properties of the graphs. On the other hand, sophisticated strategies also reflect students' understanding of the variables and their ability to interpret outcomes. Students who have a better understanding of the situation to begin with may also be better able to design informative experiments. Study 2 aims to address these questions by imposing a constraint on students' experimentation to promote efficient strategies and examining whether efficient experimenters achieve a better understanding of the *Airbags* situation.

Chapter 6: Study 2 - Effect of constraining students' experimentation

To investigate in more detail the connection between experimentation and learning, I designed a Study 2, a comparison study. This study builds on the findings of Study 1 by imposing a constraint on students' experimentation. The constraint aims to guide students toward using efficient strategies that investigate the nature of the individual variables. Constraining students to a specified number of experimental trials could improve learning outcomes for many experimenters. A constraint could encourage students who would spontaneously use uninformed or exhaustive strategies to plan their trials more carefully and intentionally investigate the roles of individual variables using controlled comparisons. The constraint could also improve learning outcomes of unengaged or confused students (who would otherwise conduct very few trials) by increasing the number and diversity of trials they conduct.

Alternatively, students in Study 1 who used efficient strategies were successful largely because those strategies were student-initiated, and using a constraint to guide students toward these strategies could be unhelpful to many students in making relevant insights about *Airbags*. The constraint could also limit students' ability to explore the experimentation variables in a way that best builds from their prior ideas and interfere with learning. This study examines how a constraint on experimentation affects students' experimentation planning, experimentation strategies, and ability to use graphs to understand the airbags situation.

Methods

I used a pretest/posttest quasi-experimental design with embedded assessments and two comparison conditions. In the *unconstrained* condition, students were instructed to conduct as many experimental trials with the visualization as they needed to answer the investigation questions. In the *constrained* condition, students were told they had enough material resources to conduct exactly twelve trials (the approximate number required for most students to investigate all three questions by isolating appropriate variables, based on Study 1). The modules used by students in each condition were otherwise identical.

Participants and Implementation

Fifty-eight students in three high school physics classes studied *Airbags*. Two teachers at different schools (Schools 2 and 3, Table 1) taught the three classes. Students worked in dyads on *Airbags*. Each of the two classes at one school was randomly assigned to one of instructional conditions. Students in the third class were assigned by their teacher using a stratified random approach to distribute their ability equally across the two conditions. Both the constrained and unconstrained condition contained 29 students. One teacher had three years of science teaching experience and had used previous versions of *Airbags*. The other teacher had more than 20 years of science teaching experience and was using *Airbags* for the first time. My role in the classroom was the same as in Study 1.

Assessments and Scoring

Study 2 used three types of assessments of student understanding: pretest/posttest assessments, embedded assessments, and logs of students' experimentation sequences with the visualization.

Pretest and posttest assessments. Pretests and posttests were divided into two subtests. The MOTION subtest consisted of ten items, most of which were used in Study 1 and which measured how well students could generalize their understanding of motion and graphs from *Airbags* to new motion contexts. The EXPERIMENTATION subtest consisted of three constructed response items that measured how well students could interpret and design controlled experiments. Figure 5 shows a sample item from the EXPERIMENTATION subtest. Administration and scoring of pretest and posttest assessments were the same as in Study 1. Figure 5 shows an example item from the EXPERIMENTATION subtest.

Embedded assessments. I used two types of embedded assessments:

- Before the experimentation step, *Airbags* prompted students to articulate a plan for using their experimental trials. I gave each dyad's experimentation plan a LOGISTICAL score and a CONCEPTUAL score. The LOGISTICAL score measured students' intent to investigate all the questions, vary the variable values, and conduct controlled experiments. The CONCEPTUAL score measured how well students connected investigation questions, variables, and anticipated outcomes. Table 7 shows the knowledge integration rubrics used to score the experimentation plans.
- Students responded to nine INTERPRETATIONS prompts, which included the six online embedded assessments from Study 1 and three additional paper-and-pencil assessments that measured students' ability to construct graphs that supported explanations of the airbags situation (Figure 6). Administration and scoring of the INTERPRETATIONS was the same as for Study 1.

Experimentation data logs. I used the reports of students' trials to score each student groups' experimentation strategy in the same three ways as in Study 1.

Overall assessment scores were computed as the sum of the individual items. Due to absences, a few students did not complete all the assessments or participate in the experimentation activity.

Analysis

The difference in pretest scores between students in the constrained and unconstrained conditions was marginally significant for both the MOTION [(M = 22.78, SD = 8.55 constrained; M = 26.34, SD = 5.46 unconstrained), $t(55) = 1.88$, $p < .1$] and EXPERIMENTATION [(M = 6.68, SD = 3.22 constrained; M = 8.10, SD = 2.27 unconstrained), $t(55) = 1.94$, $p < .1$] subtests. For this reason I used analyses of covariance (ANCOVA) to compare outcome measures from the constrained and unconstrained conditions. I used the MOTION pretest scores as a covariate when comparing students' responses to the MOTION posttest items and INTERPRETATIONS items (which focused on students' understanding of physics). I used EXPERIMENTATION pretest scores as a covariate when comparing students' experimentation strategies and their responses to the EXPERIMENTATION posttest items and the planning prompt (which focused on students' experimentation knowledge). One student who did not take the pretest and two students who missed a large portion of the module due to class absences were omitted from these comparisons. Four students who did not take the posttest were also omitted from pretest-posttest comparisons.

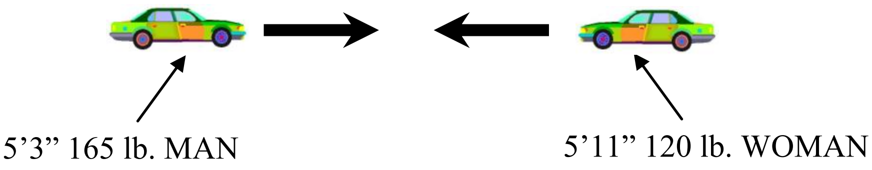
Amanda and Rosa want to know how to get the best gas mileage from their cars. They each conduct two trials and measure the gas mileage for each one. The tables show their results.

	Amanda’s test		**Rosa’s test**	
	Trial 1	**Trial 2**	**Trial 1**	**Trial 2**
Speed	45 mph	65 mph	50 mph	60 mph
Windows	UP	UP	UP	DOWN
Gas mileage	25 mpg	21 mpg	23 mpg	19 mpg

Whose test BEST shows what happens to the car’s gas mileage when the SPEED changes? Explain your answer.

Figure 5. Item from the EXPERIMENTATION subtest used for Study 2.

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 person do you think is **more likely to be injured** by an **airbag deploying**? Draw the position graphs of the **driver** of each car to illustrate your choice.

Figure 6. Graph construction INTERPRETATION paper and pencil assessment item used for Study 2.

Table 7. Knowledge integration rubric for scoring experimentation plans according to logistical and conceptual content

KI level	Score	Description (LOGISTICAL)	Description (CONCEPTUAL)
Blank	0	Nothing written	Nothing written
None	1	No mention of any planned activity	No mention of individual variables (position, velocity, or time) or investigation questions (height, speed, car crumpling)
Isolated/ Invalid/ Irrelevant	2	Plan without mention of specific strategy (individual trials, variation of variables, addressing investigation questions)	Reference to specific variables without reference to investigation questions OR reference to specific questions without reference to variables
Partial	3	At least one of the following partial connections between variables/questions and strategy: - Using trials to investigate without specifically referring to specific questions/ variables - Intention to vary variables without using CVS -	Reference to both variables and questions without clear connections between them
Basic	4	Exactly one of the following connections between variables/questions and strategy: - Plan to divide trials among individual variables or questions - Intention to employ CVS	Exactly one specific connection among variables, questions, and anticipated outcomes
Complex	5	Both of the following connections between variables/question and strategy: - Plan to divide trials among individual variables or questions - Intention to employ CVS	At least two specific connections among variables, questions, and anticipated outcomes

Results

Overall Impact of the Module

Students as a whole made significant pretest to posttest gains on both the MOTION [M = 24.62, SD = 6.45 (pre); M = 28.70, SD = 4.71 (post), $t(52) = 5.24$, $p < .001$, $d = 0.72$] and EXPERIMENTATION [M = 7.30, SD = 2.80 (pre); M = 8.58, SD = 2.26 (post); $t = 4.08$, $p < .001$, $d = 0.50$)] subtests. Students in both the constrained and unconstrained conditions made significant gains on both subtests. Learning gains in Study 2 are not easily comparable to gains in Study 1 because some of the assessments were revised and the school settings were different. Nevertheless, pretest scores, posttest scores, and effect sizes were similar across the two studies. These gains confirm the findings from Study 1 and suggest that that *Airbags* was successful in improving students' understanding of motion graphs and ability to design and interpret valid experiments.

Effect of Constraint on Experimentation Planning

Experimentation plans for the two conditions differed in both LOGISTICAL and CONCEPTUAL scores. ANCOVA using EXPERIMENTATION pretest scores as a covariate showed that plans for constrained students had significantly higher LOGISTICAL scores [M = 3.11, SD = 1.13 (constrained); M = 2.28, SD = 0.90 (unconstrained), F(1, 54) = 9.99, $p < .01$]. This difference reflects the constrained students' tendency to focus on how they planned to use precisely twelve trials. Plans for unconstrained students had marginally higher CONCEPTUAL scores [M = 2.39, SD = 1.47 (constrained); M = 3.36, SD = 1.62 (unconstrained), F(1, 54) = 3.74, $p < .1$]. This difference reflects the unconstrained students' tendency to focus on connecting the variables to the investigation questions or anticipated outcomes.

Effect of Constraint on Experimentation Strategies

Experimentation strategies for the two conditions also differed in both variability and CVS scores. ANCOVA using EXPERIMENTATION pretest scores as a covariate showed that the constrained students conducted tests with significantly higher variability [M = 65.1, SD = 22.7 (constrained); M = 48.4, SD = 23.2 (unconstrained), F(1, 51) = 8.79, $p < .01$] and CVS scores [M = 3.19, SD = 1.13 (constrained); M = 2.64, SD = 1.27 (unconstrained), F(1, 51) = 4.73, $p < .05$] than the unconstrained students. A t-test showed the mean number of trials conducted by the two groups to be nearly equal ($p = 0.3$), so the differences in variability and CVS scores are not attributable to an unequal number of trials conducted by students in the two conditions. These differences suggest that the 12-trial constraint was successful in guiding students toward isolating appropriate variables and exploring the variable values more widely.

Effect of Constraint on Learning Outcomes

Though I did not expect differences between conditions on either the MOTION or EXERIMENTATION posttest assessments, I conducted a confirmatory analysis to examine the possibility. As I expected, ANCOVA using corresponding pretest scores as a covariate revealed there were no significant differences in posttest scores between the two conditions ($p > 0.7$). Differences between the conditions on the posttest items were small most likely because the experimentation activity comprises a small part of the five total days of instruction, so both groups of students had many opportunities other than the experimentation activity to improve their understanding of motion graphs. In addition, both groups of students had opportunities to plan, design, and interpret a set of experimentation trials, all of which likely contributed to their understanding of experimentation as reflected on the posttest.

An ANCOVA using MOTION pretest scores as a covariate showed that unconstrained students significantly outperformed constrained students on the INTERPRETATIONS, [M = 27.16, SD = 4.44 (constrained); M = 29.52, SD = 4.27 (unconstrained), F(1, 54) = 4.07, $p < 0.05$]. This difference occurred even though the unconstrained students did not isolate variables as well or vary the variables as widely as the constrained students. This pattern was observed for both high prior knowledge and low prior knowledge learners (as measured by pretest scores). This finding disconfirms the hypothesis that isolating appropriate variables during experimentation led to greater learning in *Airbags*.

Discussion

In Study 2, students who were constrained to 12 trials articulated experimentation plans that focused on the logistics of experimentation, while unconstrained students focused their plans on conceptual connections between the system variables, investigation questions, and trial outcomes. Compared to unconstrained students, the constrained students experimented more widely with the variables and used CVS more extensively. However, unconstrained students demonstrated better understanding of the relationship between the *Airbags* graphs and collision events as they were depicted in the animation and crash test videos.

The most likely explanation for the results of Study 2 is that the constraint shifted students' attention during the experimentation activity away from understanding the situation and toward satisfying the constraint. The plans suggest that constrained students focused less on relationships between variables and investigation questions, instead detailing logistical considerations such as how to divide 12 trials among the three variables. Prompting students to attend to the logistical aspects of experimentation also caused students to focus on conducting "correct" or "valid" experiments. For some students, correct experiments sample a wide range of values for each variable (as illustrated by the variability scores). For other students, correct experiments involve controlled tests (as illustrated by the CVS scores). Though constrained students were more successful at isolating variables, they focused less on making sense of how the variables related to the outcomes and on the appearances of the graphs than the unconstrained group.

Unconstrained students, who in their plans focused on variable relationships and expected outcomes, may have been less likely to attend to "valid" experimentation and more likely to use diverse strategies to examine the nature of the variables. The performance of the unconstrained students on the embedded assessments suggests they had a stronger understanding of the *Airbags* situation than the constrained students. Though these students were less successful at isolating variables, their experiments were more *informative* about the *Airbags* situation than those of the constrained students.

Reid, Zhang, and Chen (2003) achieved a similar result in a study that compared *interpretive* and *experimental* support for computer-based experimentation activities about buoyancy. Interpretive support (IS) was designed to activate scientific concepts (such as balanced forces) students needed to make valid inferences about the system. Experimental support (ES) was designed to help students conduct valid experiments systematically. Reid et al.

found a significant effect for IS on students' ability to generalize their knowledge to new situations and associations between new and prior knowledge about weight, balanced forces, and buoyancy. There was no main effect for ES on students' understanding. Though their findings suggest the importance of conceptual support, their results could alternatively be attributed to the additional instruction that IS students received (in the form of conceptual questions and access to a knowledge base), rather than the specific design of the support. The Study 2 findings extend the findings of Reid et al., as only the design of guidance (and not the duration of instruction or access to information) differed between the instructional conditions.

The findings shed light on how the experimentation activity helped students augment their understanding of the airbags situation. The ability to isolate variables is important but not sufficient to ensure understanding of the situation in *Airbags*. This study demonstrates a case where students' ability to isolate appropriate variables did not lead to improved learning outcomes. Instead, the instruction that guided them toward isolating variables might have interfered with interpreting the overall meaning of their investigations. The experimentation plans that students generated in *Airbags* suggest that students who focused on conceptual, rather than logistical, details while planning gained a better understanding of the visualization. These students attended more to the processes that were responsible for the outcomes of their trials rather than the design of their experiments.

The findings highlight the need to study the full range of insights students gain during experimentation, not just whether students can isolate variables and make logical inferences. Study 2 extends prior research on how to guide students toward using CVS (e.g. Dean & Kuhn, 2007; Klahr & Nigam, 2004). In these studies, the quality of students' understanding is characterized only by inferences that require CVS. *Airbags* differs from these experimentation contexts in that experimenters can gain meaningful insights without fully isolating variables and miss important insights even when they do isolate variables. Thus, the experimentation activity can help students understand graphical representations of motion and the dynamics of airbag deployment even when they do not isolate all variables. Furthermore, students can isolate variables in ways that do not reveal all aspects of *Airbags*. The findings reinforce the idea that there is more to experimentation than isolating variables.

These results point to the need for measures of student understanding that go beyond assessing whether students can infer logical relationships between variables and outcomes. Studies should also assess students' understanding of the scientific concepts and mechanisms that govern these relationships. In Study 2, students who emphasized procedural rather than conceptual aspects of experimentation demonstrated a poorer understanding of the airbags situation, despite having isolated the appropriate variables more effectively. Measuring only whether students could control variables would not have captured this effect.

Student outcomes with the constrained and unconstrained versions of *Airbags* suggest ways for instruction to promote informative experimentation. Study 2 shows that in some situations the best guidance may focus more on relevant scientific concepts than on specific strategies, as strategies are ultimately a means to scientific understanding rather than ends in themselves. In addition, guidance that allows students' strategies to be governed by their conceptual understanding rather than procedural constraints may better help students leverage their prior knowledge, especially when experimentation is part of a larger inquiry investigation. Study 3, discussed in the next chapter, examines the impact of a form of conceptual, rather than procedural, guidance on students' experimentation strategies and inferences.

An alternative explanation for the results of Study 2 merits discussion. It is possible that the constraint increased students' cognitive load (Chandler & Sweller, 1991). Cognitive load theory suggests reasons why learners may struggle to learn using visualization tools (Chandler, 2004). Visualizations that increase extraneous cognitive load (which stems from activities not directly related to the learning goals) my interfere with learning. In Study 2, the requirement of limiting (or extending) students' spontaneous experimentation strategies to exactly 12 trials might have interfered with students' ability to make conceptual connections among the variables, outcomes, and graphs.

Cognitive load theory would suggest that guidance for experimentation should minimize extraneous cognitive load, even if it promotes "better" strategies. Instruction that asks students to use cognitive resources to implement specific strategies may divert attention away from other learning goals. By the same token, tools that reduce extraneous cognitive load may allow students to focus on target concepts. For instance, features of computer-based experimentation environments may reduce the extraneous cognitive load associated with data management by facilitating the organization of experimental results, helping students attend instead to the phenomena that produce the data and patterns in the experimental outcomes.

The results of Study 2 raise new questions. What strategies did students who did not focus on CVS use, and why were these strategies more informative? How can guidance promote informative experimentation by emphasizing conceptual rather than procedural aspects of experimentation? How do students reach insights when they do not control variables? Study 3 aims to address these questions by comparing two types of guidance for the experimentation activity in *Airbags.*

Chapter 7: Study 3 - Isolating vs. comparing variables

Study 3 examines how two different types of experimentation goals led students toward different investigation strategies and insights during *Airbags*. Study 3 examines how prompting students to *compare*, rather than *isolate*, variables in their experiments guides students toward different experimentation strategies and insights. Study 3 extends Study 2 by examining how experimentation goals can highlight the nature of the variables and make students' experimentation more informative. This study also explores how students come to understand the role of threshold values (values of a particular variable above or below which the outcome is independent of the other variables) in *Airbags*. The presence of thresholds requires students to consider factors other than covariation to achieve a complete understanding of *Airbags*. Though students may readily observe the effects of thresholds by isolating variables, I hypothesized that comparing variables would highlight the distinct characteristics of each variable, helping sophisticated learners achieve a more nuanced understanding of *Airbags* based on thresholds as well as covariation. I also expected that the more straightforward task of isolating variables would be more tractable (and thus beneficial) for less sophisticated learners.

Students may use two types of inferences to explain their findings in *Airbags*. First, over a particular range of values, each of the three variables (roughly) covaries with the time that elapses before the driver and airbag collide. Tall drivers, low speed collisions, and a large crumple zone therefore make a driver more likely to encounter a fully inflated airbag than short drivers, high speed collisions, and a small crumple zone. Second, threshold values for position and time determine situations where the likelihood of injury is invariant. For instance, short drivers who sit within an airbag's zone of deployment will *never* encounter a fully inflated airbag. Similarly, for sufficiently tall drivers, if the duration of the crumple zone exceeds the deployment time for the airbag, drivers will *always* encounter a fully inflated airbag. Responses to embedded assessments indicate whether students attribute their findings to covariation and/or thresholds.

Methods

Study design

Groups in the *isolate* and *compare* conditions received investigation questions (Table 8) that encouraged them to isolate or compare variables, respectively. Except for the investigation questions, the modules used for each condition were identical. I used a pretest/posttest experimental design with embedded assessments and two comparison conditions. I randomly assigned student groups to one of the conditions using a stratified approach to distribute their ability equally across the two conditions.

Participants

Physics students (N=166) at five high schools (Schools 2, 3, 4, 5, and 6, Table 1) studied *Airbags*. Most students worked in dyads on the module (unpaired students worked alone). Students were usually grouped with other students with nearly equal ability. Most teachers had used previous versions of *Airbags*.

Table 8. Collision factors, variables, and investigation questions for the *isolate* and *compare* conditions

Collision factor	**Variable**	**Investigation question**	
		***Isolate* condition**	***Compare* condition**
Driver height	Position	Are TALL or SHORT drivers more likely to be injured by a deploying airbag?	Does the DRIVER'S HEIGHT make the biggest difference in whether the driver is injured?
Collision speed	Velocity	Do HIGH or LOW SPEED collisions make drivers more likely to be injured by a deploying airbag?	Does the COLLISION SPEED make the biggest difference in whether the driver is injured?
Car crumpling	Time	Does MORE CRUMPLING or LESS CRUMPLING make drivers more likely to be injured by a deploying airbag?	Does HOW MUCH THE CAR CRUMPLES make the biggest difference in whether the driver is injured?

Data sources and scoring

Pretests and posttests

Ten constructed response pretest and posttest items (MOTION items used for Study 2) assessed how well students could interpret and construct graphs of one-dimensional motion. I administered pretests individually to students the day before the beginning of *Airbags* and posttests within a few days of completion. I scored the items using the same method as for Study 2. Students who did not take either the pretest or the posttest were excluded from analysis of pretest to posttest gains. A large number of students (28) did not take the posttest at school 2 because the teacher gave the posttest on a day when many students were absent due to a school field trip. The teacher was unable to take class time to administer the tests to the missing students.

Embedded assessments

- *Use of the Control-of Variables Strategy (CVS).* I used students' experimentation sequences to compute a CVS proportion, the percentage of each group's trials that were part of a controlled comparison using the variable appropriate to the chosen investigation question among successive trials. I used only trials that specified one of the three investigation questions (i.e. not *Just exploring*) for this score. This proportion score indicates the extent to which students planfully used CVS to investigate the three questions. I used a proportion score for Study 3 (rather than the knowledge integration score used for Study 1 and Study 2) in order to capture how frequently students used CVS. This measure provides a clearer indication of the extent to which students relied CVS to investigate the situation.
- *Interpreting and constructing Airbags graphs.* Twelve INTERPRETATIONS items asked students to interpret motion graphs from the visualization or generate graphs that represented a collision situation. These 12 items consisted of the nine INTERPRETATIONS items from Study 2 plus three additional items. These items were scored in the same way as the INTERPRETATIONS items in Study 2.

- *Understanding of covariation and thresholds*. Three items prompted students to explain their answers to the three investigation questions listed in Table 8. I used these responses to determine the extent to which each group attained covariation-based and thresholds-based understanding of the *Airbags* situation. To examine how frequently students attributed their findings to covariation, I counted the number of responses by each group that described covariation between the factor or variable in question and elapsed time before the driver encounters the airbag. To examine how frequently students attributed their findings to thresholds, I counted the number of responses by each group that used at least one of the thresholds in support of their finding.

Videorecords

I videorecorded 12 dyads across both the *isolate* and *compare* conditions as they engaged in the experimentation activity, so that I could closely examine the discussions that occurred during experimentation within specific dyads. I chose three of these dyads to illustrate how the investigation questions they received influenced their experimentation strategies and insights. A more detailed description of these three dyads and the reasons I chose them for analysis follows the presentation of the results of the comparison study.

Results

Impact of *Airbags* on students' understanding of motion graphs

Students made moderate, significant pretest-posttest gains [M = 29.99, SD = 6.66 (pre); M = 32.76, SD = 5.96 (post), $t(128) = 5.17$, $p < .001$ (two-tailed), $d = 0.44$]. Gains were positive for all five schools, and significant for four of the schools. Low prior knowledge learners made the greatest gains. The gains indicate that *Airbags* was successful in helping diverse learners interpret and construct motion graphs. Though the effect size of the gains is smaller than for Studies 1 and 2, in Study 3 all students had completed a kinematics unit shortly before studying *Airbags.* The gains thus represent value added to a traditionally taught kinematics unit.

The difference in posttest scores between the two conditions was not significant. As in Study 2, I did not expect differences because the experimentation activity comprised a small portion of the total time spent on *Airbags*. There were no significant differences between the conditions on the INTERPRETATIONS scores, suggesting that the *isolate* and *compare* prompts were equally effective in helping students connect the characteristics of the *Airbags* graphs to the collision events.

To examine subtle impacts of experimentation goals on students' strategies and insights, I sorted student groups into low, middle, and high tertiles according to their prior knowledge using the mean pretest score for each group. My exploratory analyses show variation in students' strategies and inferences by tertile. Though the students' different schools were not evenly distributed among the three tertiles, all three tertiles were represented by groups at each of the five schools.

Impacts of prior knowledge on students' strategies and inferences

CVS scores showed that the high tertile groups conducted a significantly higher proportion of controlled trials than the low and middle tertile groups [M = .79, SD = .30 (high), M = .45, SD = .35 (low/middle), t(79) = 4.39, p < .001 (two-tailed)]. This difference suggests that high prior knowledge students as a whole were more focused on controlling variables in their investigations.

Overall, about 30% of the students generated covariation-based explanations of the collision events. The percentage was somewhat higher for the high (38%) and middle (36%) tertile groups than for the low (19%) tertile groups, but this difference was not significant, indicating that a covariation-based understanding of *Airbags* was about equally accessible to students at all prior knowledge levels. However, just 2% of low and middle tertile students generated thresholds-based explanations, compared to 31% of the high tertile students. A Wilcoxon rank-sum test showed that the difference in the average number of thresholds-based explanations was significant [M = .59, SD = .98 (high), M = .038, SD = .28 (low/middle), U = 3.78, p < .001]. This result suggests that only sophisticated learners were able to achieve a thresholds-based understanding of *Airbags*.

Impacts of experimentation goals on students' strategies and inferences

CVS scores showed significant differences between the conditions only for the high tertile students [M = .94, SD = .06 (*isolate*), M = .70, SD = .35 (*compare*), t(27) = 2.22, p = .035 (two-tailed)]. A close examination of these data revealed that while virtually all the *isolate* groups devoted nearly all their trials to controlled comparisons between successive trials, one-third of the *compare* groups used at least half their trials for other strategies. (One of the case studies that follows will illustrate some of these strategies.) This result suggests that though most students used similar approaches to investigate the *isolate* and *compare* questions, the *compare* questions led some high prior knowledge students to use alternative strategies to CVS in their investigations.

There were no differences in the number of covariation-based explanations between the two conditions, indicating both conditions were equally effective in leading students toward a covariation-based understanding of *Airbags*. However, there were significant differences in the number of thresholds-based explanations for the high tertile students. Just 8% of the high tertile *isolate* groups generated thresholds-based explanations, compared to 44% of the *compare* groups. A Wilcoxon rank-sum test for the high tertile students showed that the difference in the average number of thresholds-based explanations was significant [M = .091, SD = .30 (*isolate*), M = .89, SD = 1.13 (*compare*), U = 2.09, p = .037]. This finding suggests that the *compare* questions helped high prior knowledge students achieve a thresholds-based understanding of *Airbags* while having no effect on lower prior knowledge students.

Case comparison: isolate vs. compare

To examine the effects that the compare questions might have had on the strategies and inferences of students, I selected the videorecords of three student dyads for detailed analysis. All student names used here are pseudonyms.

Case 1 ("Brett" and "Eric") presents a high tertile group in the *isolate* condition. Case 2 ("Joann" and "Linda") presents a high tertile group in the *compare* condition. These two cases illustrate how the compare questions might have helped high prior knowledge learners attain a threshold-based understanding. Student groups in these two cases had similar average pretest scores and embedded assessment scores and studied physics with the same teacher. Students in both groups were concurrently enrolled in calculus and thus had strong mathematics skills.

Case 3 (“Christine” and “David”) presents a middle tertile group in the compare condition. I have selected the students in Case 3 to illustrate why the *compare* questions might have had little effect on lower prior knowledge students compared to high prior knowledge students.

Case 1: Brett and Eric (*isolate* condition)

Overview

Brett and Eric studied the *isolate* version of *Airbags*. Table 9 shows their sequence of 10 trials. For their first four trials, they chose *Just exploring* as their goal, during which time they tested the default variable values, and tested highly contrasting situations across a wide range of values for each variable. In their the last six trials they conducted a pair of controlled trials for each of the three investigation questions. Their experimentation session lasted for about 10 minutes. During their experimentation session, Brett used the mouse to navigate the visualization, while Eric took the initiative in deciding on what trials to conduct. Despite rarely offering alternatives to Eric’s experimentation decisions, Brett did correct several of Eric’s misstatements, indicating that Brett was both knowledgeable and fully engaged in the experimentation activity.

Table 9. Brett’s and Eric’s experimentation sequence

Trial #	Trial Goal	Position	Velocity	Time	Outcome (safe/unsafe)
1	Just exploring	0.4	-5	0.05	S
2	Just exploring	0.32	-9	0.01	U
3	Just exploring	0.52	0	0.1	S
4	Just exploring	0.52	-1	0.1	S
5	Driver height	0.32	-5	0.05	S
6	Driver height	0.38	-5	0.05	S
7	Collision speed	0.36	-3	0.05	S
8	Collision speed	0.36	-8	0.05	S
9	Crumpling	0.36	-5	0.03	U
10	Crumpling	0.36	-5	0.08	S

Trials 1 – 4: Exploring extreme cases

Brett’s and Eric’s experimentation sequence shows that they did not test any position values below the position threshold of 0.30. As a lower boundary for the position value they used the smallest testable value above the threshold, 0.32. Their discussion as they prepared to conduct Trial 2 begins to illustrate why.

B: Short person’s closer…[Moves dummy to position of about 0.2]

E: Not THAT close, cause the airbag deploys all the way up there [Eric gestures to a point on animation where the airbag stops. Brett moves the position slider back to about 0.4 again]. Closer. That was the starting point. [Brett gradually moves the position slider to 0.32] Right there.

By design, the *Airbags* visualization allows students to test situations where drivers sit within the airbag's deployment zone, as in actuality it is the main reason for the risk of injury to short women. However, by the second trial, Eric had essentially removed the effect of the threshold from task entirely. He made clear with his description and hand gesture that he understood the nature of the airbag's inflation to the position of 0.30. However, his rationale for eliminating values below the distance threshold is not yet clear, as he did not explain why he believed they should not test position values below the threshold. His choice of the next highest testable value above the threshold suggests that the choice was not arbitrary. They revisited the rationale for establishing this minimum value as they conduct Trial 5.

Trials 2 and 3 also illustrate Brett's and Eric's prior conceptions of the role the variables play in the collision outcomes. In Trial 2 they test values for each variable that they believe contribute to an unsafe outcome, while in Trial 3 they test values that they believe contribute to a safe outcome. (Trial 4 repeats Trial 3 but for the most extreme non-stationary situation.)

E: …So he's going at high speed, he's close, and there's less crumpling of the car? OK, play it.
B: He's gonna die! [Laughter].
E: Play it. [They run Trial 2]
B: He's gonna die!
…
B: Let's do a safe one. [Laughter]
E: He's a tall dude.
B: Tall dude.
E: Moving slow, and then--
B: And a lot of crumpling.

Their exchanges illustrate the results they expected to observe (and did) for these two trials. The absence of any discussion about the reasons for the outcomes of these or any of the 10 trials suggests there was nothing they observed during the experimentation activity that challenged their expectations or prior conceptions, even though they neglected to attribute any of the outcomes to the position or time thresholds.

Trials 5 & 6: Examining the position variable

As Brett and Eric prepared to conduct Trial 5, Eric provided the only further discussion about his choice of the minimum position value.

B: All right. We need to figure out taller people or short people.
E: So let's just go short. [Brett moves the position slider to a value below the position threshold.] No, the airbag full deployment is right there at that thing [gestures toward the 0.30 mark on the animation], so if he's close, let's say that's the steering wheel, if he's really close, go back some, right, no, forw—right there. [Brett again moves the position slider below the threshold.] No, it's too cl—he's not gonna, his face isn't gonna be right at the steering wheel.

Here Eric reinforced his decision to make the lower limit of the position just outside the airbag's deployment distance. With the statement "his face isn't gonna be right at the steering wheel", Eric communicates his intent to narrow the range of testable values based on what he believed was possible for a real life situation. In this way Eric's choice illustrates a way students use an everyday understanding of the airbags situation to inform their experimentation choices. While Eric's decision to eliminate very small values of position from testing is reasonable, choosing to make the lower limit of testable values coincide with the extent of the airbag's deployment was less valid. There is no particular real life relationship between the extent of airbag deployment and a driver's seating position with respect to the steering wheel. In thise case Eric appears to have applied a critical distance value present in the Airbag's problem (the airbag deployment distance) and applied it arbitrarily to constrain his exploration of the position variable. Eric's failure to recognize the independence of the two distances effectively excluded variable values that were important for a complete understanding of the situation.

In Trial 6 they completed their controlled test for the position variable. The following exchange illustrates their rationale for designing Trial 6 and a (very) brief interpretation of the two controlled trials:

E: Short or tall. And now we have to move the guy back, cause he's taller. So we gotta keep everything except position. So move him back some. Like right there.

B: He's going to be safe, obviously.

E: He might not, let's just check. [They run trial 6] Yeah. So mark that as safe. OK, put the graphs for the previous two. [They compare the graphs of trials 5 and 6]

B: They're both safe.

E: Yeah. So let's go to the next question.

This brief interpretation of a controlled test raises questions about Brett's and Eric's goal in conducting their trials. Conducting a test whose outcome Brett found "obvious" and neglecting to discuss the reasons for the outcomes they observed suggests that they were not focused on augmenting their understanding of the situation. Conducting two tests that both have the same outcome while failing to even attempt to generate a different outcome suggests they were not trying to generate evidence for an effect of the position variable on the collision outcomes. In fact, their discussion does not provide any compelling reasons for conducting controlled trials at all. Their decision to conduct controlled trials appears to be only a formality at this point. Their discussion during the next two trials provides more insight into this possibility.

Trials 7 & 8: Examining the velocity variable

Brett and Eric aimed to isolate the velocity variable in Trials 7 and 8. The following exchange captures the entirety of their discussion about these two trials.

E: Do low or high speed more likely to be injured.

B: All right.

E: So we've gotta keep the dummy position constant.

B: I'd say, like, right here. [Moves position slider to 0.36]

E: So then velocity, let's do low speed. Alright, so. Play. [They run Trial 7.] So then, dummy veloc—dummy velocity. OK. So. So then, go to 8, negative 8. Then play. [They run Trial 8] He's still safe. It's safe. But I mean…

B: But we can like, tell where the position and velocity and time need to be, with every single type of person no matter what velocity no matter what position no matter what time.

E: Yeah, cause like, w--we're really, since we're doing like experiments, we can only change one of them, we can't change multiple ones.
B: Yeah.
E: Cause like in real life, there would be a combination of all three.
B: Mm hm.
E: So—more or less crumpling….

Their procedure for these two trials was nearly identical to the previous two trials. They varied only the target variable, achieved two safe outcomes, and engaged in almost no discussion of the results.

Brett's main comment is unclear. He could have been referring to their ability to predict the outcome for any set of variable values without having to conduct the trial, or that they could choose for any particular person the variable value required to achieve a safe outcome. The comment appears to communicate the idea that he considered conducting the trials formality that they perform for the purpose of completing the task that has been assigned to them.

Eric's rationale for his experimentation choices is more clear. He indicated his belief that strategies other than CVS were invalid, despite the uninformative nature of their approach. Furthermore, the distinction he made between their experiments and "real life" suggests he believed other strategies would be valid in a more authentic context. Though Eric did not elaborate on what he meant by "a combination of all three", the phrase suggests he is aware of possible interactions and tradeoffs that might occur between the variables. Brett's and Eric's language suggests they took a "schoolish" approach to the *Airbags* task. Their experimentation choices appear to have been governed by an effort to fulfill typical science classroom expectations, rather than to inform their understanding or generate evidence.

Trials 9 & 10: Examining the time variable

Brett and Eric used their final two trials to isolate the time variable. This exchange occurred as they decided how to conduct these two trials:

E: So--more or less crumpling. So keep the velocity constant. So then, if there's less delay, that means more crumpling. So then, yeah let's do less crumpling.
B: That's more crumpling.
E: That's—yeah. So just go low first. Like there's low crumpling, almost no crumpling. [They run Trial 9 and briefly interpret the outcome]
…
E: OK, so then, that was low speed, so this is high speed. All right.
B: More crumpling.
E: Yeah, more crumpling.

Here Eric demonstrated his attention to employing CVS, and his inattention to the nature of the variables within the context of the *Airbags* investigation. Initially, Eric incorrectly stated the relationship between the crumpling factor and the time variable. Eric's subsequent recommendation "just go low first" (combined with his correct statement of the relationship in Trials 2 and 3) suggest that this error was a result of indifference rather than a conceptual misunderstanding. Because the precise nature of the relationship between the factor and the variable (direct or inverse) would not change how they employed CVS, the only decision Eric believed they need to make is whether to test the “low” or the “high” value first. Later, even after being reminded of the correct relationship, Eric attributed the time variable to the wrong factor entirely (speed, rather than crumpling). By this time, Eric was no longer attending to the nature of the variables and appears to have almost completely sequestered the experimentation strategy from their understanding of the *Airbags* context. At this point, the variables might as well have been X, Y, and Z rather than position, velocity, and time.

Brett’s and Eric’s failure to consider the nature of the variables during their investigation could have contributed to their decision to apply the threshold distance to their range of testable position values during the early trials. Though their choice was informed by Eric’s effort to incorporate aspects of the real life situation into their strategy, a more careful consideration of the nature of the position variable might have prevented him from eliminating key variable values in *Airbags*.

Summary

Brett and Eric had productive ideas about the *Airbags* situation as they began the experimentation activity, but they did not conduct experiments in a way that built on these ideas. In at least one case, their prior ideas led them to eliminate variable values that could have substantially enhanced their understanding of the situation. Brett and Eric took a schoolish approach to experimentation by controlling variables in a way that neither produced evidence for their views nor augmented their understanding. They sequestered their experimentation strategies from the nature of the variables and the *Airbags* context, failing to make adjustments to their experimentation strategy that would have informed their understanding of *Airbags*.

Case 2: Joann and Linda (*compare* condition)

Overview

Joann and Linda (pseudonyms) studied the *compare* version of *Airbags*. Table 10 shows their sequence of 11 trials. They began their experimentation by carrying out an initial plan to isolate each variable. They quickly abandoned that approach and employed other strategies such as testing extreme values and incrementally varying individual variables. Their experimentation session lasted about 30 minutes. This analysis will focus on three excerpts that illustrate the evolution of their investigation strategies and the insights they achieved by using these diverse approaches.

Table 10. Joann's and Linda's experimentation sequence

Trial #	Trial Goal	Position	Velocity	Time	Outcome
1	Just exploring	0.02	-5	0.05	U
2	Just exploring	0.26	-5	0.05	U
3	Just exploring	0.52	-5	0.05	S
4	Driver height	0.44	-5	0.07	S
5	Collision speed	0.44	-10	0.04	S
6	Collision speed	0.44	-10	0	U
7	Crumpling	0.44	-5	0.1	S
8	Crumpling	0.34	-5	0.07	S
9	Crumpling	0.34	-5	0.06	S
10	Crumpling	0.34	-6	0.06	S
11	Collision speed	0.34	-10	0.06	S

Trials 1 – 4: Employing (then abandoning) CVS

Joann and Linda used their first three trials to isolate the position variable and test its full range. As they decided on the values for the fourth trial, Linda began to reconsider their approach.

L: I don't know. Maybe we just test ummm, like, test the position, at, like 3 different points, that just so--that just so many tests, never mind. Cause like what I was originally thinking is if we tested like the dummy position like here [points to left side of position slider], and then test it, like, test it there [left side of position slider], and have like, it--the velocity and the dummy time like still be here [left side of velocity slider]--

J: Mm hm.

L: --in both of them [left side of time slider], and then like, move it here [center of velocity slider], and test it again with dummy position, and move them here [right side of velocity slider], and test again with dummy position, and--

J: Oh, OK.

L: But that's a lot of trials, you know, to do for, the whole thing.

Here Linda appears to describe an exhaustive 27-trial sequence that would control for all three variables at three values of the other variables. She was initially discouraged by the sheer length of this approach, but after choosing intermediate values for trial 4 and discussing the outcome, their discussion about the effect of the collision speed on the other two variables prompted a change in approach:

J: See we keep--we kept all that the same, but the farther away it was, the safer. Keeping the velocity and time on track. Because I would imagine let's say you had—it was closer, and it goes right there, and you had the dummy time at like 1 full second, that would give it more time to inflate. So then the dummy wouldn't start moving until 7.5.

L: Right, OK.

J: So it's an extra half—time…or whatever… And if you decrease the velocity, they can move slower, which I'm assuming is a slow crash, like slow impact crash.

L: OK.

J: Then it all falls back to what we said originally, the crash, the speed of the crash dictates if position and dummy time, you know, the crumpling of the car, would have an effect.

L: Yeah. I don't understand why we have to do different tests for each three different sections? You know? You click on them and be like whatever trials for this, kind of—cause it looks like we're kind of figuring it out as we're looking at this.

This episode sheds light on their rationale for changing their approach. Joann summarized the results of their initial controlled trials using covariation ("the farther away it was, the safer"). Their approach appeared to change when they recognized that a simple covariation-based explanation was insufficient to address the *compare* questions. Their discussion turned to the tradeoffs between variables (e.g. reducing the position but increasing the time) and their pre-experimentation hypothesis (that the speed "dictates" the effects of the other variables). Linda's final comment suggests she believes CVS would not adequately address the *compare* questions. At this point they began conducting trials in a more spontaneous way in an effort to simply "figure it out."

Here their experimentation approach diverted substantially from Brett's and Eric's approach. Brett and Eric failed to change their approach appropriately in response to observe outcomes from previous trials and conducted trials without an apparent regard for the outcomes. Joann and Linda on the other hand recognized early that their initial strategy would not sufficiently inform their investigation of the *compare* questions. In this way the *compare* questions seem to have prompted their change in strategy.

Trials 5 - 8: Exploring extreme values and intermediate values

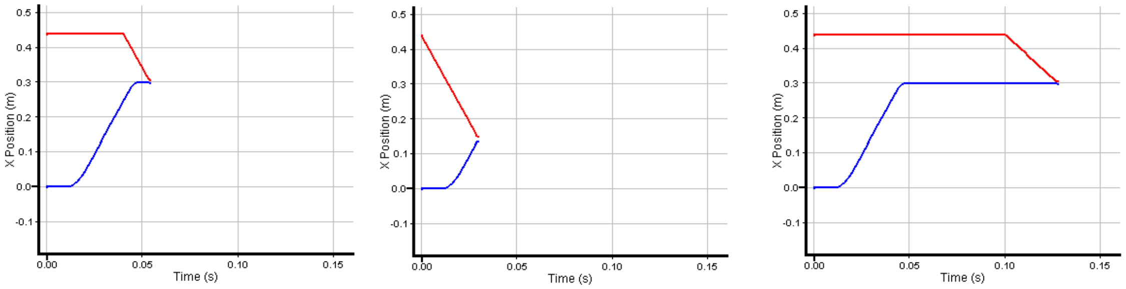
In trials 5, 6, and 7, Joann and Linda explored extreme values of the velocity and time variables (illustrated in Figure 7). Though they never explicitly stated their rationale for testing these extreme values, their approach appears to be exploratory and spontaneous, similar to Brett's and Eric's early trials. In trial 5 they simulated a "high impact crash", setting the velocity to the "fastest possible." In trial 6, they "try it with dummy time if we put it at, like, zero." In trial 7, they tested the maximum crumple zone ("All the way?"). They conducted these three trials in fairly rapid succession and with little discussion of the outcomes. This sequence culminated in this exchange after the completion of Trial 7:

L: All right, so maybe change like our dummy time?

J: Well, I guess, dummy time can also have an effect with position, cause like I'm saying, if you have no dummy time, then how close you are to the steering wheel matters a lot.

L: Yeah, so maybe we'll change that too. I mean test that. Yeah, change our dummy position.

Joann's conjecture about when the position value "matters" illustrates her realization that that the importance of driver's position depends on the value of the time variable. Her comment thus suggests not only a mindfulness of the investigation questions, but also constitutes the first mention of the interaction between the position and time variables. This comment marked the end of their spontaneous exploration of extreme values and the beginning of a more purposeful investigation of the variables' effects on the collision outcome. Their recognition and investigation of the interaction ultimately served as a stepping stone to characterizing the situation based on the thresholds. Their awareness of the interaction prompted them to explore it further in the subsequent trial.



(a) (b) (c)

Figure 7. Linda's and Joann's experimentation sequence exploring extreme values in trials (a) 5 (b) 6 (c) 7.

In Trial 8, Joann and Linda tested a position and time values near their respective threshold values to further examine the interaction between position and time. Their choices of near-threshold values for the interacting variables is important, as values near thresholds make outcomes more sensitive to changes to the variable values. Though they might not have consciously tested these values for this reason, these near-threshold values they chose for Trial 8 helped them refine their understanding of the interaction between position and time:

J: Wait a minute. Yeah. This is our other trial. [They compare Trials 7 & 8, (Figure 8)]

L: Even if you're sitting closer, that took off like a whole second, .34, .44? So if you're sitting closer, and your car has a pretty decent length crumple time I guess…

J: Oh, that's when the crumple time can play an effect.

Joann's observations continued to be informed by the *compare* questions. Here she correctly observed that a long crumple time is more critical for a driver who sits nearer to the steering wheel. Though this observation does not explicitly identify the time threshold, it approaches a thresholds-based understanding and serves as a stepping stone to their explicit understanding of the thresholds they eventually achieve.

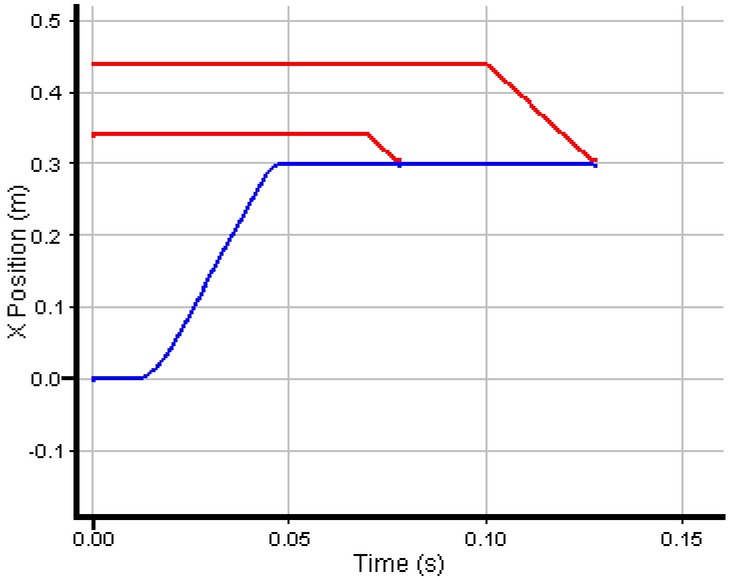


Figure 8. Joann's and Linda's comparison of Trials 7 and 8.

Trials 9 - 11: Conducting controlled trials to examine variable interactions

In their final three trials, Joann and Linda returned to conducting controlled comparisons. Unlike their initial approach, which used CVS to examine covariation across the full range of the position variable, their final approach was informed by their awareness of possible variable interactions and tested only narrow ranges of the variables near the threshold values.

L: Let's leave the position, and like, make the dummy time smaller. I was just going to just say, I kind of like playing with this, but—
J: I know, we're going to have like 100 trials! OK.
L: So dummy time maybe like, a little bit smaller. I mean it's also going to matter how fast you're going with velocity too, you know?
J: Yeah, that's very true. But I guess just keep it consistent for now.
L: OK.
J: Alright, take off .01, cause last time it--it was .07, so now .06?
L: Then it'll probably be a lot closer.

Though Joann and Linda did not articulate their motivation for conducting Trial 9, they did clearly communicate an intention to control variables. They also indicated an awareness of a possible interaction with velocity variable, which they examined in Trial 10 when the outcome of Trial 9 confirmed their expectations. They appear to use Trial 10 to examine the role of the velocity variable, which they initially conjectured as the most important factor and had still not been able to confirm or disconfirm to this point.

L: Yeah. Can we leave that [time variable] the same? So let's try and change the velocity to be like higher, to see if it does something. It would be faster.
J: Then, yeah, it would definitely meet sooner.
L: Yeah but like how fast to?

J: OK. Um. I don't know. [They adjust the velocity variable to -6 and run Trial 10.]

Their use of small adjustments to the variables in Trials 9 and 10 is important, though they do not seem to be explicitly aware of it. They seem intuitively to recognize that their trials near the thresholds yield greater insights about the variable interactions. After Trial 7 their variable choices gradually approached the threshold values, resulting in outcomes that approached the boundary between safe and unsafe.

Trial 10 illustrated to Joann and Linda the invariance of the outcome due to the velocity:

L: …So then that one doesn't make much of a difference.

J: No. OK.

L: So the speed didn't make a difference in that--

J: Oh. Doesn't that counteract what we first said?

L: Yeah.

J: I mean, contradict?

…

J: I'm thinking that, he doesn't start moving until, I don't know, I can't put it in words. Even though he's moving faster, by even a little bit, it doesn't really have much of an effect because then, the amount of time—

…

J: But I think that's also, that [crumpling time] would be more important if it was a slow, like slow crash, slow impact. If it was a greater impact, then the time probably wouldn't matter that much. Let's test it. I want to see that. [They adjust the velocity variable to -10 and run Trial 11.]

Here, Joann and Linda recognized that the speed "didn't make a difference" in this particular situation, and that this observation ran contrary to their initial hypothesis that the speed would "dictate" the outcome. Their understanding of the reason for their Trial 10 observation was tenuous, however. Rather than generalize the result of Trial 10 toward an understanding of the thresholds, Joann resisted changing her initial view and attempted to make it work for a different situation ("greater impact"). Only after this final trial were they able to articulate the nature of the time threshold. (Figure 9 illustrates the comparison of the final three trials that made the time threshold apparent.)

L: So it's the same, the speed doesn't change anything, it's the same graph we just had.

J: Oh, in that case, cause he won't start moving until—he won't start moving *period* [verbal emphasis] until—this has already been inflated no matter how fast he's going. Yeah. Because he won't start moving until .06 seconds has gone by.

L: Right.

J: So he could be moving at 100,000 miles an hour and he won't hit it until the airbag's already inflated, according to how we set this up.

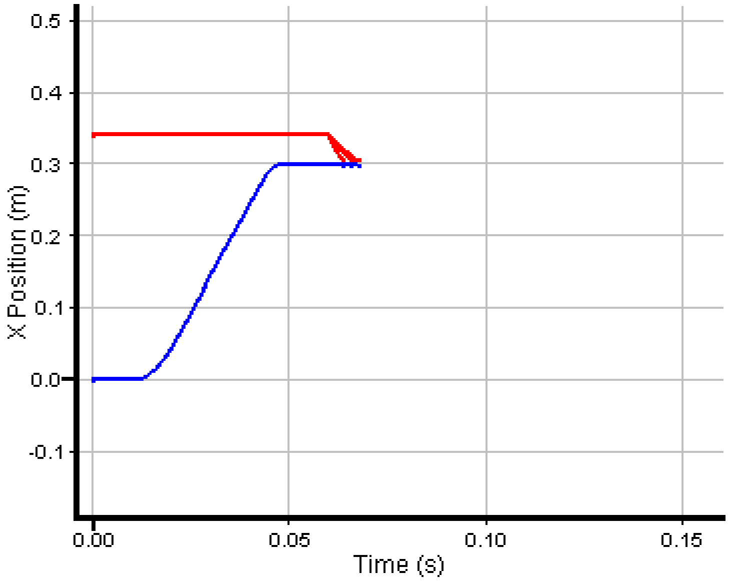


Figure 9. Linda's and Joann's variations to the velocity variable in Trials 9-11.

At this point, the recognition of one threshold appears to have catalyzed their understanding of the other threshold. They did not require any further trials to articulate the position threshold, as extending their understanding of the time threshold to the position threshold was a relatively small conceptual step.

> J: if we move it closer to the steering wheel, then dummy velocity and dummy time wouldn't matter, because as the airbag starts inflating, he'd be in the way, just him sitting here, just him.

Summary

Joann and Linda began their investigation by using CVS in the traditional way to illustrate covariation. This approach proved insufficient to answer the *compare* questions once they recognized the role of variable interactions in determining outcomes. Spontaneous exploration of extreme values then helped them refine their understanding of these variable interactions. Finally, controlled trials over narrow ranges of the variables that gradually approached the threshold values aimed to examine these interactions in more detail. These last controlled trials helped them to generalize their understanding of the interactions toward an understanding of the thresholds. Having to answer the *compare* questions forced Joann and Linda to consider the conditions under which certain variables made the biggest difference. This investigation of conditional situations and interactions ultimately served as a stepping stone to identifying the thresholds, which in *Airbags* are a special kind of variable interaction.

Discussion: Brett and Eric vs. Joann and Linda

The stark differences between the first two cases illustrate how the *isolate* and *compare* questions might have led students to reason about the *Airbags* situation. For Brett and Eric, the *isolate* questions appeared to provoke a "schoolish" interpretation of the task. They viewed the task as a simple covariation problem (a common class of problem in school science), and as a result they limited themselves to a predetermined pattern of using CVS rather than spontaneously employing diverse strategies to investigate new questions. They prioritized valid implementation of CVS over gaining insight. They sequestered their understanding of the *Airbags* situation from their investigation strategy and, more generally, from their conceptions of the *Airbags* task and the real life practice of science. Though their strategy was “valid” (as judged by criteria often imposed by classroom science), it was not especially informative. Their analysis did not go beyond a superficial characterization of the variables.

The investigation questions in the *isolate* condition did not challenge Brett and Eric to explore other strategies or seek to deepen their understanding of the situation. It is unclear whether the *compare* questions would have prompted them to use other strategies (which they believed were effectively disallowed), but the strategies they used would have been insufficient to answer the compare questions. It is possible that the *compare* questions would have encouraged Brett and Eric to go beyond characterizing the covariation relationship between each variable and the collision outcomes and to attempt to distinguish between the effects of each variable. These efforts might also have led to a discussion about at least one of the threshold values.

In contrast, the *compare* questions prompted Joann and Linda to incorporate a wider range of strategies to elucidate variation patterns. They conducted trials with the intention of understanding the relationships between variables and outcomes and the mechanisms that governed these relationships. Though their initial efforts to use CVS did not yield useful ideas, in the end they spontaneously used CVS to achieve a highly nuanced understanding of *Airbags* by building on ideas they refined using other strategies. The *compare* questions encouraged Joann and Linda to deeply consider the nature of the variables and to make important and meaningful distinctions between them. The complexity of Joann’s and Linda’s analysis illustrates why the study reveals differences in insights only for high prior knowledge students. Joann and Linda needed highly sophisticated knowledge of experimentation strategies and graph interpretation (as well as a high level of persistence) in order to reach their advanced level of understanding.

Case comparison: high vs. low prior knowledge

The previous case comparison illustrates how the *compare* questions helped sophisticated students achieve a nuanced understanding of *Airbags*. Why did the *compare* questions not help lower prior knowledge students achieve this threshold-based understanding? Two possible factors might have contributed to this effect:

> (1) In Study 1, students’ pretest scores correlated significantly with the number of trials conducted, the variability of their trials, and their CVS scores and proportions. These relationships suggest that students with less sophisticated graphing knowledge also used fewer experimentation strategies to investigate the questions. Students who could conduct only simple controlled tests might have been able to achieve a covariation-based understanding of *Airbags* in either the *isolate* or *compare* condition, but would have lacked the ability to use the strategies necessary to understand the thresholds.

(2) Lower prior knowledge students had less sophisticated graphing knowledge, as measured by pretest scores. Students who struggled to interpret the graphs might have had more difficulty interpreting and comparing the outcomes of their experiments based on the graphs in the visualization. Difficulties with graph interpretation might have precluded students in the *compare* condition from noticing the thresholds because the graphs facilitate observation of the thresholds. Covariation, on the other hand, is a topic commonly addressed in traditional science and mathematics curricula and was likely more readily observable to students from controlled experiments without sophisticated graphing knowledge.

Here I examine a third case, Christine and David (pseudonyms), to illustrate the role that graphing knowledge and experimentation knowledge might have had on some of the low and middle tertile students. The analysis of Case 3 that follows examines how a student group with less sophisticated prior knowledge struggled to achieve the insights that Joann and Linda did. Christine and David studied the *compare* version of *Airbags*. Their combined pretest scores were near the median for all groups in this study. Individually, Christine scored slightly above the median and David scored slightly below the median.

Case 3: Christine and David (compare condition)

Overview

The initial plan Christine and David articulated before conducting their trials was vague and did not communicate any intentions to use specific strategies. Table 11 shows their sequence of eight trials. An inspection of the choices they made for their eight trials suggests they deliberately employed CVS to investigate all three investigation questions, an observation that the following analysis of their verbal discussion supports. Their experimentation session lasted for about 20 minutes. Throughout the session, Christine used the computer and David recorded observations using paper and pencil.

Table 11. Christine's and David's experimentation sequence

Trial #	Trial Goal	Position	Velocity	Time	Outcome
1	Driver height	0.52	-5	0.05	S
2[†]	Driver height	0.26	-5	0.05	U
3	Driver height	0.1	-5	0.05	U
4	Collision speed	0.26	-10	0.05	U
5	Collision speed	0.26	0	0.05	U
6	Collision speed	0.26	-10	0.05	U
7	Crumpling	0.26	-10	0.1	U
8	Crumpling	0.26	-10	0.03	U

[†] Christine and David conducted Trial 2 was conducted twice consecutively with identical variable values.

Trials 1-3: CVS supported an initial covariation-based conjecture

Like most groups, Christine and David began by investigating the height variable. In Trial 1, they explored the situation for a tall driver using the maximum allowable position value. They interpreted the outcome of Trial 1 in the following exchange:

D: So, the taller you are, the more delay time it is—and the longer it takes for you to reach the airbag.
C: Yeah.
D: But it's still going to be fully inflated, so you're still going to have a comfortable, nice pillow.
C: So, the taller you are the safer you are.
D: [writing] Taller equals safer.

Christine's and David's interpretation of just one trial illustrates their propensity to attribute the experimentation outcomes to covariation. Despite not having multiple trials to show a trend, David associated the driver's height with the amount of time required to reach the airbag. This led to a characterization of the outcomes in the form more x = more y. Statements of this form appear periodically during their experimentation sequence. This interpretation of their very first trial suggests that upon beginning the experimentation activity, they expected (or assumed) that the variables would covary with the driver's risk for injury.

In Trial 2, they implemented CVS by leaving the velocity and time variables unchanged from Trial 1 and testing a "medium" value of the height variable:

C: Well, you want medium height, and then—
D: Yeah, yeah, right in the middle.
C: [Moves slider to position value of 0.26] That's medium, right?
D: Right there, right there. [They play Trial 2]
C: Dang!
D: Not safe.
C: That is NOT safe.
D: Medium to small height equals dead. [Writing] Medium to small height equals not safe.

Their choice of 0.26 for the position variable is half the maximum allowable value of 0.52. This choice of half-maximum was probably not arbitrary, as most of the variable choices that were not boundary values occurred at the half-maximum value for the other trials as well. Choosing the half-maximum value or midpoint as an intermediate value is generally sensible. Because in *Airbags* this position value lies below the position threshold, it might have led some students toward recognizing the threshold. However, Christine and David chose to interpret Trial 2 using their covariation view, possibly preventing them from recognizing the threshold. Continuing to return to the half-maximum or midpoint value as a "neutral" value ultimately proved to limit the range of outcomes they could observe during their experimentation.

In Trial 3, Christine and David confirmed their hypothesized covariation relationship by testing a small value for position ("small is dead"). They viewed the graphs from all three trials at once (Figure 10) and briefly discussed them:

C: [Pointing to the airbag graph] This is the airbag where it's supposed to be, where it ends.
D: So these are the height trials, there are the three height trials.

C: Mm hm.
D: So, what's next?

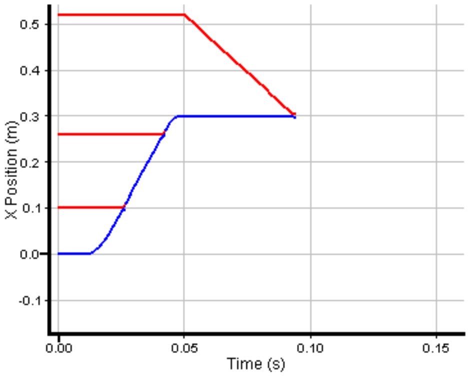


Figure 10. Christine's and David's comparison of their first set of controlled trials. Trial 1 is the upper red line, Trials 2 is the middle red line, and Trial 3 is the lower red line.

The trial comparison shown in Figure 10 illustrates a missed opportunity to identify the position threshold. Even though Christine remarked where the airbag is "supposed to be" with respect to Trials 2 and 3, they failed to attribute the outcomes at small position values to the threshold. This failure appears to stem from their effort to merely confirm the covariation relationship they expected from the beginning, rather than to investigate the dynamics of the situation in greater detail. Their interpretations of all three trials to this point focus only on the outcomes of the trials, rather than the events and conditions that led to the outcomes. Furthermore, they had no reason to revise their initial conjecture ("taller equals safer") because their subsequent trials did not generate unexpected results. Their failure to recognize the threshold at this point led to a series of uninformative trials to follow.

Trials 4-6: Generating and interpreting identical outcomes

In trials 4-6, Christine and David conducted trials with outcomes identical to those of Trial 2 (Figure 11). As they decided on values for Trial 4, they discussed the appropriate value of the position variable:

C: So let's put our dummy back in the middle, so wherever he was, it was…where was it?
D: Um…
C: Hold on, reset.
D: I think that's it, right there. That might be the neutral.
C: Neutral?
D: Well, the starting—where it started at.
C: Well, they were all in the middle.
D: Press reset again. Mm. Alright, just leave it right there [position value of 0.26]. That'll be our neutral.

Rather than a position value that would allow them to observe some variation in the outcomes with the velocity variable, they again chose the half-maximum value as their "neutral" position and time values for the next three trials. As a result, their three controlled trials for the velocity variable generated exactly the same outcome (Figure 11). Their choice of 0.26 for position confirms their failure to recognize the role of the position threshold in determining the collision outcomes.

Whether David uses the word “neutral” to mean “unbiased” or “in the middle is unclear from this exchange. However, their frequent choice of the half-maximum (88% of all non-boundary value choices during the entire experimentation sequence) suggests that this choice is not arbitrary. The half-maximum choice is likely part of David’s beliefs about valid experimentation to examine covariation.

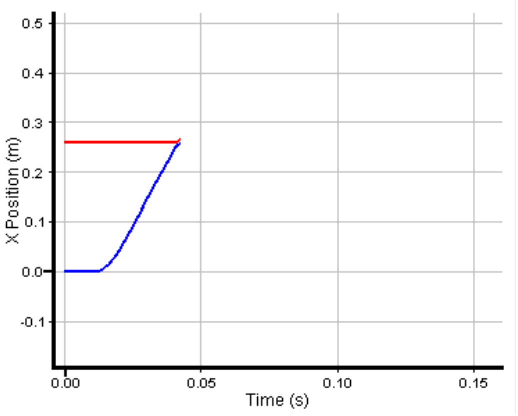


Figure 11. The identical graph Christine and David generated in Trials 2, 4, 5, 6, and 7.

Following Trial 4, Christine appears to suggest that their current experimentation approach may not be addressing the complexity of the situation:

C: Or do you want to do, um, let’s say like, I mean, you don’t have to do a whole bunch of trials to end the possibility, for, you know, it’s going to be like, it’s going to the fastest while the person’s really short, or gonna be the person’s really tall and it’s slow, or something like that you know, it’s always different possibilities, so it’s not definite, but, I don’t understand how we’re supposed to account for each one.

D: Yeah, just do another trial, I’ll see, see how that one is.

To this point in their experimentation sequence, Christine and David had about equal input in choosing their variable values. Here, Christine appears to suggest here that their current approach of changing one variable while holding the other two variables at intermediate values may not illustrate the full range of possibilities. Her articulation of two alternative trials (“fastest while the person’s really short”, “really tall and it’s slow”) that do not conform to their current approach suggest a vague awareness of variable interaction. Unfortunately, Christine was not able to articulate the idea clearly enough for David to understand, or for either of them to act on.

Christine’s observation here resembles Joann’s observation of variable tradeoffs following their set of controlled trials. Joann’s and Linda’s sophisticated understanding of the variables helped them explore variable interactions using a new experimentation approach. In contrast, Christine’s and David’s more tenuous understanding of the variables left these more complex ideas about the *Airbags* situation hanging in midair. Their difficulty conceiving of more complex variable relationships led them back to their initial approach rather than toward spontaneous exploration or a new strategy.

After Trial 5, David noticed the identical outcome to Trial 4:

D: Um, wait, go back to the other graph? Isn't that kind of like the same?
C: Yeah. Yeah. So…
D: Mm.
C: Why are they the same?
D: All right. Faster speed equals less safe.

Christine and David had distinctly different responses to the outcome of Trial 5. Christine appears to be interested in what causes the results—this was the first time during the experimentation activity that either of them asked a "why" question. David did not seem interested in why, however. He continued to seek evidence for covariation despite producing two identical trials. His repeated attempts to characterize all outcomes in the form "more x = more y" suggest he was unaware of any other types of relationships that could exist between variables. Trial 6, though a repeat of Trial 4, finally broke David's tight grasp on using covariation to explain his observations:

C: …It still looks the same!
D: Well that's like the airbag hitting *him* [verbal emphasis]. So he's like, driving driving driving driving driving, and the airbag's coming, psh, and it's hitting him. And then it will be like stopped, and he didn't go into it, it just kind of blew into his face, so that means that he'd be…not good.
C: Yeah, but I mean like, it still looks the same as if it was going, slower. Still the same effect, the position doesn't change so that person stays the same height obviously.
D: All right, so… really, all right, so., the slower he went, it was pretty much the same right?
C: Huh?
D: When we did it really slow it was like the same as really fast right?
C: Mm hm.
D: So slower is still less safe.
C: The reality, wouldn't the slow, I mean, wouldn't the slow, if it's slow then the way the person's moving—
D: I think I kind of get it, like. Like, um, our hypothesis was, you know, for this, the height made a difference, like the taller you are, then the safer you're going to be, and the smaller you are, the not safer you're going to be. And we thought it was really the speed that was going to affect it, but, whether you're going slower or faster, the airbag coming out and hitting you [gestures hand toward face], you know—
C: The same.
D: It's gonna be the same.
C: Oh OK, I get you.
D: So really, the speed doesn't affect it.
C: OK.
D: And—

C: The speed, the airbag does, but the speed of the car wouldn't. We were thinking along the speed of the car.
D: But right now, it's more the height.

This exchange illustrates how David was able to connect the outcomes of the experiment to the collision events. In this episode he was able to integrate two previously isolated ideas: (1) The driver is at risk for injury when the airbag strikes the driver and (2) the outcome is the same, regardless of the collision speed. Though David attempted once more to force these ideas into a covariation view ("slower is still less safe"), a new "invariance" view appears to have suddenly clicked into place. Furthermore, his explanation evolves out of an effort to determine which variable "made a difference", suggesting the *compare* prompts were integral in his ability to generate an explanation that accounted for his observations. After achieving an unexpected result, David was able to revisit his initial hypothesis and add invariance alongside covariation to his repertoire of ideas for interpreting experiments.

Though in this exchange Christine and David reached a level of understanding about the *Airbags* situation that many student groups did not reach, this episode still represents a missed opportunity to incorporate the distance threshold into their understanding. Christine and David appear to have overgeneralized their interpretation of Trials 4-6 to all situations, rather than only to situations where the driver is positioned below the threshold. This overgeneralization likely occurred because of their failure to explore interactions between the variables. Joann and Linda were successful at investigating the effect of one variable on another variable's role in determining the outcome (e.g. "if you have no dummy time, then how close you are to the steering wheel matters a lot"). Christine was unable to articulate the rationale for testing these scenarios, and David did not feel a need to consider them at all. Even when prompted to *compare* the variables, Christine and David still *isolated* the variables from each other in almost every sense.

Christine's and David's ability to distinguish invariance from covariation, apparently facilitated by the *compare* prompts, was a significant advance in their knowledge of experimentation. In this way the *compare* prompts offer promise as a method for guiding less sophisticated students toward more informative experimentation.

Trials 7-8: Interpreting the *Airbags* graphs

During their first six trials Christine's and David's interpretation of the visualization graphs appears to have played a small role in their understanding of the situation for two main reasons: (1) the range of outcomes they achieved was very narrow, and (2) they found the outcomes they generated easy to interpret using the animation. In Trials 7 and 8, their interpretations of the collision events were informed largely from their interpretations of the graphs. Why they turned their attention to the graphs in these last two trials is not entirely clear, but their difficulty with the time variable might have led them to the use the graphs to help them interpret the collision events.

C: Crumpling. Dummy time? What's that for?
D: Uh, I think that's um, the delay time, which is, uh--
C: So you want to do delay time, or not?
D: Yeah.
C: Yeah?

D: Well I think that's still like on the crumpling cause the crumpling is like, how um, like, the time in which he reaches the steering wheel and airbags.

Even though *Airbags* defines the delay time explicitly for students before the experimentation activity, Christine and David struggled to achieve a precise understanding of the time variable. Though their interpretation of the delay time did not differ critically from how the variable is defined in *Airbags*, their exchange suggests they were more uncomfortable with the nature of this variable than the other two.

Two exchanges are particularly revealing about how their interpretations of the *Airbags* graphs informed their understanding. First, using the graph to interpret Trial 7 (which was identical to the graphs of Trials 2, 4, 5, and 6) led them to a different understanding than when they used the animation in the previous trials:

D: All right so…the airbag, was moving, was accelerating, starting to move at a constant speed when he was acc—was beginning his acceleration, it looks like. So—

C: So they basically hit each other at the same time he started to accelerate, right?

D: Yeah.

C: This one's a lot more difficult.

Even though the outcome of Trial 7 replicated Trial 6, their interpretation differed. After Trial 6, David remarked "he didn't go into [the airbag], it just kind of blew into his face." After Trial 7, however, both Christine and David agreed that the driver and airbag "basically hit each other." The discrepancy appears to stem from the driver's small movement resulting from the impact of the airbag. In Trial 6 they interpreted this movement correctly, but in Trials 7 and 8 they interpreted the movement as having occurred *before* the collision.

This difference in interpretation of the same outcome suggests a failure to make two types of knowledge links that help lead to success in the *Airbags* activity. First, it suggests they viewed their three sets of controlled experiments as completely distinct. Even though they achieved the same outcome in Trial 7 as in their previous set of controlled tirals, they did not take advantage of the knowledge they had developed conducting these trials. Second, it suggests a disconnect between their views of the graph and the animation. Their discussion of the velocity graph of Trial 8 (Figure 12) further illustrates how they sequestered their interpretations of the graph and the animation.

C: [studying the velocity graph] These two are the same, they're on the same line, but then he stayed at a constant speed longer than the airbag, the airbag started out immediately, then accelerated—

D: Mm hm.

C: —then went to a constant speed, then decelerated. Then he stayed at a constant speed, accelerated, because he's going in the opposite direction, stayed at a constant speed, then decelerated.

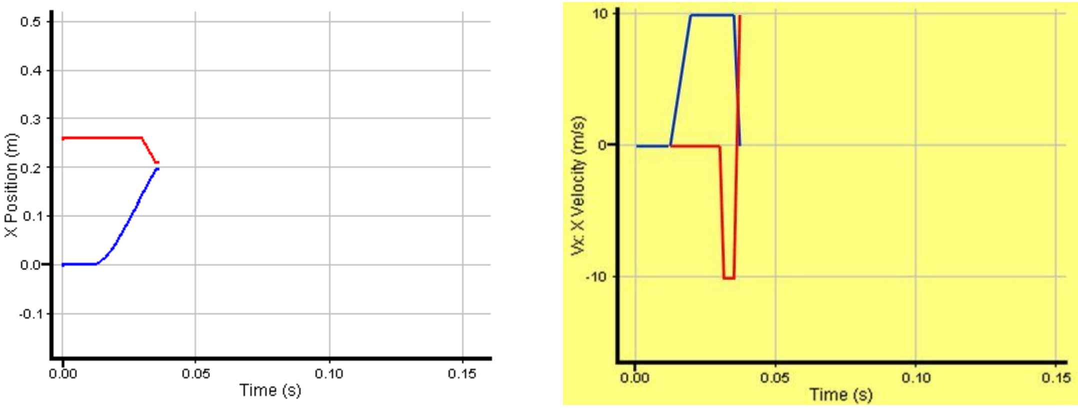
D: So, would he be safe?

C: [Laughs] I don't know. They bounced off of each other.

D: So, all right, well…

C: I think we should replay it.

D: All right. Well so as it’s decelerating, he’s accelerating, which means that it’s already, you know, it’s already out and it’s decelerating, you know, coming to a stop. So it’s already pretty much out as he’s accelerating into it.



(a) (b)

Figure 12. Christine and David’s Trial 8 (a) position graph (b) velocity graph

Christine was able to use the velocity graph to describe the kinematics of the airbag and the driver with reasonable accuracy. However, as with Trial 7, her kinematical interpretation neglects to refer to relevant collision events—namely, whether the airbag had finished deploying when it encountered the driver. David then misattributed motion resulting from the airbag’s impact with the driver to the airbag completing its deployment. The ease with which both Christine and David interpreted other trials with similar outcomes earlier in the activity (e.g. Trials 2 and 3, Figure 10) without even referring to the graph suggests the graphs were more confusing than helpful. They appear to have struggled to map their understanding of the collision events onto the characteristics of the graph. They were therefore unable to use the graphs to deepen their understanding of the situation.

The conclusion of their investigation merits a brief discussion. Christine’s and David’s attention to the *compare* prompt during their investigation of the velocity variable suggests they understood the task before them. However, they ended their investigation without even the slightest effort to compare the time variable to the other variables. That is to say, they appeared simply to ignore the task they had been assigned. Their written summary of their findings about the time variable following the experimentation activity confirms their inattention to the question at hand: “With less delay time, the driver is decelerating as the airbags are decelerating; therefore making a safer impact. So if the car crumples then the airbags deploy quicker and vice versa.” Here there is no mention of whether the crumple time “makes the biggest difference,” even though they responded directly to the compare prompt in generating this explanation.

Why did Christine and David cease to investigate the *compare* questions and instead answer as if they had investigated the *isolate* questions? The most likely explanation is that their use of CVS failed to yield any information that would distinguish between two covariation relationships. They were successful comparing the position and velocity variables because they observed covariation for one and invariance for the other. Though their approach could distinguish between the presence or absence of an effect of one variable on the outcome, it could not distinguish the nature of the effects using interactions or thresholds. Christine and David probably didn't answer the question posed to them because they simply didn't have any evidence with which to answer it. Instead, they answered with what they did discover using CVS, which was the nature of the covariation relationship.

Summary

Christine and David began the experimentation activity expecting to examine covariation relationships between individual variables and collision outcomes. They conducted a single set of controlled trials using a prescribed pattern for each of the three investigation questions, and they chose not to divert from their initial pattern in response to their observations. Though Christine appeared to show a vague understanding of variable interactions and see value in exploring other strategies, she was unable to articulate a way to incorporate the outcomes of these strategies into their understanding. David's strong commitment to CVS thus overrode Christine's tentative suggestions. Christine and David also struggled to connect characteristics of the visualization's graphs to the collision events, precluding them from using the graphs to refine their understanding of the situation. Though their use of CVS allowed them to distinguish between the presence and the absence of a covariation relationship, their approach did not allow them to make the finer distinctions between the variables' effects on the collision outcomes that would have led to understanding of the thresholds.

Discussion: Joann and Linda vs. Christine and David

Discussions between the dyads suggest Joann and Linda were able to achieve a more nuanced, thresholds-based understanding of the *Airbags* situation than Christine and David largely because of their (1) more sophisticated experimentation knowledge and (2) greater proficiency with graph interpretation.

Experimentation knowledge. Joann's and Linda's broad range of available strategies experimentation knowledge helped them reach sophisticated insights with *Airbags*. The multitude of strategies they were capable of using allowed them to examine variable interactions, which were critical to investigating the *compare* questions. They used their trials to examine specifically how an individual variable's effect on the outcome were influenced by values of the other variables. They conducted trials using values near threshold values and gradually approached these threshold values as they refined their understanding of the situation. In addition to testing extreme values, they conducted trials using incremental adjustments and similar outcomes in order to test subtle effects. They demonstrated a willingness to make "midcourse corrections" to their initial strategy, changing their approach in response to achieving unexpected outcomes and reaching new insights.

In contrast, Christine and David, limited themselves to a single strategy—a single set of controlled trials that tested boundary values for each variable, holding the other variables at intermediate values (generally the half-maximum). Though these controlled tests could reveal the presence or absence of a covariation relationship, they could not reveal subtle distinctions between the variables' effects, variable interactions, or threshold values. Christine and David were unable (or unwilling) to make adjustments to their strategy to examine new questions or accommodate new insights. Their experimentation approach appears to reflect their prior conceptions about experimentation as a way to illustrate covariation relationships. As a result, their strategy was indistinguishable from a group whose goal was to investigate the *isolate* questions.

Proficiency with graph interpretation. Joann and Linda interpreted the *Airbags* graphs by referring to relevant collision events rather than simply providing contextless, kinematical descriptions of motion. Their strong understanding of the graphs allowed them to make distinctions between trials and situations that would have been difficult to make using only the animation. Comparing multiple graphs made apparent fine distinctions between effects of the individual variables and between trials with similar outcomes. The graphs were essential to Joann's and Linda's ability to make sense of the *Airbags* situation by observing extreme values, examining variable interactions, and ultimately understanding the thresholds. Christine and David, on the other hand, though able to use the graphs to provide kinematical descriptions of motion, struggled to connect these descriptions to the collision events. As a result, their graph interpretations at times conflicted with their interpretations of the animation, hindering their sense-making. The graphs were designed to help students refine their understanding of the variables, but for Christine and David they were a source of confusion.

Discussion

The three cases studies in this chapter describe ways students conducted experiments and made insights with the *Airbags* visualization. This study incorporated diverse students who studied physics in different learning contexts, and no single group can represent the breadth of experiences students had with the *Airbags* module. These cases, along with the quantitative analysis, illustrate how the *isolate* and *compare* versions of *Airbags* led students toward different strategies and inferences. They also show how the level of students' prior knowledge about experimentation and graph interpretation could either augment or limit students' investigations.

The quantitative analysis showed that students who investigated the *compare* questions made more threshold based inferences and conducted fewer controlled trials than students who investigated the *isolate* questions. This difference occurred only for high prior knowledge students. For Brett and Eric, the *isolate* version of *Airbags* fell short in challenging their prior conceptions in two main ways. First, the *isolate* questions did not compel Brett and Eric to look beyond covariation for an explanation of their observations. Thus they settled into a simple pattern of controlling variables that failed to reveal important details about the *Airbags* situation. Second, the *isolate* questions led Brett and Eric to observe the norms of school science rather than fully explore the complexity of the situation. Though Brett and Eric recognized the presence of this additional complexity, they consciously neglected to investigate it in order to adequately meet perceived classroom expectations. The study's quantitative findings show that virtually all the high tertile *isolate* students shared Brett's and Eric's adherence to using controlled trials to investigate *Airbags*.

Joann's and Linda's discussions illustrate ways that the *compare* questions can guide students toward a sophisticated investigation that made key distinctions between the variables and examine variable interactions. Controlled tests designed only to examine covariation relationships could not have revealed the detailed insights that Joann and Linda reached during their investigation. In addition, their ability to interpret the graphs of each trial in terms of collision events greatly helped them continually refine their understanding during the course of their experimentation sequence. The quantitative findings suggest that a sizeable portion of the higher tertile *compare* students also employed diverse strategies to reach deeper insights about *Airbags*.

Though Christine and David received the same prompts as Joann and Linda and also entertained notions of examining variable interactions (however briefly), their understanding of possible variable relationships beyond simple covariation was tenuous. Joann and Linda lacked the requisite experimentation and domain knowledge to even begin to investigate these more complex relationships. As a result, they fell back to what they had likely learned in school science and performed their investigation as if they had been responding to the *isolate* questions. They conducted controlled tests in a similar way to Brett and Eric, even though their trials did not generate sufficient evidence to fully answer the *compare* investigation questions. The quantitative findings also reflect Christine and David's limited investigation approach, as the strategies of the *isolate* and *compare* students were statistically indistinguishable by all these study's experimentation measures for the low and middle tertile students.

The *compare* condition is effective for some students because the strategy students learn in traditional science leaves students without answers to the questions. They must continue to investigate in order to make the necessary distinctions to answer the questions. Understanding of the thresholds is one way to distinguish between the variables. However, the *compare* prompts were not effective for students who lacked the requisite knowledge about experimentation, graphing, or the airbags situation to make the necessary distinctions.

This study suggests ways that the emphasis of traditional science instruction on CVS may interfere with informative experimentation. For instance, David repeatedly sought evidence for covariation even when the outcomes of their trials suggested a different relationship. To David's credit, he was able to change his view by incorporating new evidence. His initial views, however, reflect the emphasis of school science on relationships simply characterized by "more x leads to more y."

In *Airbags*, where threshold values are important for a complete understanding of the situation, students who limit their investigations to isolating variables are less likely to make nuanced distinctions between the variables such as thresholds. Controlled trials at opposite extremes outcomes may fail to describe specific nature of the variable in question if it does not exhibit linear covariation with the outcome. A single controlled test also fails to examine possible variable interactions. Students who are reluctant to change their experimentation approach "on the fly" may be hindered by poor initial variable choices, as Christine and David where when repeatedly testing position values below the threshold.

Joann's and Linda's investigation illustrates the benefits of a balance between spontaneous exploration and a diverse array of controlled trials. Controlled trials that examine opposite extremes as well as similar values that approach thresholds provide unique insights into the situation. Joann's and Linda's willingness to make midcourse corrections was critical to examining new questions that were raised during their investigation. The strategic flexibility that Joann and Linda demonstrated might have had a bootstrapping effect. Adjusting one's strategy in order to examine new questions can lead to new insights, which in turn could help learners make increasingly more informed variables choices. That Joann's and Linda's variable choices gradually approached the threshold values for both position and time suggests they might have benefited from this sort of bootstrapping.

This study points to the value of providing opportunities for multiple approaches to experimentation rather than guiding students only to isolate variables. The findings suggest that instructional designers should balance guidance designed to promote CVS with opportunities to explore the nature and meaning of the variables. Designers of instruction should select problems where subtle distinctions such as thresholds are necessary for complete understanding. Many everyday problems such as decisions about drug dosage require an understanding of thresholds. This study demonstrates the value of connecting experimentation to real-life contexts such as airbags, where students can appreciate connections between science instruction and everyday life.

One limitation of this study was that the effect of the compare prompts was significant only for the most sophisticated learners. How might *Airbags* have guided less sophisticated students toward deeper insights? Upcoming versions of WISE software will be capable of providing immediate, automated guidance on students' experimentation strategies. Certain aspects of students' experimentation could serve as clear red flags for uninformative experimentation, such as not testing values on both sides of a critical value, not testing multiple values of variables to investigate interactions, or producing only one of several possible outcomes. Algorithms could respond to these red flags and provide just-in-time hints that would aim to nudge students toward a more informative approach. Algorithms based only students' variable choices are very limited, as discerning students' intentions often requires a more detailed examination of their thinking.

Alternatively, guidance for experimentation could explicitly emphasize aspects of variable relationships that go beyond covariation. More specific prompts than the *compare* prompts focus students on key differences between the variables could benefit less sophisticated students. For instance, prompts could ask students to identify specific variable values that are particularly important to the outcomes in order to attune them to the role of thresholds.

Additionally, Christine's and David's investigation suggests the connections students make between the domain of physics and real-life situations can be weak. *Airbags* could further strengthen these connections. I designed the activities in *Airbags* prior to the experimentation activity to strengthen students' connections between graphs and collision events. Students spent two class periods observing animations of the motion of the airbag and driver separately, describing the motion they observed, and sketching and assessing their own graphs. Though Christine and David generated strong responses to the prompts in these activities, they struggled to apply their understanding of the individual graphs to their experimentation. It is possible that the combination of the airbag and driver raised issues that were not present in the individual situations (such as their misinterpretation of the motion resulting from the airbag-driver collision). Their difficulties connecting graphs to the Airbags situation suggests that an additional scaffolding activity on the airbag-graph interaction prior to experimentation might have been helpful for Christine and David. This could have helped them use the graphs more to their advantage during experimentation.

Chapter 8: Discussion

Each of the three empirical chapters of this dissertation discusses issues relevant to the research questions of that particular study. Collectively, these three studies with *Airbags* also have more general implications for the design of inquiry instruction, visualizations, knowledge assessments, and classroom research studies. This chapter will examine how my five years of classroom studies with the *Airbags* module shed light on five large issues in science education research:

- *Design of inquiry instruction*. *Airbags* makes careful use of design patterns to help students integrate ideas. What patterns were particularly effective in helping students develop understanding of motion? How did year to year refinements to *Airbags* enhance their effectiveness? How did teachers complement the *Airbags* curriculum?
- *Design of dynamic visualizations*. The *Airbags* visualization combines multiple representations of motion, record keeping tools, and opportunities for student-initiated investigation. How did this design contribute to students' understanding of *Airbags*? How did refinements to this design enhance students' experience with the visualization?
- *Support for experimentation. Airbags* prompts aim to guide students toward conducting informative experiments. What strategies did students find informative in *Airbags*? How can guidance help students achieve the right balance of structure and individual initiative in their investigations? How can experimentation build on productive scientific ideas?
- *Design of assessments. Airbags* made use of diverse assessment instruments such as pretests/posttests, embedded prompts, logs of students' experimentation. Students generated written interpretations of graphs and constructed graphs to describe situations. How do these assessments capture the diversity and complexity of students' ideas? What are the benefits and limitations of experimentation logs?
- *Classroom contexts*. Conducting studies in school classrooms presented specific research challenges. What were the challenges of implementing *Airbags* in high school physics classrooms? How did the culture and expectations of classroom science influence students' interactions with the module?

This chapter will focus on aspects of design that I observed to be effective, as well as ways the designs of the module, visualizations, forms of guidance, and the assessments could be improved for future research with *Airbags*.

Design of inquiry instruction

Airbags benefited learners across the distribution of prior knowledge, albeit in different ways. Generally learners with initially poor understanding of motion graphs made large gains in their abilities to interpret and construct motion graphs, and they were able to generalize their knowledge from *Airbags* to other motion contexts. Students with higher levels of prior knowledge made smaller gains on their understanding of motion graphs but gained insights about the applications of physics to airbag safety.

I attribute the overall success of *Airbags* in diverse learning contexts to three main things: use of the knowledge integration design patterns in developing and refining the module, the scaffolding for the experimentation visualization, and the role of the instructors in facilitating discussions among students. Because the study designs did not vary basic aspects of the curriculum design, the quantitative data from the three studies do not provide specific evidence for the effectiveness of the patterns and the teaching decisions of the instructors. Rather, these observations are based on hundreds of interactions and conversations with students and teachers over five years of implementing *Airbags* in school classrooms.

Design patterns

The patterns provided a framework to make the diverse activities in *Airbags* cohere. Here I discuss some specific examples of how four of the patterns (*orient and elicit*, *predict-observe-explain-compare*, *experiment*, and *construct an argument*) help students link physics concepts to the theme of the *Airbags* investigation, distinguish between key physics concepts, test their own ideas, and evaluate the quality of their own learning.

Orient and elicit. The early steps in *Airbags* were successful in motivating the topic for study. The dramatic videos depicting crash tests and the real-time airbag deployment demonstration contributed to capturing students' interest, as did examples of visualizations that professional engineers used to study the safety of automobiles. The impact of these videos was likely enhanced for students who were of legal driving age, which included most students in the eleventh and twelfth grades. Many students told me they thought about the *Airbags* module every time they got into their car during the week I conducted the module in their classrooms. The topic of car safety thus resonated with these students at a personal level.

The success of the *orient and elicit* pattern goes beyond capturing students' interest. The videos and fatality statistics prompted students to consider the speed of the airbag deployment as a tradeoff between protecting and injuring the driver, depending on the conditions of the accident. Class discussions helped engage students in debating the design features of cars and airbags in light of this tradeoff. The limited information students had to debate this issue at the beginning helped make the use of physics to analyze the situation consequential.

It was interesting to observe the ways students interpreted the statistics that showed women to be at greater risk for injury from airbags. Initially, many students attributed the gender difference to factors such as driving skill or inattention to driving (resulting from applying make-up, for example). Many of these students also seemed aware that greater insurance premiums for men compared to women conflicted with these initial interpretations, motivating them to seek evidence for alternative explanations. Students' large, significant pretest to posttest gains on the Study 1 item (Appendix A, item 3) that asks students to explain whether a short man or tall women is at greater risk from an airbag injury illustrates the success of the module in this regard [M = 2.05, SD = 0.96 (pre); M = 3.76, SD = 1.17 (post), $p < 0.001$].

Predict-observe-explain-compare. Conversations I had with students as they generated their predictions and explanations suggested this pattern was extremely effective at helping students connect the graphs and the observed motion. *Airbags* augmented the POE sequence by further asking students to compare contrasting parts of the graph (e.g. slowing down vs. speeding up) or to compare the graphs of two different objects (e.g. motion in opposite directions). These prompts helped students distinguish different types of motion on the basis of the graph characteristics. Articulating these distinctions was key to students overall success on the module.

A common interaction I had with students occurred during the third activity as they sketched their prediction of the position and velocity graphs of the driver's motion. After having struggled with the graphs of the airbag's motion in the previous activity, many students drew very good driver graphs that neglected only to represent the correct direction of motion (opposite that of the airbag). When these students observed the computer-generated graphs (which were inverted versions of their own graph), the difference between the two graphs not only highlighted how both position and velocity graphs represent the direction of motion, but also reinforced the idea that the value of the velocity graph is the slope of the position graph. Students noted these ideas in their resulting explanations.

Experiment. I later devote an entire section of this chapter to how experimentation contributed to students' learning. Here I make a couple of general observations about how students benefit from the experimentation activity. Students' responses to assessments showed that experimentation helped challenge students' initial views of the car and airbag safety. For instance, before conducting experiments many students held the belief that drivers would be better protected in a car that failed to crumple than in a car that exhibited a crumpling zone. Responses to embedded notes show that many students were able to revise this view based on the results of their experiments, articulating how the crumpling of the car could affect the motion of the driver. Some students in the Study 3 *compare* condition (such as Joann and Linda) initially predicted on the basis of everyday experience that the vehicle speed would affect the driver's risk for injury from an airbag. The experiments helped some of these students recognize the important role of the other factors relative to the vehicle speed, and to distinguish the nature of the injury risk presented by inflating airbags from other collision dangers.

The ability of the experimentation activity to help students revise their everyday views illustrates the importance of providing students with opportunities to test their own ideas and evaluate evidence that they generate with experiments. Combining the *construct an argument* pattern with experimentation required students to sort through the evidence in order to support a particular point of view, increasing the effectiveness of the experiments themselves.

Unlike the previous two patterns, which appeared to be effective for most students, the benefits of experimentation appear to depend significantly on the quality of students' experimentation knowledge and choices. This observation speaks to the importance of guidance for experimentation, particularly in a complex context such as *Airbags*, where a single predetermined strategy does not provide students with complete insight about the situation. I discuss the design of the guidance for *Airbags* in more detail later in this chapter.

Construct an argument. As I mentioned above, the opportunity to construct arguments helped students solidify what they learned from their experiments. I observed students reaching insights from discussions with their working partner in efforts to reach consensus. These discussions were likely enhanced at least in part by the design of the prompts. Rather than simply asking students to explain an observation or phenomenon, the prompts provided students with just two or three choices and asked them to defend just one of the views. For instance, some prompts asked students to explain why a graph illustrated a safe or an unsafe outcome. Other prompts asked students to choose the collision factor that was most responsible for the outcome. Requiring each dyad to take a particular position helped students recognize when they were disagreement with their partner and required them to achieve consensus. Students often asked me to resolve disagreements between group members as they attempted to generate a unified response to these prompts. In these situations I would ask each student to summarize their own point of view and engage them in a mini-debate. Sometimes I would instruct students to revisit evidence such as the videos or their experimental results to better support their views.

It is worth nothing that all the above patterns prompted students to distinguish ideas, a process often neglected by traditional science instruction. *Orient and elicit* made students consider their everyday experiences about cars and motion, then reconcile them with accident statistics or ideas from their classmates. *Predict-observe-explain-compare* made students distinguish different characteristics of motion and how these characteristics are represented graphically. *Experimentation* requires students to distinguish among the outcomes of multiple trials and the collision factors that lead to these different outcomes. *Constructing arguments* require students to distinguish among multiple situations or collision factors in explaining a scenario. In this way *Airbags* demonstrates the process of making distinctions as critical to helping students integrate ideas.

Year to year iterative refinements contributed to success of design patterns

Airbags benefited from many iterative refinements since its initial design. I altered the motion profile of the airbag to provide more subtle differences between the motion of the airbag and driver (other than the direction of motion). These differences helped students distinguish motion with constant velocity from motion with constant acceleration. I also revised the explanation prompts to highlight these distinctions. This revision was successful in challenging more sophisticated learners.

I added a planning step to enhance students' experimentation. Early versions of *Airbags* did not prompt students to consider the design of their experiments until they began the experimentation activity. As a result, I observed students to be less purposeful in the way they conducted their experiments and how they connected their experimental design to the investigation questions. Along with revisions to the experimentation visualization itself, the planning step encouraged students to try to connect the investigation and the variables in advance. The results of Study 2 suggest that the nature of students' planning could have played a role in what they learned from their experiments.

Early versions of *Airbags* urged students to consider the general role of models and simulation in science. Students viewed examples of professional simulations of car crashes and reflected on their benefits and limitations for illustrating scientific phenomena. However, these steps seemed to interfere with students' understanding of the goals of the *Airbags* module. Students struggled to integrate the ideas from these steps with the overall investigation on airbag safety. Removing these steps made *Airbags* more streamlined and more clearly focused on the relationships among motion, graphs, and the variables governing airbag safety.

Previous versions of *Airbags* followed the experimentation activity with a short activity on the design of safer airbags. Though students appeared to find this activity interesting, like the examples of professional simulations, they did not prompt students to reconsider their ideas about motion and graphs to construct their explanations. I replaced the airbag design activity with the prompts requiring students to interpret and construct graphs to support their explanations. This revision required students to distinguish among the different characteristics of graphs and how the graphs represent changes to the experimentation variables in order to support their arguments, deepening their understanding of how graphs represent motion.

A design improvement to *Airbags* could help students give careful thought to their graph predictions. During early implementations of *Airbags*, students who I left to navigate through the module at their own pace would spend very little time constructing a graph that captured subtle aspects of the airbag's and driver's motion. Students would proceed quickly to the computer generated graph without having considered the nature of the observed motion very carefully. Though WISE lacks features that prevent students from advancing to a subsequent step without approval from an instructor, I was successful in simply asking students not to proceed past the prediction step until they had discussed their graph with either me or their teacher. Most of the time we asked students to clearly articulate the reasons for their predictions and to revise their graphs at least once before advancing to the next step. Improvements to WISE will soon be able to lock subsequent steps until students have appropriate discussions with an instructor. Alternatively, the Concord Consortium (who develops much of the visualization software for use in WISE) has developed software tools that can analyze student-generated graphs. The software could prevent students from proceeding past the prediction step until their graphs meet specific criteria.

Visualization scaffolding

A second important design consideration of *Airbags* is the set of four visualizations that scaffold the experimentation environment (Figure 2). These visualizations introduce the motion of the airbag and driver one at a time in a way that is similar to model progression (Swaak, Van Joolingen, & de Jong, 1998). Because of the complexity of the experimentation visualization, these four visualizations constitute a fading scaffold in that they provide a decreasing amount of support for understanding the sequence of collision events.

Embedded within the *predict-observe-explain-compare* pattern, these activities were effective in helping students connect graphs and observed motion and making distinctions between different types of motion. However, many students struggled to apply their understanding of the individual motion profiles of the airbag and driver to interpreting the collision outcomes. Christine and David (Case 3 of Study 3 in this dissertation) demonstrated facility interpreting both motion graphs kinematically, but did not always correctly apply the criteria for determining the safety of the collision. An additional activity that combined the two motion profiles and prompted students to interpret the collision might have helped students like Christine and David map the experimentation visualization on to the collision events.

Other scaffolding methods are also possible. In a design framework for scaffolding complex tasks, Merriënboer, Kirschner, & Kester (2003) distinguish between part-task and whole-task scaffolding methods. They suggest that though part-task methods can effectively prevent cognitive overload, learners can struggle to integrate these part-tasks. Whole-task approaches on the other hand “attend to the coordination and integration of constituent skills from the beginning” (p. 6). This distinction between part-task and whole-task scaffolding may help to explain the Christine’s and David’s proficiency with kinematical graph interpretation and difficulty with interpreting the collision events. The scaffolding approach for the *Airbags* visualization is whole-task with regard to graph interpretation, because the complexity of the motion the visualization presents remains the same throughout the activity. However, the scaffolding is part-task with regard to the airbag, driver, and their physical interaction. This approached worked well for most students in achieving proficiency with generating and interpreting motion graphs, but not as well for interpreting collision outcomes.

An alternative approach would be to present the airbag and driver together from the beginning and scaffold another aspect of the system’s complexity, such as the contribution of the individual variables or the types of outcomes that resulted. This approach might help students better integrate their understanding of the airbag’s and driver’s motion, though it might also obscure other important aspects of the situation. Comparing scaffolding approaches could be an area for future study with *Airbags*.

Role of the instructors and professional development in supporting the patterns

Most of the teachers engaged in at least one summer professional development session on facilitating technology-based inquiry learning. These sessions helped teachers monitor students' learning, use assessment tools, provide useful feedback to students, guide students interactions with visualizations, and facilitate timely whole class discussions. As the module designer as well as an experienced teacher, I also modeled the process for teachers, gradually turning leadership over to the teachers either toward the end of the module or for the following year's implementation. Conversations with student groups and whole class discussions were particularly important to students' success with *Airbags*.

Conversations with individual student groups occurred when students asked for assistance or when either the teacher or I recognized students needed a nudge in the right direction. We asked students probing questions to elicit relevant ideas. We would occasionally add appropriate ideas, particularly for students who did not read well, such as English language learners. We could highlight important distinctions by helping students make relevant comparisons. We could prompt students to construct arguments by playing "devil's advocate" to their points of view, helping students seek supporting evidence. Teachers and I periodically led whole class discussions which support the knowledge integration processes in many of the same ways. Class discussions elicit key ideas from students and serve as a source of new ideas from one student to another. They can also promote argumentation and debate between students where when different class members hold different views. In a sense, conversations and discussions constitute a form of the *collaborate* design pattern (Linn & Eylon, 2006).

Individual conversations and class discussions are particularly important for students in WISE projects (and likely for computer-based curricula generally) because they are adaptive to students' immediate needs in ways that software generally is not. Good instructors can assess students' current level of understanding, respond to the needs of individuals or a whole class, elicit or add the right ideas at the right time, focus students on appropriate distinctions, or suggest they revisit certain concepts. In this way, professional development for WISE aims to help teachers complement WISE to optimize instruction and discourages the view that WISE acts as a replacement for teachers.

Visualization design

In addition to the design of the module as a whole, design features of the visualization also contributed to students' learning. As with the overall module design, there are no comparison data that illustrate the effectiveness of the visualization's design features, though the case studies provide some helpful illustrations. As with many design efforts, my continuous evaluation of the design occurred through user feedback, observations in authentic contexts, and discussions with teachers, students, researchers, and developers. In particular, my interactions with students illustrated what features were helpful for students' learning and what students struggled with.

In general, the diversity of students in my studies presents challenges for visualization design. Many studies find that learning outcomes from dynamic visualizations depend on learners’ prior knowledge and spatial ability (e.g. ChanLin, 2001; Hays, 1996; Yang, et al., 2003). How can the design of visualization tools accommodate a wide range of learner abilities? I designed the visualization with several design principles in mind that have applicability to a wide range of learners. Here I discuss the success of four design principles in fostering student learning in *Airbags*: (1) minimize extraneous complexity to highlight learning goals, (2) provide coordinated, multiple representations, (3) provide record keeping tools, and (4) provide opportunities for learner-initiated exploration and interaction. I will also point to refinements that improved the visualization’s use of each of these principles.

Design principles for dynamic visualizations

Minimize extraneous complexity in order to highlight learning goals

Unnecessary complexity in visualizations can distract students from attending to key concepts and focus on peripheral details instead (Ainsworth, 2006). Visualization details that are unrelated to the learning goals can also impart extraneous cognitive load, interfering with learning (Chandler, 2004). In designing the *Airbags* visualization, I aimed to focus on details that were germane to the learning goals concerning motion, graphs, and the investigation questions.

At first glance the visualization appears complex with its three distinct areas that operate simultaneously. My classroom observations, however, suggest that visual complexity was not an obstacle to most students’ understanding. Some students even suggested that I introduce more powerful data management tools, which would have made the visualization even more complex. Students’ ability to use the visualization without being cognitively overwhelmed demonstrates that the important features were adequately highlighted and that the visualization as a whole was scaffolded effectively by the early activities. Whatever complexity the multiple representations introduced appear to have been appropriate to the learning goals and not extraneous.

The collision animation design is minimal. Simple icons represent the appearance of the driver’s head and the airbag’s surface. The animation selects the most relevant time period of the collision in order to focus students on the key events that inform the investigation. I de-emphasized the distance scales, both on the graph axes and in the animation. The visualization was successful in emphasizing conceptual and qualitative rather than quantitative comparisons.

My observations revealed that some students struggled with distance scales and might have benefited from a more faithful visual depiction of the collision. For instance, the abstractness of the animation did not allow students to judge from the visualization what constitutes being “close” to the airbag. A representation of the driver and airbag that was more to-scale might have helped students leverage their everyday ideas more productively.

Students also struggled to map the elapsed time in the simulation to real life. To simplify students’ interaction with the tool, I did not provide students with control over the viewing speed of the animation. As a result, even the most sophisticated students made time errors by more than an order of magnitude. Eric (from Study 3, Case 1) referred to time intervals on the order of eight to 10 seconds, even though the entire collision event represented in the animation lasted for about one-tenth of a second. Joann (from Case 2) made a similar time-scale error. These errors probably occurred from students seeing the collision only in slow-motion and never in real-time. These difficulties suggest that an additional level of control allowing students to view the simulation at two different speeds (slow and real-time), combined with videos in real time, could benefit students’ understanding of the time scale of the events in question.

Provide coordinated, multiple representations to facilitate understanding

The *Airbags* visualization offers students two types of motion representations, an animation and a graph. Students' abilities to connect these two representations were mixed. In Study 3, Joann and Linda went seamlessly between the two representations during the course of their discussions. They gestured to the animation to illustrate spatial relationships in the car (such as position of the dummy relative to the steering wheel) and to the graphs to compare the outcomes of multiple trials. For Joann and Linda the two representations served mainly complementary roles (Ainsworth, 1999). Christine and David on the other hand relied primarily on the animation. They interpreted the graphs kinematically but failed to connect these interpretations to their understanding of the collision dynamics. Though the two representations were synchronized, Christine's and David's struggles suggest that synchronization alone was insufficient for some students to make the necessary connections. These students could benefit from guidance or an additional activity that would connect representational features more explicitly. For example, students could annotate the graphs and/or the animation to highlight how the two representations illustrate key collision events, or they could capture snapshots of the visualization to support a narrative of events. The newest version of WISE will support features such as annotations and snapshots.

The crash test video provided a third representation of the collision, though links between the visualization and the video were difficult because the video could not be changed to represent different experimental trials with different motion parameters. Because the video could display only one collision scenario, students could not link their investigation of the variables to the video. Kozma (2003) reports success linking videos to symbolic, graphical, and molecular representations in the learning environment 4M:Chem. However, these representations were not manipulable by users as the *Airbags* experimentation environment is. How to help learners link manipulable representations to real-life representations such as videos is a topic for further research.

The experimentation environment also allows students to toggle between two types of graphical representations, position and velocity graphs. An early version of *Airbags* provided students only with the ability to view position-time graphs. I later added the ability to view velocity-time graphs in response to students' difficulties with the position graphs. Even though the addition of a new type of graph added an additional mode of interaction to the visualization, this addition helped students who found interpreting velocity graphs easier than position graphs. These students could then use the velocity graph to constrain their interpretation of the position graphs (Ainsworth, 1999). Having two interchangeable types of graphs gave students more points of entry into understanding the visualization, meeting the needs of a more diverse group of learners.

Provide record keeping tools

Addition of the experimentation history was the most significant revision I made to the *Airbags* visualization. In initial versions of *Airbags,* students documented their own experimentation choices and outcomes in a text-based journal. Students' completion of the journal was perfunctory and did little to augment their understanding of the experiments. When WISE technology supported logging of students' experimentation choices, I added the experimentation history to the visualization. The history automated recording for the variable values for each trial, allowing students to devote their main effort to designing and interpreting their experiments. Students often used the checkboxes that indicate whether the trial was safe, helping them review previous trials and see patterns in the outcomes. I also observed the vast majority of students to use the graph comparison feature, usually to compare multiple trials that investigate the same variable. The comparison feature helped students distinguish safe from unsafe outcomes and the effects of one variable from another on the characteristics of the graph.

The data management feature was so useful to some students that many requested enhancements to its design. Most of these suggestions concerned greater control over organizing the data in the trial history. Students wanted to be able to sort trials according to other criteria than the order in which they were conducted. Being able to sort the trials using safe/unsafe outcomes, the investigation questions, or values of a specific variable could indeed have helped students use the data to identify patterns or even threshold values.

I added a feature that allowed students to "record findings". This feature opened a window where students could complete fields in a table that included any experimental findings, as well as the trials that led to those findings. This tool was too cumbersome to be useful—it needed to be more tightly linked to the trial history. Students' findings were better captured using reflection notes after the experimentation activity. Students suggested that I add a field to the history that allowed them a short annotation for each trial to remind them of what they learned from that trial. Unfortunately, there was insufficient screen real-estate to accommodate annotations without a drastic change to the layout of the entire visualization. This enhancement could allow students to reflect on each trial they conducted without disrupting the flow of their overall investigation. Future research with WISE modules featuring experimentation will examine the benefits these organization tools have on students' interpretations of the data.

Provide opportunities for learner-initiated exploration and interaction

Experimentation with variables necessarily allows students to initiate their own exploration. In *Airbags*, students choose an investigation question, change the variable values, observe the outcome, and conduct comparisons. This process allowed students to test their own hypotheses, explore the outcomes of extreme variable values, and examine the effects of minute changes to the variables. I will discuss the specific role of experimentation in students' learning in more detail in the next section.

The *Airbags* visualizations offer students another way to control the information. Research suggests benefits for providing learners with control over the playback of dynamic material can improve learning outcomes by reducing the need to keep information in memory (Moreno & Mayer, 2007; Schwan & Riempp, 2004). In accordance with this principle, I provided students with controls that could play and pause the animations. This feature not only allowed students to effectively slow down the animation to a speed they could manage, but it was also useful for incrementally advancing the animation to observe accelerated motion. Advancing the animation in a step-wise fashion illustrated the fundamental nature of acceleration as students observed the different distances covered by the object during successive, equal time increments. Though students usually required prompting to take advantage of the play/pause control in this way, it greatly enhanced students' understanding of the nature of acceleration and it's representation on a position-time graph. Students would likely benefit from a "step forward" button that would advance the simulation by a prescribed time interval.

Visualizations should support the knowledge integration design patterns

There is a growing literature on characterizing the effectiveness of dynamic visualizations for learning (Chandler, 2004; Hegarty, 2004), including two recent meta-analyses (Chang, et al., in preparation; Höffler & Leutner, 2007). Research studies examine design features of visualizations such as visual characteristics, instructional domain, level of abstraction, duration, instructional setting, level of interactivity, and many others. Characterizing visualizations according to these features fails to capture important details about the instructional conditions in which they are used. In many situations, such as *Airbags*, these instructional conditions are what determine the effectiveness of the visualization.

The benefits of the *Airbags* experimentation visualization goes beyond the four design principles I discuss above. Though the airbags visualization meets criteria for dynamic visualization, it is really an experimentation environment that, along with the rest of the *Airbags* module, engages students in the knowledge integration processes. I co-designed the visualization with the module as a whole to support the knowledge integration patterns and help students develop coherent knowledge. The module along with the visualization thus function as a single instructional unit and their individual contributions toward students' learning outcomes are not entirely separable. The *Airbags* visualization supports students as they test hypotheses, conduct experiments, make comparisons, and examine the data. The module subsequently guides students in reporting findings and using the data to construct arguments, activities that are central to scientific inquiry. Specific design features make the visualization easier to use, but the visualization promotes deep learning about science because it engages students in scientific inquiry by supporting established curriculum design patterns.

The approach I have used co-designing visualizations and the curricula that support them is an area that is ripe for further research. The complexity of learning that is possible with the *Airbags* visualization points to the need to develop a new set of criteria for evaluating the design of visualizations. Traditional design principles that govern specific features of visualizations are important, but design principles that are similar to the ones used to evaluate all curricular designs are needed to evaluate the potential for visualizations to support complex science learning. Visualizations should be evaluated as a subset of science curriculum materials that must also build on students' everyday ideas and prior conceptions of science, add new ideas in appropriate ways, help student distinguish ideas by promoting comparisons, and support reflection and revision of new knowledge.

Supporting informative experimentation

Varma (2010) conducted a study that examined support for middle school students conducting controlled experiments using a visualization. They found that instructing students to vary one variable at a time improved the quality of their inferences about environmental factors that lead to climate change. Studies with *Airbags*, however, suggest that a focus on controlling variables can interfere with learning.

Varma's study on the climate change visualization differs from studies on *Airbags* in two important ways. First, research on children's experimentation (Klahr & Nigam, 2004; D. Kuhn & Dean, 2005) show that younger students have difficulty applying CVS to classroom experimentation situations without prompting. Students who studied *Airbags* were high school students enrolled in physics and had had two or three previous years of high school science instruction. Many students in my studies on *Airbags* were thus able to employ CVS without any prompting. One explanation for the discrepant results is the superior general science knowledge of high school students compared to middle school students. This knowledge could include facility with CVS.

Second, learning outcomes in the climate change module, like others that examine students' propensity to isolate variables, involve inferences that can be made entirely on the basis of controlled tests. My studies with *Airbags* on the other hand measure students' ability to understand motion graphs in the airbags context as well as the role of threshold values, ideas that are not necessarily revealed by isolating variables. This difference illustrates the distinction I make in this dissertation between controlled and informative experimentation. In situations where the learning goals consist of describing covariation between variables and outcomes, a predetermined method of isolating variables can be sufficiently informative to achieve the learning goals. In these cases inferences can be based solely on observed outcomes between controlled trials and do not necessarily require an understanding of the experimentation context. My studies on *Airbags* show that prescribed methods of controlling variables may be insufficient for uncovering the nuances of the *Airbags* situation. In order to achieve a complete understanding, students need to leverage a wide range of ideas. Students must be able to connect the variables and the graphs in the *Airbags* visualization to events in the collision. Students should be aware of the types of relationships variables can have to outcomes in addition to covariation, such as invariance, variable interactions, and thresholds. Students must also possess a range of strategies (other than CVS) to investigate these myriad relationships.

Many students may have sophisticated ideas about experimentation in their repertoire, but they may be suppressed by expectations they have for experimentation activities or by their classroom culture. As a result, these students may not be able to fully interpret their experimental findings or make distinctions necessary for a nuanced understanding of the situation. For instance, Brett and Eric were clearly aware of complexity in the *Airbags* activity that they chose not to investigate, possibly because of their beliefs about what constitutes a valid experimentation approach. Although they clearly had knowledge of strategies other than CVS (which they used during their informal exploration of the variables) they used exclusively CVS to investigate the questions, even when other strategies would have been more informative. David and Christine did not look to explain their findings using thresholds, even though examples of thresholds are prevalent in both everyday life and in their physics curriculum (e.g. the amount of force required to overcome static friction between an object and a surface). Both of these groups could have benefited from instruction that brought these ideas to the forefront of their repertoires.

Instruction should guide students through the knowledge integration processes to help students benefit from experimentation. Guidance should elicit ideas about variable relationships and experimentation strategies (or add these ideas if students lack them altogether) prompt students to make key distinctions, then help students refine their understanding of the investigation. I explore the implications for experimentation guidance in the next section.

Designing guidance to make experimentation more informative

The case studies illustrate how students' propensity to seek evidence only for covariation can have an impact on students' experimentation strategies. To encourage students to look beyond covariation to interpret the outcomes of their experiments, instruction could elicit other types of relationships from students. Klahr and Dunbar (1988) found that prompting learners to generate multiple hypotheses led them to conduct trials that could discriminate between their hypotheses. In *Airbags*, eliciting hypotheses about interactions or thresholds could encourage students to explore the variable space in more detail than a few pairs of controlled tests. Eliciting a broad range of strategies in connection with these hypotheses could help students prioritize conducting informative trials over adhering to a prescribed pattern of CVS. Though the planning step of *Airbags* did elicit strategies students intended to use, it did not explicitly ask students to articulate a range of possible strategies.

Guidance during the course of experimentation could also keep students mindful of the purposes of their trials. Kuhn and Phelps (1982) found that students who conducted trials with a purpose in mind were more successful on an experimentation task. In *Airbags*, students choose one of their three investigation questions before conducting each trial. Though this action appears to remind students about the overall investigation, more detailed prompts could engage students more deeply. Prompting students to articulate what they expect will happen in their upcoming trial could challenge them to consider the purposes of their trials more carefully and to reflect on the implications of unexpected results. Too much prompting during experimentation could be counterproductive and interrupt the flow of their investigation, however. Further research should examine the trade-offs between prompting and continuity during experimentation.

Airbags shows that the right experimentation goals can prompt students to explore subtle variable effects, extending research by Dunbar (1993). The *compare* prompts were successful with some students in part because they required students to look beyond relationships between individual variables and outcomes and distinguish between the effects of the variables. More specific prompts might have been more successful with less sophisticated learners, however. Lower prior knowledge learners might benefit from having to identify any “special” values in the airbag’s motion profile and use their experiments to explain the role of these values in determining collision outcomes. These prompts could direct students’ attention to critical values and highlight the piecewise nature of the relationships between the variables and outcomes. Students would also be compelled to distinguish the relationships that exist above and below these critical values.

Airbags illustrates the challenges in designing experimentation guidance for individuals. The wide range of knowledge and beliefs students have about experimentation makes anticipating the type of guidance individual students need a difficult task. Furthermore, students’ experimentation choices alone are insufficient to determine students’ intentions. I discuss the value of the experimentation reports in more detail in the next section on assessments.

Emphasis on CVS in traditional science curricula

The propensity of some students to use exclusively CVS, even when it is not especially informative, raises issues about the role of CVS in traditional science instruction. Brett and Eric demonstrated their belief that CVS is the only valid strategy for conducting “experiments.” They prioritized using CVS over gaining further insight. Christine and David appeared to employ CVS because they lacked knowledge of other approaches that would test their ideas. Both groups employed CVS according to predetermined patterns and with almost no variation. Both groups neglected to alter their approach in response to unexpected results or new findings. Furthermore, their approaches suggest they don’t distinguish between different ways to control variables.

Though I cannot say where Brett, Eric, Christine, and David acquired their knowledge and beliefs about CVS, the emphasis of traditional instruction on the validity of CVS likely contributed to their views.

Joann and Linda illustrate that not all controlled tests are equally informative. They learned little from their initial controlled tests by using a predetermined strategy (high-middle-low). After spontaneous exploration, however, their controlled tests were much more informative because they found a range of values was consequential to the outcome. Joann and Linda appeared to benefit from being presented with investigation questions that encouraged them to explore alternative strategies. Their initial set of controlled trials suggests that using CVS was their default view, and the compare questions combined with a productive orientation toward science provided a sufficient impetus for abandoning the traditional approach. This type of guidance alone was not sufficient for all students to abandon the default approach. Further research on helping students prioritize learning over procedure in experimentation activities is needed, particularly in classroom contexts.

These nuances of experimentation, such as identifying consequential values and the multiple ways one can control variables, are unlikely to be addressed in traditional science instruction, as they are specific to individual experimentation contexts. They require specific knowledge about the investigation context. I believe however that the idea that it is permissible to explore variables in an unstructured way in order to further one’s understanding of the situation is accessible to typical science students. An approach to teaching CVS that presents it as one of many valid strategies (rather than a panacea) and that includes discussing its limitations in addition to its usefulness would help students learn better from situations like *Airbags*. I address other general implications for classroom science instruction in the Conclusions chapter.

Assessment design

Pretests/posttest and embedded assessments

Airbags used both pretest/posttest and embedded assessments to capture students’ understanding of motion and graphs in different ways. Pretests captured students’ initial level of understanding, embedded assessments captured students’ understanding within the context of the *Airbags* investigation, and posttests assessed whether students could generalize this understanding to other motion contexts. Domain understanding both within and outside of the investigation context are important. Because this study aimed to examine the benefits of subtle forms of experimentation guidance on students’ understanding of the *Airbags* situation, the embedded assessments were necessary to examine effects of the guidance on students’ understanding of the experimentation activity. Posttest assessments were necessary to examine the overall effect of the design of *Airbags* on students’ understanding. Though as I expected, the posttest scores were insensitive to the different forms of guidance students received, the posttests illustrated that students’ knowledge gains were not limited to the investigation context.

Both kinds of assessments included two types of generation items that were important to measuring students’ understanding of motion graphs. Explanation items captured how students link the characteristics of graphs to motion in ways the multiple choice items cannot. Graph construction items provided students with weak writing skills to generate these ideas in a non-verbal way.

Both sets of assessments benefited from iterative refinement. Pretest/posttest items were revised not only to correspond with changes in the module design, but also to better target specific ideas such as distinctions between position, velocity and acceleration or motion in opposite directions. Early versions of embedded prompts designed to assess students' interpretation of the visualization graphs measured only whether students could distinguish safe from unsafe outcomes. I augmented these prompts to measure students' ability to distinguish between the position, velocity, and time variables and how students attributed the outcome to one of these variables.

I added the paper and pencil graphing activity after the experimentation activity to provide students a way to use graphs to support an argument. These items also focused on making distinctions between the variables and the connections between the individual variables and the outcomes. I used the paper and pencil format for graph sketching because the WISE drawing tool was cumbersome for students who wished to generate neat, precise graphs. WISE graphing tools currently in development will facilitate graph sketching, not only obviating the need for paper and pencil alternatives, but also providing tools that can automatically analyze the characteristics of students' graphs and provide appropriate feedback.

Experimentation reports

The reports of students' experimentation provided detailed information on each group's experimentation choices. Many aspects of students' experimentation choices were readily visible from these reports, such as students' propensity to conduct controlled trials, to test boundary values, to repeat trials, and to investigate each inquiry question. To a certain extent, the number of trials students' conducted reflected students' level of engagement with the activity. Furthermore, because at least a few controlled trials were necessary for most students to understand the situation, the presence of controlled trials within students' experimentation sequences reflects some degree of systematic inquiry. In these ways, reports allowed students to be generally classified at a glance as unengaged or engaged, haphazard or systematic. The immediate availability of this kind of information can be useful at a broad level to characterize how students approach their experimentation.

The experimentation reports were helpful in illustrating students' intentions to control variables. The CVS score I used for Study 1 and Study 2 could identify students who consistently controlled variables across all the investigation questions, while the CVS proportion I used for Study 3 illustrated how heavily students relied on CVS in their investigations. This measure was particularly illustrative for students whose CVS proportion approached 100%, revealing the extent to which high prior knowledge students chose to isolate variables at the expense of less systematic approaches.

Clearly the experimentation reports have limitations. Most notably, the reports lack important information on students' rationales for their variable choices. In many (if not most) cases, the reasons behind students' variable choices, and not the values themselves, determine how informative a strategy is. For instance, analysis of their conversations reveal that Brett and Eric tested boundary values of all the variables in an effort to observe contrasting outcomes, while Joann and Linda tested boundary values in order to examine tradeoffs and variable interactions. Furthermore, students can isolate variables for different reasons and in different ways. Brett and Eric conducted controlled tests without aiming to further their understanding, while Joann and Linda used controlled tests to approach critical values they identified during their previous trials. Only a detailed analysis of students' discussions could have revealed these distinctions between similar strategies.

Despite the limitations of the experimentation reports, my close analysis of the three cases suggests ways to refine algorithms to analyze students' experimentation choices. Algorithms could flag strategies that are unlikely to inform the investigation, such as neglecting to test variable values on both sides of threshold or achieving identical outcomes for many consecutive trials. Algorithms could also identify when students adhere very strictly to common patterns of implementing CVS without variation or if students fail to generate sufficient evidence to explore variable interactions. Future research will examine ways of characterizing students' inquiry moves more precisely using experimentation reports.

Experimentation environments could also prompt students for more information about their intentions or insights as they conduct individual trials. As I discussed earlier, my initial efforts to incorporate a way for students to record findings during the course of experimentation were unsuccessful. Less intrusive ways of soliciting students' ideas about individual trials could still be useful, however. Even yes-or-no responses to prompts such as "Did this trial help you answer the question?" or "Was this what you expected to see?" could provide researchers with some additional information on students' intentions.

Upcoming features in WISE will enable a wide variety of ways to capture students' ideas. For instance, students will be able to take snapshots of visualizations, annotate the snapshots, add them to an online lab notebook, and use the notebook entries to support their arguments. These entries would provide researchers and teachers with information about students' choices and insights during the course of experimentation. Furthermore, the journal would provide a compelling way for students monitor their own inquiry progress that is similar to the practice of authentic real-life science.

School settings

Airbags was consistently effective within the course of the regular classroom physics curriculum at improving students' understanding of motion graphs. Teachers used *Airbags* either near the end of their unit on kinematics as a capstone unit or late in the year as a way to review mechanics concepts. My studies on *Airbags* demonstrate the extent to which design principles for inquiry instruction, dynamic visualizations, and experimentation guidance translate from laboratory to classroom settings. *Airbags* was also effective in diverse classroom contexts and with students exhibiting wide ranging mathematics ability. One of the most important advantages of inquiry instruction is its potential to help students across learning contexts and prior knowledge levels. I attribute the success of *Airbags* to the compelling investigation context combined with the knowledge integration patterns.

Study 3 in particular illustrates the importance of conducting science education research within classroom contexts. Brett and Eric's beliefs about classroom science inquiry appear to strongly influence their experimentation approach. Brett and Eric were enrolled in a science and mathematics program that had high standards for admission and that served the strongest students in their metropolitan area. Their approach to the *Airbags* task was naturally affected by factors other than their knowledge about experimentation or the investigation context. Their discussions reveal how their roles as high achieving, college-bound science students imposed goals that were unrelated to the *Airbags* task. It is quite possible that Brett and Eric would have used alternate strategies if they had been in a learning setting (such as a laboratory) where scholastic achievement did not impose these other goals. My findings point to the need for instruction to consider alternative goals students may have when they engage in classroom inquiry. Further research should examine how instruction can help students engage in authentic science inquiry in the face of goals that may interfere with deep learning. I continue to explore the implications of science classroom culture and traditional science instruction on the design of learning environments in the concluding chapter.

Conducting classroom studies with *Airbags* presented several challenges. First, working with teachers and students in their regular classroom settings obliges researchers to aim to optimize learning conditions for all students, even at the expense of collecting precise data on student learning. Years of classroom studies with WISE indicate students are most successful when they work in dyads because they benefit from discussions with their partner. One consequence of having students work in dyads is that all responses to embedded prompts, as well as students' experimentation choices, reflect the consensus of both students. In actuality, students' decisions and insights may more accurately be attributed to one student or the other. Using the student dyad as the unit of analysis thus interferes with the researcher's ability to connect students' inquiry moves to learning.

Second, conducting studies in high school physics classes led to a relatively small total sample size for each study. Nearly all classroom implementations of *Airbags* involved the school's total physics enrollment, as enrollments at some schools were sometimes less than 10 students. Small class sizes also made class effects difficult to account for. As previously discussed, taken together the diverse contexts in the *Airbags* studies approximate the total population of US students who study high school physics. The findings thus illustrate the effectiveness of *Airbags* across a wide distribution of learners and classroom contexts.

Third, teachers were not always inclined to use strong inquiry-based teaching strategies while teaching with *Airbags*. Teachers who taught with *Airbags* participated in professional development, which promotes best-practices for teaching with technology-supported inquiry. Even with professional development, some teachers have difficulty making the transition from their traditional techniques to these best-practices in their own classrooms. I addressed this issue by guiding teachers with my own example, then gradually turning responsibility of facilitating student discussions over to teachers. Most teachers who taught with *Airbags* for multiple years showed large improvements in their inquiry-based teaching practices in their second year and beyond.

Finally, difficulties with unique aspects of school technology resources are unavoidable in classroom research. Security firewalls, poor internet connectivity, old computers, or outdated school software provided frequent obstacles to smooth implementations. These events sometimes resulted in delays, occasionally interfering with students' ability to complete the *Airbags* module. When possible, I brought TELS computers and equipment to classrooms in order to minimize these problems. Though these problems can be frustrating for researchers, they are also highly educative, as they represent real-life obstacles schools face when adopting new technologies into their curricula. Sometimes these issues forced me to make subtle changes (such as reducing the length of a video clip to minimize download time) to improve students' experience with *Airbags*. These design decisions contribute to practical design knowledge that can lead to better classroom implementations in the future.

Chapter 9: Conclusion

Summary

Overall impact

This dissertation explores the impact of *Airbags*, a computer-based inquiry curriculum module, on high school physics students' scientific understanding. *Airbags* guides students through an investigation about the personally and socially relevant context of airbag safety. Students interact with dynamic visualizations that allow them to use their knowledge of physics to examine the dynamics of airbags deployment in detail. Students use a visualization to conduct experiments that provide insight about the safety of airbags. In three studies spanning three years, students consistently made moderate to large pretest to posttest gains in their ability to interpret and construct graphs representing one-dimensional motion, with low prior knowledge learners generally making the greatest gains. Students also gained relevant insights about the safety of airbags in head-on collisions.

Design patterns

Airbags makes use of four empirically tested design patterns to help students integrate their scientific ideas about graphs, motion, and the airbags context. The patterns go beyond traditional instruction particularly in their capacity to help students distinguish key ideas. The *orient and elicit* pattern highlights distinctions between students' everyday ideas about the deployment and safety of airbags and new forms of evidence such as videos and fatality statistics. The *predict-observe-explain* pattern forces students to distinguish predictions from observations. Explanation prompts led students to distinguish the ways graphs represent different types of motion. The *experiment* pattern requires students to distinguish the effects of one variable from one another, or one value of a variable from another value, on experimental outcomes. *Constructing arguments* make students distinguish among two or more points of view and among the pieces of evidence that support those views.

Design of visualizations for science learning

Research identifies pitfalls students may encounter when learning from dynamic visualizations, but the design of the *Airbags* visualizations appears to overcome many of these pitfalls. Virtual experiments allow students to test their own ideas, building on everyday ideas about science and contributing to deep learning. Gradually introducing elements of the experimentation environment over the course of the early activities illustrated how graphs could represent the collision events, scaffolding the complexity of the experimentation environment. A record keeping tool helps students compare previous trials, keep track of outcomes, and monitor their progress. Most importantly, *Airbags* incorporates the visualizations within the framework of the design patterns. The patterns help students connect the new ideas from the visualizations to prior knowledge and prompt students to use evidence they generate from their experiments to support subsequent arguments. The success of the visualizations in *Airbags* suggests that researchers should use criteria for curriculum design in addition to visualization design to evaluate the quality of visualizations for science learning. These criteria should evaluate how well visualizations combine with surrounding instruction to engage students in authentic inquiry activities, promoting knowledge integration.

Guidance for experimentation

The main research focus of this dissertation is on how students learn from experimentation activities within inquiry investigations. My studies suggest that guidance for experimentation should focus students on distinguishing concepts, and that direct instruction of the control of variables strategy may not be effective in complex situations where learners must incorporate domain knowledge to gain insight. The results of Study 2 and Study 3 illustrate that students can benefit from experimentation strategies that do not isolate variables. Students may use alternative strategies that test extreme variable values or narrow the range of consequential values. Furthermore, not all controlled experiments are equally informative. Spontaneous exploration can help students test new questions that arise from unexpected results, informing the design of controlled tests that better reveal subtle characteristics of the variables.

Guidance that encourages students to compare rather than isolate variables may have important benefits. Comparing variables requires students to consider not only the existence of a variable's effect on the outcome, but also its magnitude relative to the other variables. Comparing variables also forces students to look for other ways to distinguish the variables, such as the presence of threshold values, leading to new insights.

Study 3 found the *compare* guidance to be effective only for high prior knowledge learners and to have no effect on lower prior knowledge learners. Making the nuanced interpretations necessary to distinguish the variables required sophisticated knowledge about physics, experimentation strategies, and the *Airbags* context. Future research should examine ways to guide less sophisticated learners toward these detailed insights.

Questions for future research

This dissertation raises new questions for further research on improving science instruction that incorporates experimentation activities. Joann and Linda illustrate how strategies other than CVS helped them gain insight about the variables in *Airbags*. Further research should systematically examine how alternative strategies to CVS can contribute to understanding in other realistic experimentation contexts. What strategies do learners use in these situations to refine their understanding? How can instruction help promote the use of these strategies?

Airbags also demonstrates ways that students may come to understand the role of threshold values in the outcomes of multivariable systems. Thresholds appear frequently in everyday science situations such as drug dosages, diodes, and static friction. Traditional science instruction may not explicitly introduce threshold-based relationships, and many students may lack sufficient knowledge about experimentation to systematically investigate their role in variable-outcome relationships. How do students come to understand threshold values, and how can instruction guide students toward understanding thresholds in a range of investigation contexts?

Data logs of students' experimentation choices provided detailed information about how students initiated their investigations. These data have the potential to be useful in many ways other than purely for research. A new NSF-funded research effort called Logging Opportunities for Online Projects in Science (LOOPS) examines ways teachers can use reports of the data to make teaching decisions. Data on students' inquiry moves can inform teachers' choices of assessments, curriculum customizations, and guidance for individual students and whole classes. Curriculum materials themselves may also take advantage of the data to provide immediate, automated guidance for students as they explore visualizations and conduct experiments. Well-designed algorithms may be able to identify uninformative patterns of experimentation and provide students with a nudge in the right direction. Research should examine whether these interventions can improve learning outcomes.

Students' experimentation with the *Airbags* visualization illustrates ways that their understanding of the investigation context, knowledge of physics and experimentation, and their representational competence with graphs intersect. There remains much more to be understood about how learners (experts and novices alike) draw on a diverse array of ideas to initiate investigations about relevant science topics. Much current research on learning from experimentation still largely neglects the role of domain knowledge in students' sense-making during experimentation. Further research must continue to examine how learners incorporate multiple sources of evidence to investigate complex science problems.

Finally, the Study 3 case studies suggest that social factors, in addition to scientific understanding, may contribute substantially to how students learn from inquiry-based curriculum materials. Brett and Eric provide an interesting case of students who can clearly articulate the complexity of the problem in front of them but appear consciously to ignore this complexity. The roles of classroom culture and the goals of traditional science instruction on students' inquiry moves and insights merit additional study. I reflect further on this issue in the next section.

"We can only change one of them, we can't change multiple ones."

This quote by Eric, though it explains much about his approach to experimentation, raises questions about how students' beliefs affect their inquiry decisions in a classroom context. The obvious question for Eric in response to his statement is of course, "*Why* can't you 'change multiple ones'?" Unfortunately there is insufficient information in the conversation between Brett and Eric to answer this question definitively, and I did not have the opportunity to interview them. Here I speculate on two issues that might have contributed to Eric's thinking.

What is an experiment?

Children hold many views of what constitutes a scientific experiment (Carey et al., 1989). From my own personal experience, I can attest that science professionals also carry divergent ideas about the nature of an experiment, some of which differ markedly from the definition implied by school science standards. The American Association for the Advancement of Science (1993) grades 6-8 standards for scientific inquiry include the following:

> "If more than one variable changes at the same time in an experiment, the outcome of the experiment may not be clearly attributable to any one variable. It may not always be possible to prevent outside variables from influencing an investigation (or even to identify all of the variables)." (p. 12)

The science content standards for California public schools (California Department of Education, 2000) include the following:

> For Grade 5: "Identify the dependent and controlled variables in an investigation" and "Identify a single independent variable in a scientific investigation and explain how this variable can be used to collect information to answer a question about the results of the experiment." (p. 17)
>
> For Grade 6: "Distinguish between variable and controlled parameters in a test." (p. 29)
>
> Grades 9 through 12: "Identify possible reasons for inconsistent results, such as sources of error or uncontrolled conditions" and "Recognize the issues of statistical variability and the need for controlled tests." (p. 52)

These standards emphasize views of experimentation (or scientific investigation generally) similar to those held by practitioners of clinical medicine and appear to include controlled tests as a fundamental criterion for an experiment.

Other professional views of what an experiment is differ from this view. For instance, my wife conducts research in the biochemistry department of a major research university. Some days she calls me to tell me she must stay late to "finish an experiment." In response to my query about what she means precisely by "experiment," she described the process of preparing a sample of cells for observation with a microscope and capturing digital images of the sample. Though she refers to this procedure as an experiment, her current research project has yet to use a controlled comparison design. My wife's conception of an experiment is consistent with my own previous research in the field of materials science. In materials research, one does either "experiments" or "theory." Experiments involve tests of real materials using machines, while theory involves modeling material behavior using equations or a computer simulation. Though materials researchers do conduct controlled comparisons, they are not a necessary part of the colloquial use of the term "experiment."

Ideally, science instruction would address the appropriateness of different methods for different situations rather than endorse one specific view of experimentation. Eric's statement suggests that he holds a narrow idea of what constitutes an experiment, possibly instilled by the emphasis on controlled tests during years of classroom science. With his experimentation strategy, Eric followed the "letter of the law" as he understood it, but neglected to uphold its "spirit," which prioritizes gaining insight over following procedures. (Joann and Linda, on the other hand, appeared to have their priorities in order.)

The contrived nature of Brett's and Eric's investigation approach suggests that *Airbags* fell short for these two students in achieving one of the overarching goals of all WISE modules: to bring a measure of authentic science to the classroom. This goal includes engaging students in an effort to deepen their understanding of a relevant socio-scientific issue. Eric's statement about experimentation suggests that the goal of deepening one's knowledge may conflict with other goals he may have as he practices classroom science. This brings me to the second issue.

Motives and goals for classroom science

A wide research literature explores the sociocultural perspective on learning. I do not attempt to review it here; rather, I quote Strike and Posner (1992) to succinctly summarize the relevant issue: “A wider range of factors [than scientific conceptions] needs to be taken into account in attempting to describe a learner's conceptual ecology. Motives and goals and the institutional and social sources of them need to be considered.” (p.162) Eric’s statement raises questions about what motivated his experimentation choices and how they manifest his broad goals as a student in physics.

As they participated in Study 3, Brett and Eric attended a magnet program for gifted science and math students. Virtually all students in the program year after year attend four year colleges, and many plan to major in science, math, and engineering fields. Though the explicit goal of a enrolling in a physics class is to learn physics, students’ classroom decisions are informed by other personal goals, such as:

Getting a good grade. This goal may ultimately be a sub-goal of other goals such as getting into one’s favorite college or pleasing one’s parents. Students know that, particularly in traditional science instruction, grades are earned by following directions and getting the right answer. As such, for many students learning physics is only an intermediate goal toward success on performance measures. In conducting a precise pattern of controlled experiments, Eric could have merely been fulfilling criteria he believed his teacher would use to evaluate his performance on *Airbags*.

Efficiency. Traditional science instruction often rewards students for being able to solve short problems quickly. Minimizing the time spent on tasks also permits students to spend time on other scholastic goals. Brett’s and Eric’s highly efficient approach of exactly one set of controlled trials per investigation represents the minimum number of trials needed to demonstrate a complete investigation.

Minimize cognitive effort. When school culture values performance (grades) over genuine learning, students may wish to expend the minimum effort required to complete the task to a level that is satisfactory to their teacher. For Brett and Eric, repeating the identical pattern of controlled trials three times did not require them to adjust their strategies in response to new ideas they encountered, or think of any new strategies at all.

Airbags (and WISE modules generally) aim to convince students to temporarily suspend their real-life roles as students in favor of a new role as a scientist (or simply a learner) by presenting relevant problems, valuing students’ ideas, and encouraging independent and creative thinking. Brett and Eric illustrated that the *isolate* version of *Airbags* was unsuccessful in encouraging them to prioritize inquiry over performance and efficiency. It is possible that the *compare* version would have been as successful to this end with Brett and Eric as it was with Joann and Linda. It is also possible that a fundamental change to their orientations toward classroom science would have been necessary for them to set new priorities.

Preparing students to be lifelong learners

Measuring students’ understanding of science by their ability to recall facts is becoming largely irrelevant. Information of all sorts, scientific or otherwise, is readily available at any time to anyone with an advanced wireless telephone. People can choose to be inundated by an enormous diversity of opinions on an endless range of topics via social media such as blogs, Facebook profiles, and Twitter posts. As this volume of information and multitude of views continue to grow, students will require the skills to incorporate these data into their own coherent view of the world. Students must learn to view this information critically and skeptically, to evaluate information on the basis of evidence, to determine which views are uninformed, and to integrate multiple perspectives.

New technologies not only give students access to the ideas of others, but they also provide students with the means to make their own voice heard over the din. As youths, students may want their voice to be heard only among their friends and colleagues. Later in life, these students may want to reach a wider audience with ideas that could have meaningful impacts on society. In order to gain this audience, students will need to be able to communicate accessibly, persuasively, and coherently.

Technology-based science instruction has not only the opportunity but also the responsibility to prepare students for this mass exchange of information. Engaging students in authentic inquiry requires students to sort through evidence, interpret data, evaluate divergent opinions, construct arguments in support of their views, and share their ideas with their classmates or science experts. These inquiry skills can promote lifelong learning.

I wish to close with an example concerning the development of the sport of baseball, one of my few hobbies. Baseball is currently undergoing an economic revolution driven by modern quantitative analysis methods collectively known as “sabermetrics.” Until about ten years ago, players were evaluated primarily based on descriptive and often subjective accounts by scouts. Only a small handful of basic statistics were commonly used to characterize players’ myriad skills. Today, statistics capture minute details of a players’ every swing or pitch. Video cameras can now track the trajectories of every pitch and batted ball and the movements of every player on the field during a game, data that will lead to new insights about hitting, pitching, baserunning and defense. Analysts use these enormous databases to develop statistical models that can predict player and team performance with an accuracy that is unprecedented in the history of professional sports. Some models even capture aging patterns that are specific to players with a particular set of physical attributes. The generation and refinement of predictive models that are based on repeated observations and that incorporate increasingly sophisticated measurement techniques to further our understanding of the sport can be summarized in one simple statement: baseball has become a science.

One of the products of the sabermetric revolution has been a new genre of baseball writing (often via blogs) that uses a style that combines traditional commentary with modern quantitative analysis. In an interview in November 2009, the popular sportswriter Bill Simmons of espn.com discussed the impact of sabermetrics on the way he writes about and appreciates baseball:

> In my opinion, it’s not even that fun to follow baseball anymore, because you’re not allowed to have any opinions. You have to look up every opinion you’re supposed to have. “Oh, is [New York Yankees third baseman Alex Rodriguez] clutch? Let me look that up. Yes, he’s hitting .356 in the clutch. So I guess that means he’s clutch.” What’s fun about it? It’s like algebra.[1]

---

[1] Tobias, S. (2009, November 12). *Interview: Bill Simmons.* Retrieved April 23, 2010, from A.V. Club website: http://www.avclub.com/articles/bill-simmons,35319/

The public availability of data and the ubiquity of modern analysis methods have raised the standard for baseball journalism. Simmons came to realize that his purely subjective (and self-admittedly biased) evaluations of a player's ability were no longer sufficient to reach a sophisticated baseball audience. Without supporting his views on the sport with data, he was unable to communicate effectively with many of his readers. By his own description, Simmons' reluctance to achieve proficiency with modern analysis methods caused him to stop writing seriously about baseball for two years.

In April of the current 2010 major league season, Simmons provided an update in his column on his progress toward understanding sabermetrics:

> Cautiously, nervously, I started researching the advanced stats… I spent March reading and surfing sabermetrics … because the advanced formulas weren't nearly as intimidating as I had expected. Full disclosure: I, um ... I—I kinda like them. I even understand why stat junkies take it so personally whenever a mainstream [journalist] spouts out an uninformed baseball opinion. It's too easy to be informed these days. Takes a lot less time than you might think.[2]

Though Simmons found a way to re-enter the baseball conversation, his self-imposed absence was unfortunate and unnecessary for both him and his readers. In the April column, Simmons described the ways modern analytics gave him a new understanding of the impact of Freddie Lynn, his favorite player from childhood. Simmons regarded Lynn's baseball card so frequently that he has the traditional statistics from Lynn's greatest seasons memorized. However, modern statistics were able to give Simmons new insights into the ways Lynn's performance stood out among his peers from those years. Sabermetrics validated beyond any subjective account that Lynn was indeed as good as Simmons' had perceived him to be.

Many educators point to the ways science instruction can prepare students for careers in science, mathematics, and engineering. Yet the benefits of science education extend far beyond careers and into learners' appreciation of everyday life. Science education must prepare students to welcome new sources of information and methods of analysis rather than fear them and to engage more deeply in their favorite conversations rather than withdraw from them as a result of ignorance. Science instruction must aim to provide students with the tools they need to engage productively within a community, be a contributing member of society, and gain a new appreciation for the world around them.

---

[2] Simmons, B. (2010, April 2). *Finally joining the revolution.* Retrieved April 23, 2010, from ESPN Page 2 website: http://sports.espn.go.com/espn/page2/story?page=simmons/100402

References

Ainsworth, S. (1999). The functions of multiple representations. *Computers & Education, 33*(2-3), 131-152.

Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction, 16*(3), 183-198.

American Association for the Advancement of Science (1993). *Benchmarks for Science Literacy*. New York: Oxford University Press.

Asaria, P., & MacMahon, E. (2006). Measles in the United Kingdom: can we eradicate it by 2010? *British Medical Journal, 333*(7574), 890.

Barab, S., Hay, K., Barnett, M., & Keating, T. (2000). Virtual solar system project: Building understanding through model building. *Journal of Research in Science Teaching, 37*(7), 719-756.

Bell, P., & Linn, M. (2000). Scientific arguments as learning artifacts: designing for learning from the web with KIE. *International Journal of Science Education, 22*(8), 797-817.

Brown, A. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *Journal of the Learning Sciences, 2*(2), 141-178.

Brown, A., & Campione, J. (1996). Psychological theory and the design of innovative learning environments: On procedures, principles, and systems. *Innovations in learning: New environments for education*, 289-325.

Buckley, B., Gobert, J., & Horwitz, P. (2006). Using Log Files to Track Students' Model-Based Inquiry. In S. Barab, K. Hay & D. Hickey (Eds.)*, Making a difference: Proceedings of the Seventh International Conference of the Learning Sciences* (pp. 57-63). Mahwah, NJ: Lawrence Erlbaum Associates.

California Department of Education (2000). *Science Content Standards for California Public Schools*. Sacramento, CA: California Department of Education.

Carey, S., Evans, R., Honda, M., Jay, E., & Unger, C. (1989). 'An experiment is when you try it and see if it works': a study of grade 7 students' understanding of the construction of scientific knowledge. *International Journal of Science Education, 11*(5), 514-529.

Carey, S., & Smith, C. (1993). On understanding the nature of scientific knowledge. *Educational Psychologist, 28*(3), 235-251.

Chandler, P. (2004). The crucial role of cognitive processes in the design of dynamic visualizations. *Learning and Instruction, 14*(3), 353-357.

Chandler, P., & Sweller, J. (1991). Cognitive Load Theory and the Format of Instruction. *Cognition and Instruction, 8*(4), 293-332.

Chang, H., Chiu, J., McElhaney, K., & Linn, M. Can Dynamic Visualizations Improve Science Learning: A Synthesis Study. *Manuscript in preparation*.

ChanLin, L. (2001). Formats and prior knowledge on learning in a computer-based lesson. *Journal of Computer Assisted Learning, 17*(4), 409-419.

Chen, Z., & Klahr, D. (1999). All other things being equal: Acquisition and transfer of the control of variables strategy. *Child Development, 70*(5), 1098-1120.

Chi, M., De Leeuw, N., Chiu, M., & LaVancher, C. (1994). Eliciting Self-Explanations Improves Understanding. *Cognitive Science, 18*(3), 439-477.

Chinn, C., & Brewer, W. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of Educational Research, 63*(1), 1-49.

Chiu, J., & Linn, M. (2008). Self-Assessment and Self-Explanation for Learning Chemistry Using Dynamic Molecular Visualizations *International Perspectives in the Learning Sciences: Cre8ting a Learning World. Proceedings of the 8th International Conference of the Learning Sciences* (Vol. 3, pp. 16-17). Utrecht, The Netherlands: International Society of the Learning Sciences, Inc.

Clark, D., & Jorde, D. (2004). Helping students revise disruptive experientially supported ideas about thermodynamics: Computer visualizations and tactile models. *Journal of Research in Science Teaching, 41*(1), 1-23.

Clark, D., & Linn, M. (2003). Designing for knowledge integration: The impact of instructional time. *Journal of the Learning Sciences, 12*(4), 451-493.

Clark, D., & Sampson, V. (2007). Personally-seeded discussions to scaffold online argumentation. *International Journal of Science Education, 29*(3), 253-277.

Clark, D., & Slotta, J. (2000). Evaluating media-enhancement and source authority on the internet: the Knowledge Integration Environment. *International Journal of Science Education, 22*(8), 859-871.

Clement, J. (1993). Using Bridging Analogies and Anchoring Intuitions to Deal with Students' Preconceptions in Physics. *Journal of Research in Science Teaching, 30*(10), 1241-1257.

Cognition and Technology Group at Vanderbilt (1992). The Jasper series as an example of anchored instruction: Theory, program description and assessment data. Educational Psychologist, 27, 291-315.

Collins, A., Joseph, D., & Bielaczyc, K. (2004). Design research: Theoretical and methodological issues. *Journal of the Learning Sciences, 13*(1), 15-42.

Crouch, C., Fagen, A., Callan, J., & Mazur, E. (2004). Classroom demonstrations: Learning tools or entertainment? *American Journal of Physics, 72*(6), 835-838.

Dean, D., & Kuhn, D. (2007). Direct instruction vs. discovery: The long view. *Science Education, 91*(3), 384-397.

diSessa, A. (1988). Knowledge in pieces. In G. Forman & P. Pufall (Eds.), *Constructivism in the computer age* (pp. 49-70). Hillsdale, NJ: Lawrence Erlbaum Associates.

diSessa, A., Elby, A., & Hammer, D. (2002). J's Epistemological Stance and Strategies. In G. Sinatra & P. Pintrich (Eds.), *Intentional conceptual change* (pp. 237-290). Mahwah, NJ: Lawrence Erlbaum.

diSessa, A., & Sherin, B. (1998). What changes in conceptual change? *International Journal of Science Education, 20*(10), 1155-1191.

Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education, 84*(3), 287-312.

Dunbar, K. (1993). Concept discovery in a scientific domain. *Cognitive Science: A Multidisciplinary Journal, 17*(3), 397-434.

Edelson, D. (2001). Learning-for-use: A framework for the design of technology-supported inquiry activities. *Journal of Research in Science Teaching, 38*(3), 355-385.

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), 391-450.

Fortus, D., Dershimer, R., Krajcik, J., Marx, R., & Mamlok-Naaman, R. (2004). Design-based science and student learning. *Journal of Research in Science Teaching, 41*(10), 1081-1110.

Gilbert, J., & Boulter, C. (2000). *Developing models in science education*. Dordrecht, The Netherlands: Kluwer.

Gupta, R., Best, J., & MacMahon, E. (2005). Mumps and the UK epidemic 2005. *British Medical Journal, 330*(7500), 1132.

Hays, T. (1996). Spatial Abilities and the Effects of Computer Animation on Short-Term and Long-Term Comprehension. *Journal of educational Computing research, 14*(2), 139-155.

Hegarty, M. (2004). Dynamic visualizations and learning: Getting to the difficult questions. *Learning and Instruction, 14*(3), 343-351.

Hegarty, M., Kriz, S., & Cate, C. (2003). The roles of mental animations and external animations in understanding mechanical systems. *Cognition and Instruction, 21*(4), 325-360.

Höffler, T., & Leutner, D. (2007). Instructional animation versus static pictures: A meta-analysis. *Learning and Instruction, 17*(6), 722-738.

Howe, C. (1998). *Conceptual structure in childhood and adolescence: The case of everyday physics*. London: Routledge.

Inhelder, B., & Piaget, J. (1958). *The Growth of Logical Thinking*. New York: Basic Books.

Jackson, S., Stratford, S., Krajcik, J., & Soloway, E. (1994). Making dynamic modeling accessible to precollege science students. *Interactive Learning Environments, 4*(3), 233-257.

Kanari, Z., & Millar, R. (2004). Reasoning from data: How students collect and interpret data in science investigations. *Journal of Research in Science Teaching, 41*(7), 748-769.

Keselman, A. (2003). Supporting inquiry learning by promoting normative understanding of multivariable causality. *Journal of Research in Science Teaching, 40*(9), 898-921.

Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science: A Multidisciplinary Journal, 12*(1), 1-48.

Klahr, D., Fay, A., & Dunbar, K. (1993). Heuristics for scientific experimentation: A developmental study. *Cognitive Psychology, 25*(1), 111-146.

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

Klahr, D., Triona, L., & Williams, C. (2007). Hands on what? The relative effectiveness of physical versus virtual materials in an engineering design project by middle school children. *Journal of Research in Science Teaching, 44*(1), 183-203.

Klayman, J., & Ha, Y. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review, 94*(2), 211-228.

Kolodner, J., Camp, P., Crismond, D., Fasse, B., Gray, J., Holbrook, J., et al. (2003). Problem-Based Learning Meets Case-Based Reasoning in the Middle-School Science Classroom: Putting Learning by Design™ Into Practice. *Journal of the Learning Sciences, 12*(4), 495-547.

Koslowski, B. (1996). *Theory and evidence: The development of scientific reasoning*: The MIT Press.

Kozma, R. (2003). The material features of multiple representations and their cognitive and social affordances for science understanding. *Learning and Instruction, 13*(2), 205-226.

Krajcik, J., Blumenfeld, P., Marx, R., & Soloway, E. (1999). Instructional, curricular, and technological supports for inquiry in science classrooms. In J. Minstrell & E. van.Zee (Eds.), *Inquiry into inquiry: Science learning and teaching* (pp. 283-315). Washington, DC: AAAS Press.

Kuhn, D., Black, J., Keselman, A., & Kaplan, D. (2000). The development of cognitive skills to support inquiry learning. *Cognition and Instruction*, *18*(4), 495-523.

Kuhn, D., & Dean, D. (2005). Is developing scientific thinking all about learning to control variables? *Psychological Science, 16*(11), 866-870.

Kuhn, D., Garcia-Mila, M., Zohar, A., Andersen, C., White, S., Klahr, D., et al. (1995). Strategies of knowledge acquisition. *Monographs of the Society for Research in Child Development, 60*(4), 1-157.

Kuhn, D., & Phelps, E. (1982). The development of problem-solving strategies. *Advances in child development and behavior, 17*, 1-44.

Kuhn, T. (1970). *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Latour, B., & Woolgar, S. (1986). *Laboratory Life: The Construction of Scientific Facts*. Princeton, NJ: Princeton University Press.

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

Lehrer, R., & Schauble, L. (2004). Modeling natural variation through distribution. *American Educational Research Journal, 41*(3), 635-679.

Leinhardt, G., Zaslavsky, O., & Stein, M. (1990). Functions, graphs, and graphing: Tasks, learning, and teaching. *Review of educational research, 60*(1), 1-64.

Lewis, E. (1996). Conceptual change among middle school students studying elementary thermodynamics. *Journal of Science Education and Technology, 5*(1), 3-31.

Linn, M. (2005). WISE Design for Lifelong Learning—Pivotal Cases. In P. Gardenfors & P. Johansson (Eds.), *Cognition, Education, and Communication Technology* (pp. 223-255). Mahwah, NJ: Lawrence Erlbaum Associates.

Linn, M., Clement, C., & Pulos, S. (1983). Is It Formal If It's Not Physics? (The Influence of Laboratory and Naturalistic Content on Formal Reasoning). *Journal of Research in Science Teaching, 20*(8), 755-770.

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

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

Linn, M., & Hsi, S. (2000). *Computers, teachers, peers: Science learning partners*: Mahwah, NJ: Lawrence Erlbaum Associates.

Linn, M., Lee, H., Tinker, R., Husic, F., & Chiu, J. (2006). Teaching and assessing knowledge integration in science. *Science, 313*(5790), 1049-1050.

Linn, M., & Songer, N. (1991). Teaching thermodynamics to middle school students: What are appropriate cognitive demands? *Journal of Research in Science Teaching, 28*(10), 885-918.

Linn, M., & Swiney, J. (1981). Individual differences in formal thought: Role of expectations and aptitudes. *Journal of Educational Psychology, 73*, 274-286.

Mayer, R., & Sims, V. (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of educational psychology, 86*(3), 389-401.

McCloskey, M. (1983). Naive theories of motion. In D. Gentner & A.Stevens (Eds.), *Mental models* (pp. 299-324). Hillsdale, NJ: Lawrence Erlbaum Associates.

McDermott, L. (1990). Research and computer based instruction: Opportunity for interaction. *American Journal of Physics, 58*(5), 452-462.

McDermott, L., Rosenquist, M., & Zee, E. (1987). Student difficulties in connecting graphs and physics: Exampies from kinematics. *American Journal of Physics, 55*, 505-513.

McIntyre, P., & Leask, J. (2008). Improving uptake of MMR vaccine. *British Medical Journal, 7647*, 729.

Merriënboer, J. J. G. v., Kirschner, P. A., & Kester, L. (2003). Taking the Load Off a Learner's Mind: Instructional Design for Complex Learning. *Educational Psychologist, 38*(1), 5-13.

Minstrell, J. (2001). The Role of the Teacher in Making Sense of Classroom Experiences and Effecting Better Learning. In S. M. Carver & D. Klahr (Eds.), *Cognition and Instruction: Twenty-Five Years of Progress* (pp. 121-149). Mahwah, NJ: Lawrence Erlbaum Associates.

Moreno, R., & Mayer, R. (2007). Interactive multimodal learning environments. *Educational psychology review, 19*(3), 309-326.

National Council of Teachers of Mathematics (2000). *Principles and standards for school mathematics.* Reston, VA: NCTM.

National Research Council. (1996). *National Science Education Standards*. National Committee on Science Education Standards and Assessment. Washington, DC: National Academy Press.

Njoo, M., & Jong, T. (1993). Exploratory learning with a computer simulation for control theory: Learning processes and instructional support. *Journal of Research in Science Teaching, 30*(8), 821-844.

Parnafes, O., & diSessa, A. (2004). Relations between types of reasoning and computational representations. *International Journal of Computers for Mathematical Learning, 9*(3), 251-280.

Quintana, C., Krajcik, J., & Soloway, E. (2001). Exploring a description and methodology for learner-centered design. In W. Heineke & L. Blasi (Eds.), *Methods of evaluating educational technology* (Vol. 1, pp. 125–146). Greenwich, CT: Information Age Pub Inc.

Quintana, C., Reiser, B., Davis, E., Krajcik, J., Fretz, E., Duncan, R., et al. (2004). A scaffolding design framework for software to support science inquiry. *Scaffolding: A Special Issue of the Journal of the Learning Sciences, 13*(3), 337-386.

Reid, D., Zhang, J., & Chen, Q. (2003). Supporting scientific discovery learning in a simulation environment. *Journal of Computer Assisted Learning, 19*(1), 9-20.

Reiser, B., Tabak, I., Sandoval, W., Smith, B., Steinmuller, F., & Leone, A. (2001). BGuILE: Strategic and conceptual scaffolds for scientific inquiry in biology classrooms. In S. M. Carver & D. Klahr (Eds.), *Cognition and instruction: Twenty-five years of progress* (pp. 263-305). Mahwah, NJ: Lawrence Erlbaum Associates.

Sandoval, W., & Millwood, K. (2005). The Quality of Students Use of Evidence in Written Scientific Explanations. *Cognition and Instruction, 23*(1), 23-55.

Sandoval, W., & Reiser, B. (2004). Explanation-driven inquiry: Integrating conceptual and epistemic scaffolds for scientific inquiry. *Science Education, 88*(3), 345-372.

Scardamalia, M., & Bereiter, C. (1994). Computer support for knowledge-building communities. *Journal of the Learning Sciences, 3*(3), 265-283.

Schauble, L. (1990). Belief revision in children: The role of prior knowledge and strategies for generating evidence* 1. *Journal of Experimental Child Psychology, 49*(1), 31-57.

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

Schwan, S., & Riempp, R. (2004). The cognitive benefits of interactive videos: learning to tie nautical knots. *Learning and Instruction, 14*(3), 293-305.

Shipstone, D. (1985). Electricity in simple circuits. In R. Driver, E. Guesne & A. Tiberghien (Eds.), *Children's ideas in science* (pp. 33-51). Milton Keynes: Open University Press.

Siegler, R., & Liebert, R. (1975). Acquisition of formal scientific reasoning by 10-and 13-year-olds: Designing a factorial experiment. *Developmental Psychology, 11*(3), 401-402.

Spitulnik, M., Krajcik, J., & Soloway, E. (1999). Construction of models to promote scientific understanding. *Modeling and simulation in science and mathematics education*, 70–94.

Stigler, J., & Hiebert, J. (1999). *The Teaching Gap*. New York: Free Press.

Stratford, S., Krajcik, J., & Soloway, E. (1998). Secondary students' dynamic modeling processes: Analyzing, reasoning about, synthesizing, and testing models of stream ecosystems. *Journal of Science Education and Technology, 7*(3), 215-234.

Strike, K., & Posner, G. (1985). A conceptual change view of learning and understanding. In L. West & A. Piines (Eds.), *Cognitive structure and conceptual change*. Orlando, FL: Academic Press.

Strike, K., & Posner, G. (1992). A revisionist theory of conceptual change. In R. Duschl & R. Hamilton (Eds.), *Philosophy of science, cognitive psychology, and educational theory and practice* (pp. 147-176). Albany, NY: SUNY.

Swaak, J., Van Joolingen, W., & de Jong, T. (1998). Supporting simulation-based learning; the effects of model progression and assignments on definitional and intuitive knowledge. *Learning and Instruction, 8*(3), 235-252.

Tao, P. (1997). Confronting students' alternative conceptions in mechanics with the Force and Motion Microworld. *Computers in Physics, 11*(2), 199-207.

Thagard, P. (1992). *Conceptual Revolutions*. Princeton, NJ: Princeton University Press.

Triona, L., & Klahr, D. (2003). Point and Click or Grab and Heft: Comparing the Influence of Physical and Virtual Instructional Materials on Elementary School Students Ability to Design Experiments. *Cognition and Instruction, 21*(2), 149-173.

Trowbridge, D., & McDermott, L. (1980). Investigation of student understanding of the concept of velocity in one dimension. *American Journal of Physics, 48*(12), 1020-1028.

Trowbridge, D., & McDermott, L. (1981). Investigation of Student Understanding of the Concept of Acceleration in One Dimension. *Am. J. Phys., 49*(3), 242-253.

Tschirgi, J. (1980). Sensible reasoning: A hypothesis about hypotheses. *Child Development, 51*(1), 1-10.

van Zee, E., & Minstrell, J. (1997). Reflective discourse: Developing shared understandings in a physics classroom. *International Journal of Science Education, 19*(2), 209-228.

Varma, K. (2010). Using Interactive Models to Support Content Learning through Scientific Reasoning *Proceedings of the 9th International Conference of the Learning Sciences*. Chicago, IL: International Society of the Learning Science, Inc.

Varma, K., Husic, F., & Linn, M. (2008). Targeted support for using technology-enhanced science inquiry modules. *Journal of Science Education and Technology, 17*(4), 341-356.

Vollmeyer, R., Burns, B., & Holyoak, K. (1996). The impact of goal specificity on strategy use and the acquisition of problem structure. *Cognitive Science: A Multidisciplinary Journal, 20*(1), 75-100.

Wakefield, A., Murch, S., Anthony, A., Linnell, J., Casson, D., Malik, M., et al. (1998). Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. *The Lancet, 351*(9103), 637-641.

White, B. (1993). ThinkerTools: Causal models, conceptual change, and science education. *Cognition and Instruction*, *10*(1), 1-100.

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

Wu, H., Krajcik, J., & Soloway, E. (2001). Promoting conceptual understanding of chemical representations: Students’ use of a visualization tool in the classroom. *Journal of Research in Science Teaching, 38*(7), 821-842.

Yang, E., Andre, T., & Greenbowe, T. (2003). Spatial Ability and the Impact of Visualization/Animation on Learning Electrochemistry. *International Journal of Science Education, 25*(3), 329-349.

Zhang, H. (2010). Exploring Drawing and Critique to Enhance Learning from Visualizations *Proceedings of the International Conference of the Learning Sciences*. Chicago, IL: International Society of the Learning Sciences.

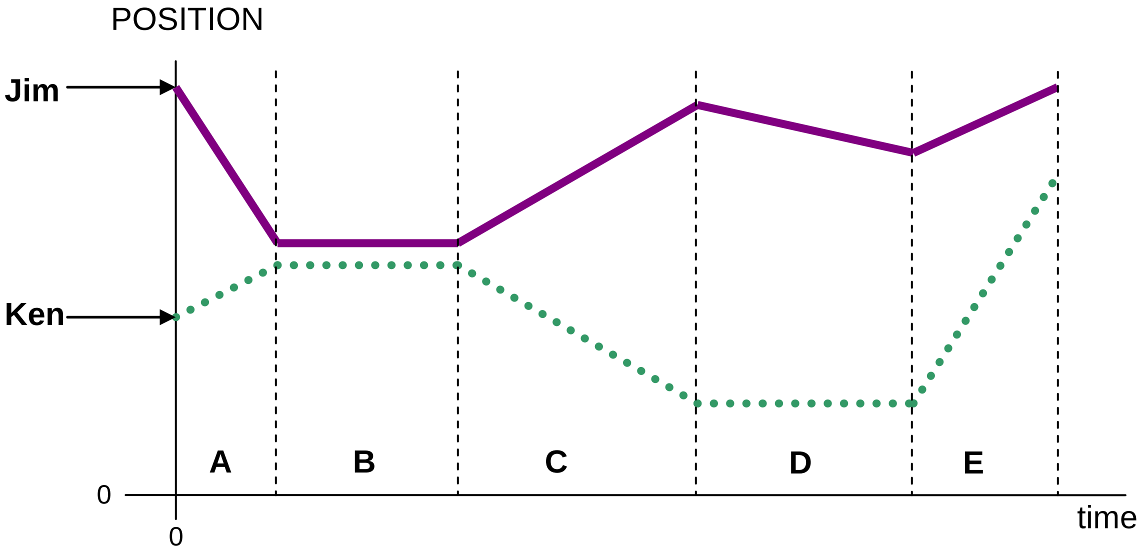
Appendices

The following appendices contain the full text and accompanying diagrams of each pretest, posttest, and embedded assessment item used for all three empirical studies, as well as rubrics used to score the assessments. Rubrics for posttest assessments are not included, as they are conceptually the same as the ones used for the corresponding pretest items.

Appendix A

Pretest and posttest for Study 1

1. The graph shows the motion of two soccer players, Jim and Ken, who move in a straight line on the field. The graph shows their POSITION versus time. Jim's graph is on top, Ken's graph is on the bottom.



a) [SOCCER_1] During which time segments does Jim travel faster than Ken? (Circle **all that apply**.)

A B C D E

Explain your answer:

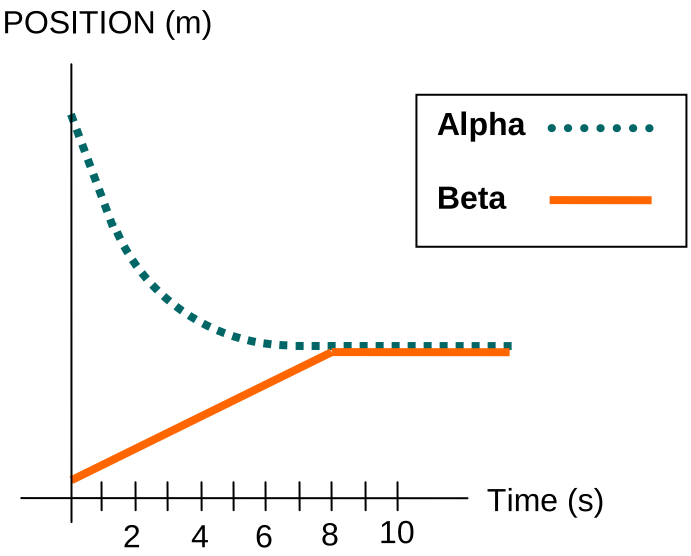
b) [SOCCER_2] Which of the following statements applies to Jim's and Ken's motion during time segment **A**? (Check **all that apply**.)

__ Jim and Ken travel toward each other
__ Jim and Ken travel away from each other
__ Jim and Ken travel in the same direction
__ Jim is not moving
__ Ken is not moving

Explain your answer:

2. A technology company researches “smart-cars” equipped with motion devices that can help cars avoid collisions. When the device works properly, it **slows the car down gradually** to avoid a collision.

A graph below shows the POSITION vs. time of two cars, Alpha and Beta, in a head-on collision testing situation.



a) [SMART_ALPHA] According to the graph did the motion device work properly for the **Alpha** car? (Circle one.)

Yes No

Explain your answer:

b) [SMART_BETA] According to the graph did the motion device work properly for the **Beta** car? (Circle one.)

Yes No

Explain your answer:

3. [COLLISION] 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.

5’3” 165 lb. male

5’11” 120 lb. female

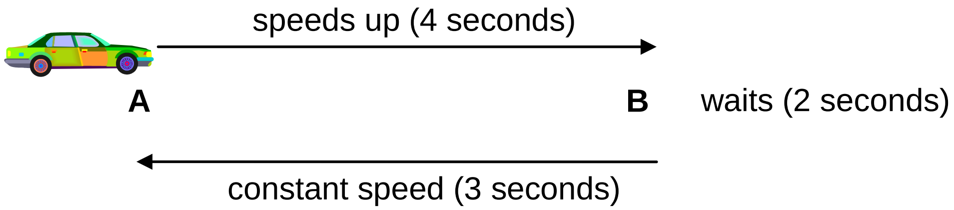
Which person do you think is **more likely to be injured** by an airbag deploying? (Circle one.)

The MAN on the left

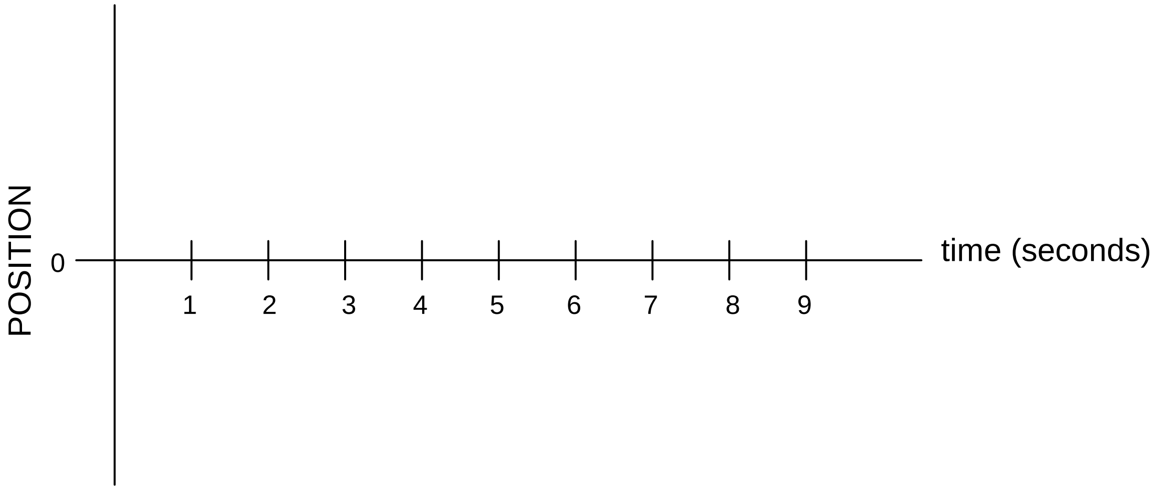
The WOMAN on the right

Explain your answer:

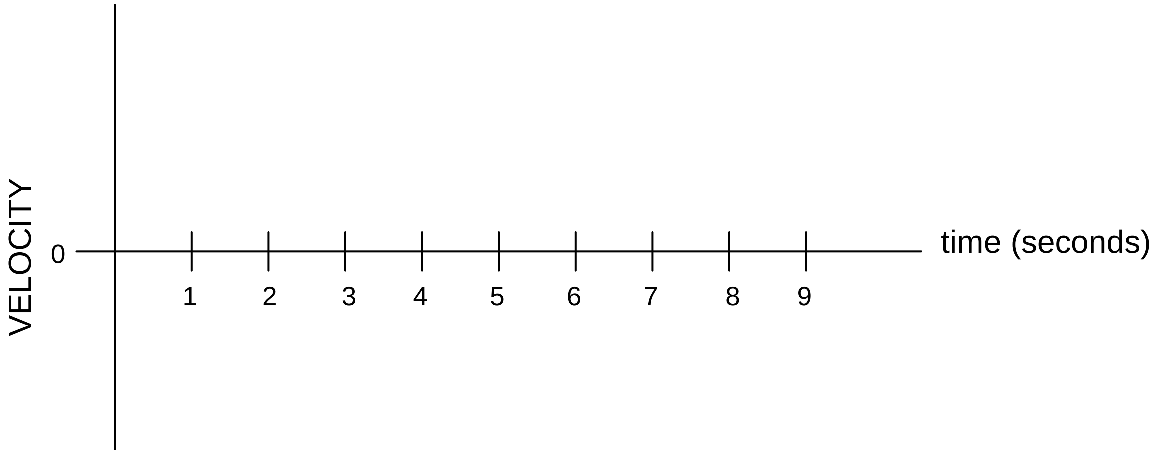
4. A car starts at point A and speeds up at a constant rate until it reaches point B in 4 seconds, where it suddenly stops. The car waits at point B for 2 seconds. It then travels at constant speed in the opposite direction, reaching point A again in another 3 seconds.



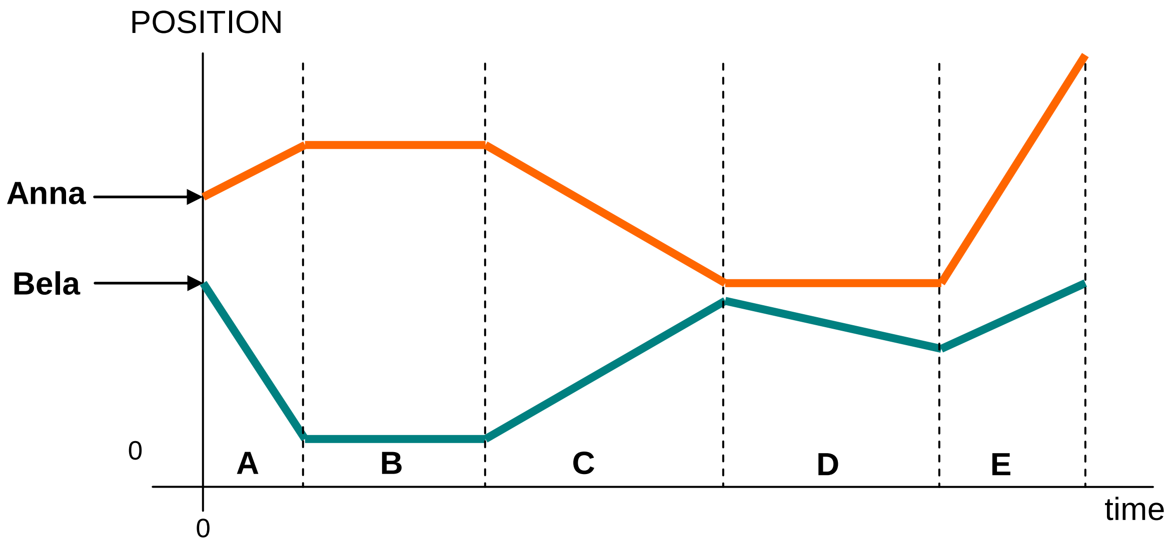
[SKETCH_POSITION] Sketch a **POSITION-TIME graph** of the motion during these 9 seconds.



[SKETCH_VELOCITY] Sketch a **VELOCITY-TIME graph** of the motion during these 9 seconds.



1. The graph shows the motion of two soccer players, Anna and Bela, who move in a straight line on the field. The graph shows their POSITION versus time. Anna's graph is on top, Bela's graph is on the bottom.



a) During which time segments does Bela travel faster than Anna? (Circle **all that apply.)**

A B C D E

Explain your answer:

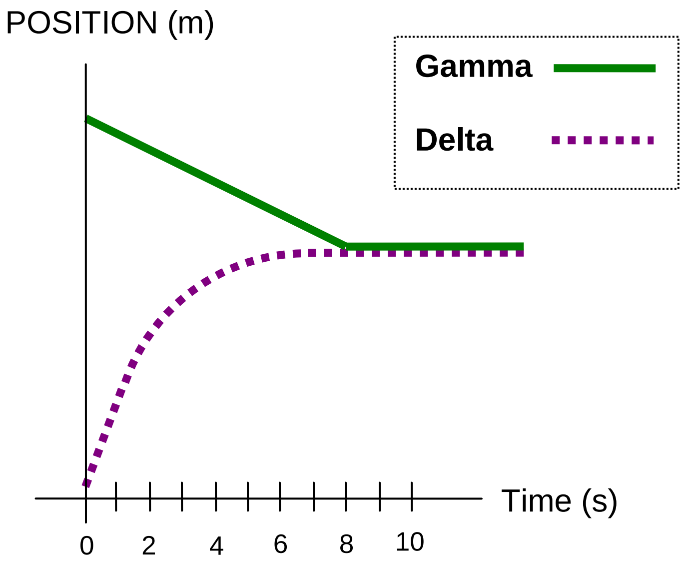
b) Which of the following statements applies to Anna's and Bela's motion during time segment **A**? (Check **all that apply**.)

__ Anna and Bela travel toward each other
__ Anna and Bela travel away from each other
__ Anna and Bela travel in the same direction
__ Anna is not moving
__ Bela is not moving

Explain your answer:

2. A technology company researches “smart-cars” equipped with motion devices that can help cars avoid collisions. When the device works properly, it **slows the car down gradually** to avoid a collision.

A graph below shows the POSITION vs. time of two cars, Gamma and Delta, in a head-on collision testing situation.



a) According to the graph did the motion device work properly for the **Gamma** car? (Circle one.)

Yes No

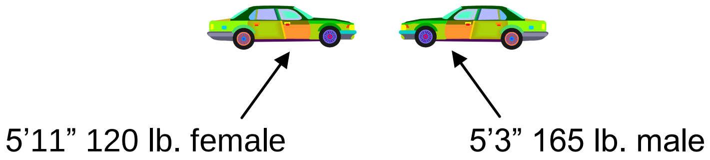
Explain your answer:

b) According to the graph did the motion device work properly for the **Delta** car? (Circle one.)

Yes No

Explain your answer:

3. 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.

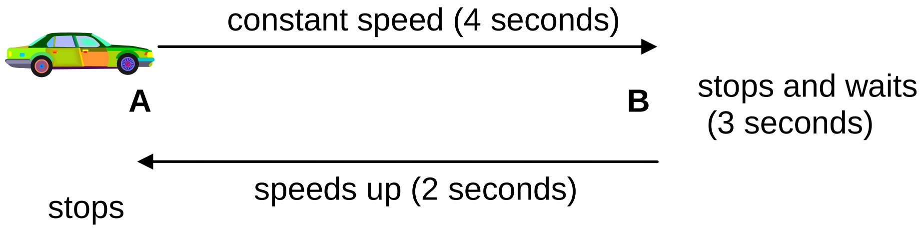


Which person do you think is **more likely to be injured** by an airbag deploying? (Circle one.)

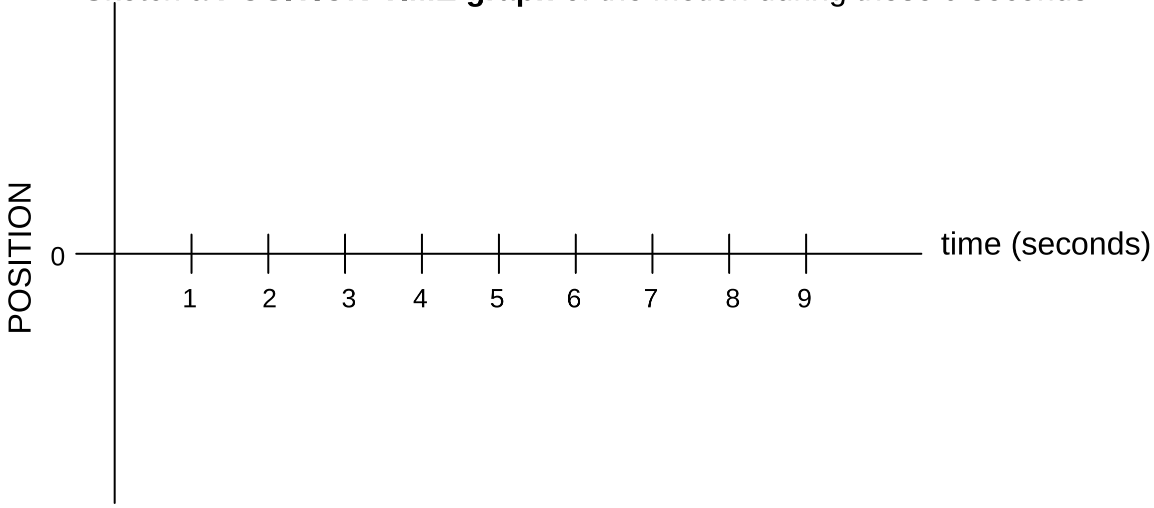
The WOMAN on the left The MAN on the right

Explain your answer:

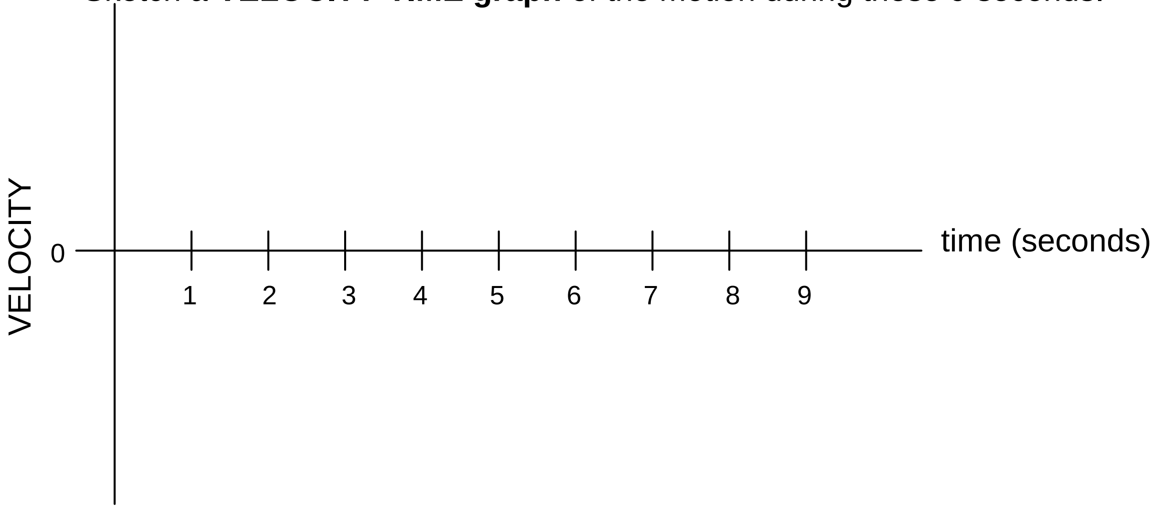
4. A car starts at point A and travels at constant speed until it reaches point B in 4 seconds, where it suddenly stops. The car waits at point B for 3 seconds. It then speeds up at a constant rate in the opposite direction, reaching point A again in another 2 seconds.



Sketch a **POSITION-TIME graph** of the motion during these 9 seconds.



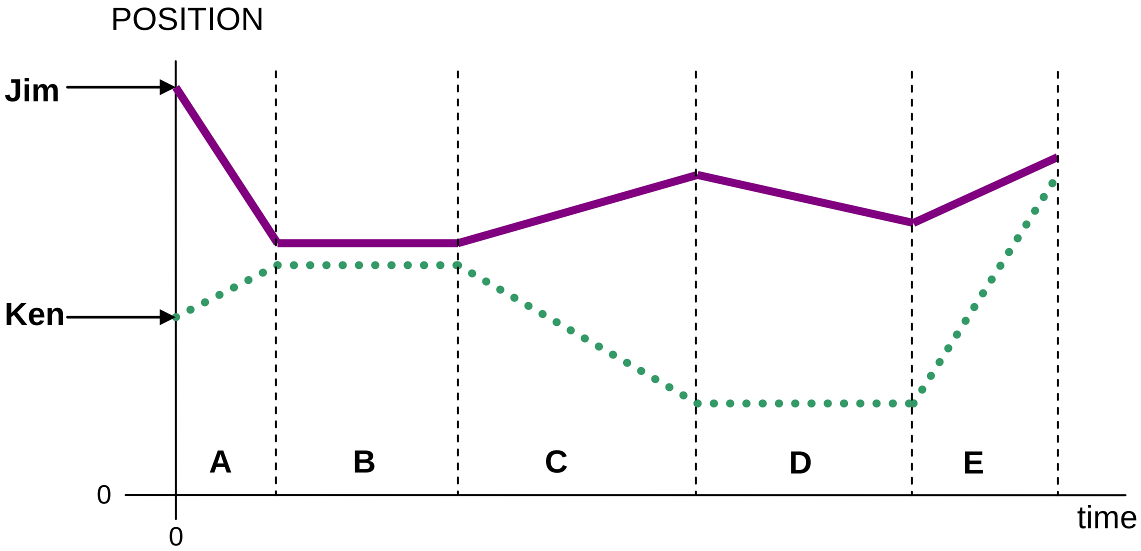
Sketch a **VELOCITY-TIME graph** of the motion during these 9 seconds.



Appendix B

Pretest and posttest for Study 2 and Study 3

1. The graph shows the motion of two soccer players, Jim and Ken, who each move in a straight line on the field. The graph shows their POSITION versus time. Jim's graph is on top, Ken's graph is on the bottom.



(a) During which time segments does Jim travel faster than Ken? **(Circle __all that apply__.)**

A B C D E

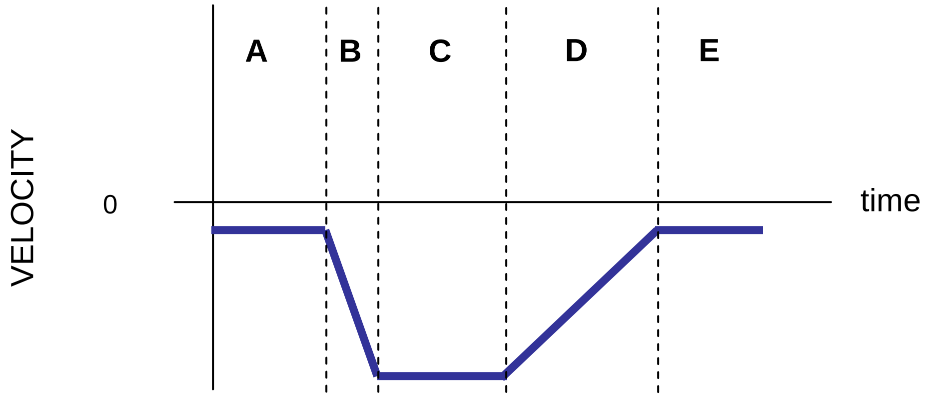
Explain your answer:

(b) During which time segments do Jim and Ken move in the same direction? **(Circle __all that apply__.)**

A B C D E

Explain you answer:

2. Some researchers are using motion sensing devices to study the motion patterns of wild animals. The devices provide graphs of an animal's VELOCITY vs. time as it tracks its prey. The graph below shows a segment of an animal's VELOCITY graph.



(a) [ANIMAL_1] Which of the following describes the motion of the animal during segment **D**? **(Check __one__.)**

___speeding up
___slowing down
___moving at constant speed
___not moving

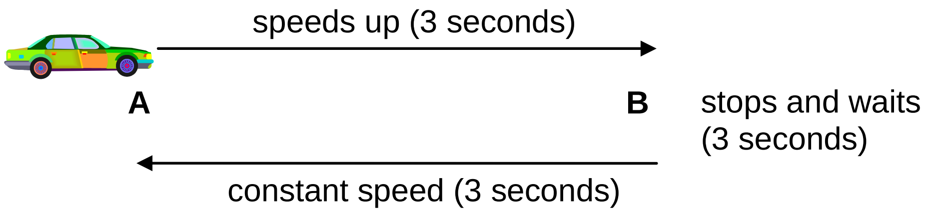
Explain your answer:

(b) [ANIMAL_2] Which of the following describes the motion of the animal during segment **C**? **(Check __one__.)**

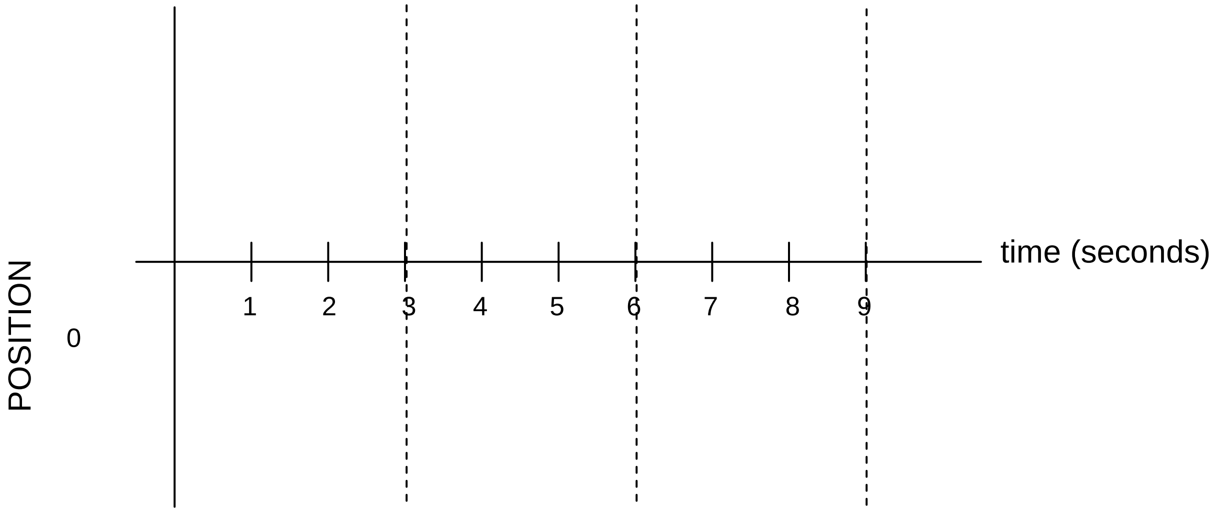
___speeding up
___slowing down
___moving at constant speed
___not moving

Explain your answer:

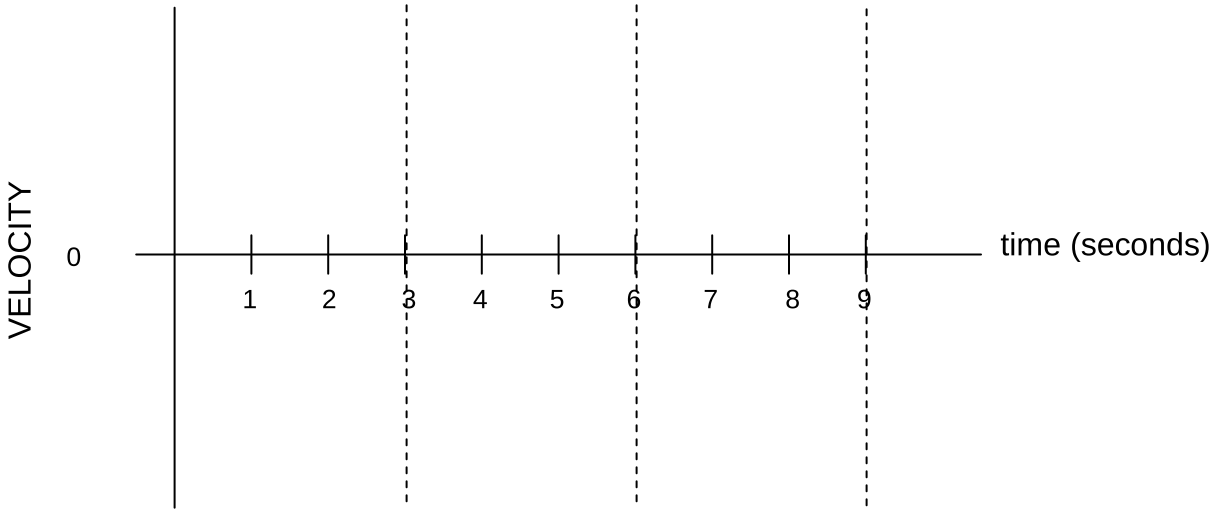
3. A car starts at point A and <u>speeds up</u> at a constant rate in the positive direction until it reaches point B in 3 seconds, where it suddenly stops. The car <u>waits</u> at point B for 3 seconds. It then travels at <u>constant speed</u> in the opposite direction, reaching point A again in another 3 seconds.



(a) Sketch a **POSITION-TIME graph** of the motion during these 9 seconds.



(b) Sketch a **VELOCITY-TIME graph** of the motion during these 9 seconds.



4. Amanda and Rosa want to know how to get the best gas mileage from their cars. They each conduct two trials and measure the gas mileage for each one. The tables show their results.

Amanda's test

	Trial 1	**Trial 2**
Speed	45 mph	65 mph
Windows	UP	UP
Gas mileage	25 mpg	21 mpg

Rosa's test

	Trial 1	**Trial 2**
Speed	50 mph	60 mph
Windows	UP	DOWN
Gas mileage	23 mpg	19 mpg

(a) [EXPERIMENT_1] Whose test BEST shows what happens to the car's gas mileage when the SPEED changes? **(Circle one.)**

Amanda's Rosa's Both

Explain your answer:

(b) [EXPERIMENT_2] According to the data, **what happens to a car's gas mileage** as a car's **speed** INCREASES? **(Circle one.)**

Increases Decreases Doesn't change

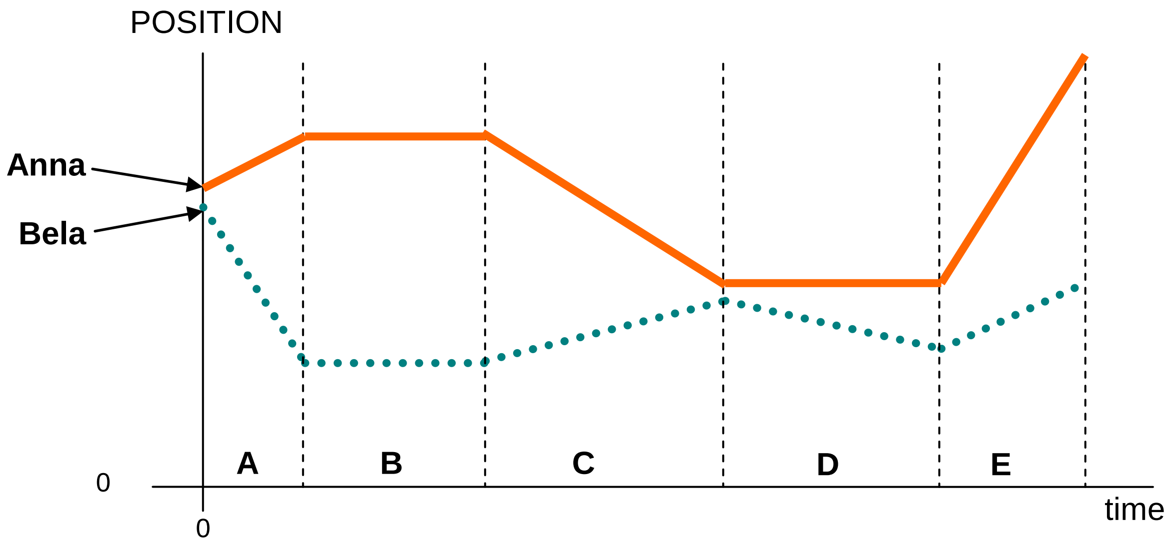
Explain using evidence from the data tables:

(c) [EXPERIMENT_3] What should Amanda and Rosa do to find out **how the WINDOWS (UP/DOWN) affect the gas mileage?** Complete the table below.

	Trial 1	**Trial 2**
Speed		
Windows		

Explain your answer:

1. The graph shows the motion of two soccer players, Anna and Bela, who move in a straight line on the field. The graph shows their POSITION versus time. Anna's graph is on top, Bela's graph is on the bottom.



(a) During which time segments does Bela travel faster than Anna? **(Circle <u>all that apply</u>.)**

A B C D E

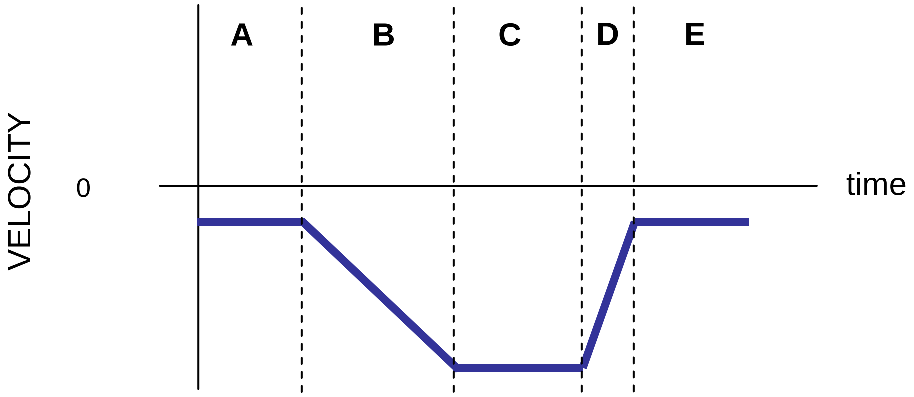
Explain your answer:

(b) During which time segments do Anna and Bela move in opposite directions? **(Circle <u>all that apply</u>.)**

A B C D E

Explain your answer:

2. Some researchers are using motion sensing devices to study the motion patterns of wild animals. The devices provide graphs of an animal's VELOCITY vs. time as it tracks its prey. The graph below shows a segment of an animal's VELOCITY graph.



(a) Which of the following describes the motion of the animal during segment **B**? **(Check <u>one</u>.)**

___speeding up
___slowing down
___moving at constant speed
___not moving

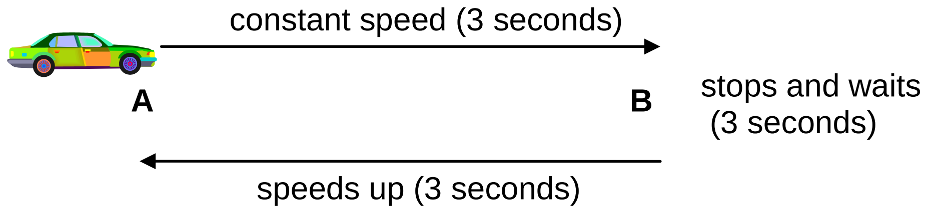
Explain your answer:

(b) Which of the following describes the motion of the animal during segment **C**? **(Check <u>one</u>.)**

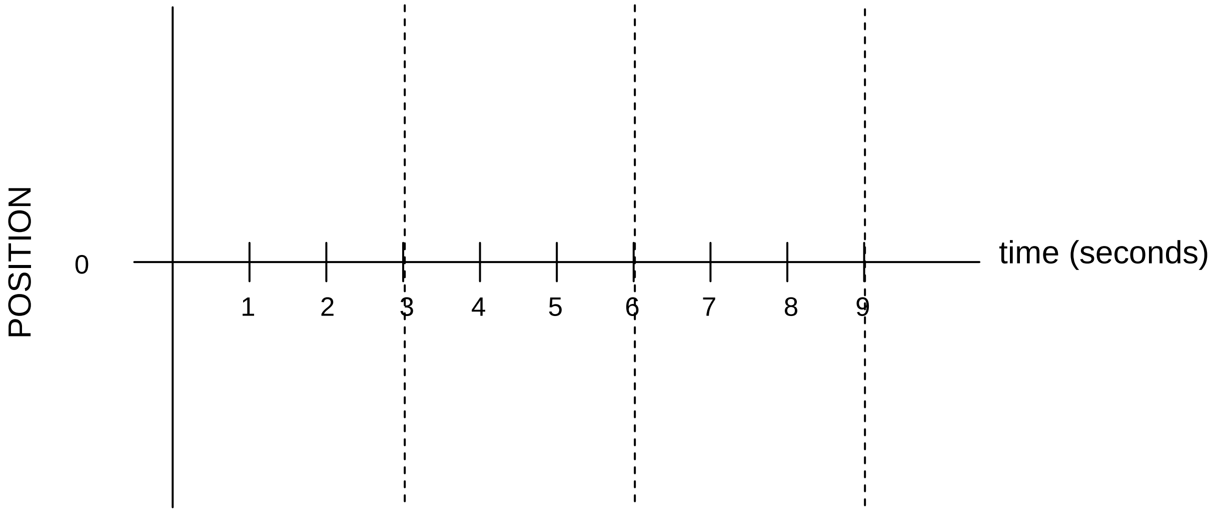
___speeding up
___slowing down
___moving at constant speed
___not moving

Explain your answer:

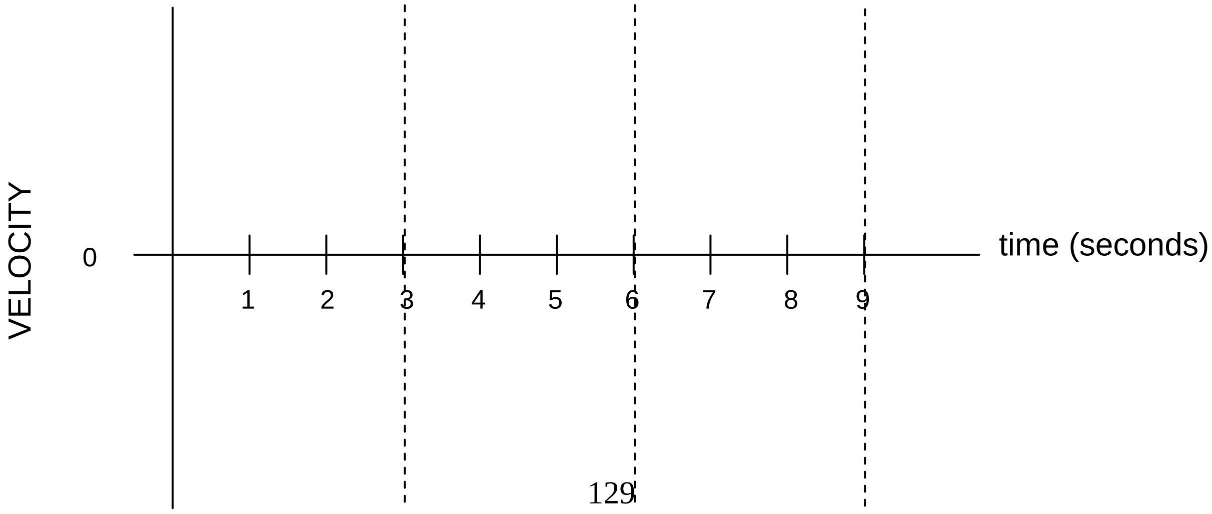
3. A car starts at point A and travels at constant speed in the positive direction until it reaches point B in 3 seconds, where it suddenly stops. The car waits at point B for 3 seconds. It then speeds up at a constant rate in the opposite direction, reaching point A again in another 3 seconds.



Sketch a **POSITION-TIME graph** of the motion during these 9 seconds.



Sketch a **VELOCITY-TIME graph** of the motion during these 9 seconds.



4. Chris and Pat want to know how to get the best gas mileage from their cars. They each conduct two trials and measure the gas mileage for each one. The tables show their results.

Chris's test

	Trial 1	**Trial 2**
Roof rack	ON	OFF
Tire pressure	26 psi	28 psi
Gas mileage	23 mpg	27 mpg

Pat's test

	Trial 1	**Trial 2**
Roof rack	ON	ON
Tire pressure	24 psi	30 psi
Gas mileage	22 mpg	26 mpg

(a) Whose test BEST shows what happens to the car's gas mileage when the TIRE PRESSURE changes? **(Circle <u>one</u>.)**

Chris's Pat's Both

Explain your answer:

(b) According to the data, **what happens to a car's gas mileage** as a car's **tire pressure** INCREASES? **(Circle <u>one</u>.)**

Increases Decreases Doesn't change

Explain using evidence from the data tables:

(c) What should Chris and Pat do to find out **how the ROOF RACK (ON/OFF) affects the gas mileage?** Complete the table below.

	Trial 1	**Trial 2**
Roof rack		
Tire pressure		

Explain your answer:

Appendix C

Scoring rubrics for pretest items

SOCCER_1

Score	**KI level**	**Description**
0	Blank	No response
1	None	Off task answer No scientific ideas
2	Invalid/ Irrelevant/ Isolated	Associating length of line with speed (longest line) Interpretation of graph as velocity-time (highest value) No reference to slope or distance-time relationship Reference to acceleration
3	Partial	*Incomplete connection between slope and speed* - (positive slope is faster than negative slope) or implication that positive slopes are faster - Reference to larger, greater, or steeper slope/distance without clear attribution to Jim or Ken - Correct interpretation without reference to graph - Jim is faster in D because Ken is at rest
4	Basic	*Complete connection between slope and speed* - Reference to relationship between distance and time (greatest distance per time) - Correct comparison between slopes of Jim and Ken - Jim has the greatest change in position

SOCCER_2

Score	KI level	Description
0	Blank	No response
1	None	Off task answer No scientific ideas
2	Invalid/ Irrelevant/ Isolated	The lines go toward/away from each other Parallel lines go in the same direction “map” interpretation (“upward”) The move straight
3	Partial	*Incomplete connection between slope and direction* Lines have the same slope correct interpretation of motion in segment E without mention of slope (moving forward)
4	Basic	*Complete connection between slope and direction* Lines have the same sign slope Both lines have positive slope Both are increasing their position positions move in the same direction

SMART_ALPHA

Score	KI level	Description
0	Blank	No response
1	None	Off task answer No scientific ideas
2	Invalid/ Irrelevant/ Isolated	Observation that the car slows down gradually No reference to graph slope or shape Reference to the graph of the Beta car Graph decreases gradually
3	Partial	*Incomplete connection between magnitude of slope and speed* - graph is curved
4	Basic	*Complete connection between magnitude of slope and speed* - slope flattens out gradually - slope becomes less and less steep/negative

SMART_BETA

Score	KI level	Description
0	Blank	No response
1	None	Off task answer No scientific ideas
2	Invalid/ Irrelevant/ Isolated	Observation that the car changes speed suddenly No reference to graph slope or shape Reference to the graph of the Alpha car
3	Partial	*Incomplete connection between magnitude of slope and speed* - graph changes suddenly
4	Basic	*Complete connection between magnitude of slope and speed* - slope becomes abruptly flat - slope becomes abruptly less steep/negative

ANIMAL_1

Score	KI level	Description
0	Blank	No response
1	None	Off task answer No scientific ideas
2	Invalid/ Irrelevant/ Isolated	*No connection between function magnitude and speed* - reference to constant speed (interpretation as position graph) - line goes up - reference to changing acceleration
3	Partial	*Incomplete connection between function magnitude and speed* - interpretation of positive slope as increasing velocity or speeding up
4	Basic	*Complete connection between function magnitude and speed* - connection between function become decreasingly negative and slower speed

ANIMAL_2

Score	KI level	Description
0	Blank	No response
1	None	Off task answer No scientific ideas
2	Invalid/ Irrelevant/ Isolated	*No connection between function magnitude and speed* - reference to no movement (interpretation as position graph) - negative velocity is slowing down - lines are straight
3	Partial	*Incomplete connection between function magnitude and speed* - correct interpretation (constant speed) without reference to graph or velocity - zero slope/acceleration/flat - velocity is nonzero - “no change” without specifying what is not changing
4	Basic	*Complete connection between function magnitude and speed* - connection between graph/velocity not changing and constant speed

SKETCH_POSITION

Part 1

Score	KI level	Description
0	Blank	No response
1	None	Off task answer/No scientific ideas not a graph, not a function
2	Invalid/ Irrelevant/ Isolated	horizontal line curve that both increases and decreases
3	Partial	partial connection between graph features and motion - linearly increasing/decreasing graph - gradually decreasing steepness - curve that shows both concave up and concave down sections
4	Basic	full connection between graph features and motion - gradually increasing steepness

Part 2

Score	KI level	Description
0	Blank	No response
1	None	Off task answer/No scientific ideas not a graph, not a function
2	Invalid/ Irrelevant/ Isolated	line that is not horizontal
3	Partial	partial connection between graph features and motion - horizontal line at zero or value other than the end of previous line
4	Basic	full connection between graph features and motion - horizontal line at a value of end of previous line

Part 3

Score	KI level	Description
0	Blank	No response
1	None	Off task answer/No scientific ideas not a graph, not a function
2	Invalid/ Irrelevant/ Isolated	horizontal line curve that both increases and decreases
3	Partial	partial connection between graph features and motion - curve either gradually increases or decreases but not both - linear graph with positive slope - linear graph with negative slope that does not return to original position
4	Basic	full connection between graph features and motion - linear graph with opposite sign slope as part 1 that returns to original position

SKETCH_VELOCITY

Part 1

Score	KI level	Description
0	Blank	No response
1	None	Off task answer No scientific ideas
2	Invalid/ Irrelevant/ Isolated	horizontal line curve that both increases and decreases
3	Partial	partial connection between graph features and motion - curve either increases - linear function that at some point is negative
4	Basic	full connection between graph features and motion - increasing linear function

Part 2

Score	KI level	Description
0	Blank	No response
1	None	Off task answer No scientific ideas
2	Invalid/ Irrelevant/ Isolated	any increasing or decreasing function
3	Partial	partial connection between graph features and motion - horizontal, non-zero function
4	Basic	full connection between graph features and motion - horizontal function equal to zero throughout

Part 3

Score	KI level	Description
0	Blank	No response
1	None	Off task answer No scientific ideas
2	Invalid/ Irrelevant/ Isolated	non-horizontal line any curved function
3	Partial	partial connection between graph features and motion - horizontal line at positive or zero value
4	Basic	full connection between graph features and motion - horizontal line at negative value

EXPERIMENT_1

Score	**KI level**	**Description**
0	Blank	No response
1	None	Off task answer No scientific ideas
2	Invalid/ Irrelevant/ Isolated	*No connection between changing variables and valid tests* - mileage decreased without explanation - cites Rosa's test as valid, even in addition to Amanda's test (no distinction between the two tests)
3	Partial	*Partial connection between changing variables and valid tests* - Bigger change in speed for Amanda - Reference to Amanda's "constant" test without reference to specific variables
4	Basic	*Complete connection between changing variables and valid tests* Connection between Amanda's test being better than Rosa's test and ONE of the following: - Explanation that Amanda's test is valid (e.g. held windows constant, changed only one thing) - Explanation that Rosa's test is invalid (e.g. changed the windows, changed both variables)
5	Complex	*Complex connection between changing variables and valid tests* Connection between Amanda's test being better than Rosa's test and BOTH of the following: - Explanation that Amanda's test is valid (e.g. held windows constant, changed only one thing) - Explanation that Rosa's test is invalid (e.g. changed the windows, changed both variables)

EXPERIMENT_2

Score	**KI level**	**Description**
0	Blank	No response
1	None	Off task answer No scientific ideas
2	Invalid/ Irrelevant/ Isolated	*No connection between variables and valid tests* - Does not cite data - Cites Rosa’s data, even in addition to Amanda’s data (no distinction between the two tests)
3	Partial	*Incomplete connection between changing variables and valid tests* - Cites only Amanda’s data, but without explanation
4	Basic	*Complete connection between data and valid tests* Connects data from Amanda’s test and ONE of the following: - Amanda held windows constant/changed one thing between trials - States that Rosa’s test was invalid - Amanda’s test ensures the effect on mileage is from only the tire pressure
5	Complex	*Complete connection between data and valid tests* Connects data from Amanda’s test and at least TWO of the following: - Amanda held windows constant/changed one thing between trials - States that Rosa’s test was invalid - Amanda’s test ensures the effect on mileage is from only the tire pressure

EXPERIMENT_3

Score	**KI level**	**Description**
0	Blank	No response
1	None	Off task answer No scientific ideas
2	Invalid/ Irrelevant/ Isolated	*No connection between variables and valid tests* - Invalid experimental design - Valid experimental design without explanation
3	Partial	*Incomplete connection between changing variables and valid tests* Connects valid experimental design with ONE of the following - Must change the windows between trials
4	Basic	*Complete connection between changing variables and valid tests* Connects valid experimental design with ONE of the following - Must keep speed constant between trials - Don’t want speed to influence the outcome/want to be sure it’s only the windows that affect the outcome
5	Complex	*Complex connection between changing variables and valid tests* Connects valid experimental design with BOTH of the following - Must keep speed constant between trials - Don’t want speed to influence the outcome/want to be sure it’s only the windows that affect the outcome

COLLISION

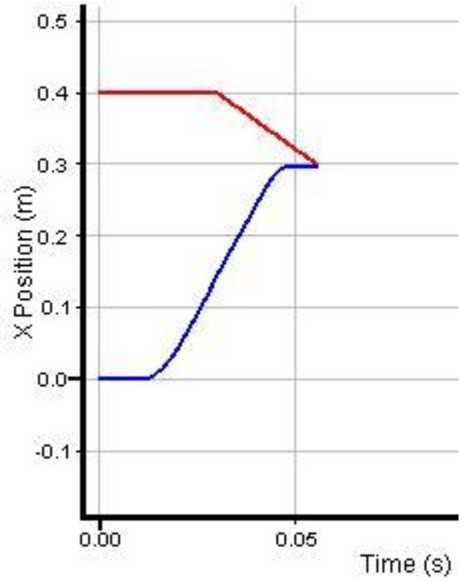
Score	KI level	Description
0	Blank	No response
1	None	Off task answer No scientific ideas
2	Invalid/ Irrelevant/ Isolated	The woman is weaker/lighter/worse driver
3	Partial	*Incomplete connection between height of person and risk for injury* - The woman is taller - no connection to how the drivers are seated
4	Basic	Connection between height of drivers and distance seated from the steering wheel OR Connection between distance seated from steering wheel and being in the path of the airbag as it deploys
5	Complex	Connection between height of drivers and distance seated from the steering wheel AND Connection between distance seated from steering wheel and being in the path of the airbag as it deploys

Appendix D

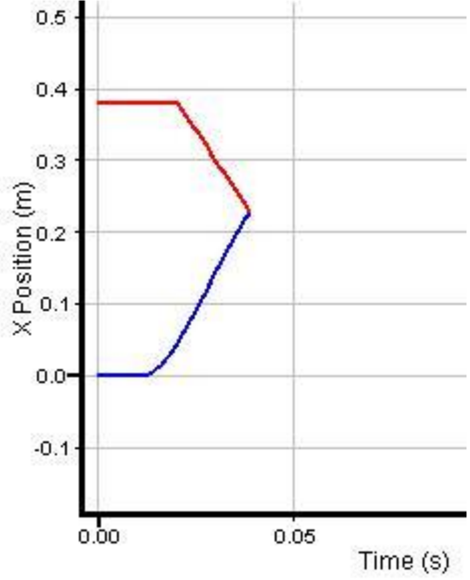
Embedded assessments for Study 1

Study 1: INTERPRETATIONS

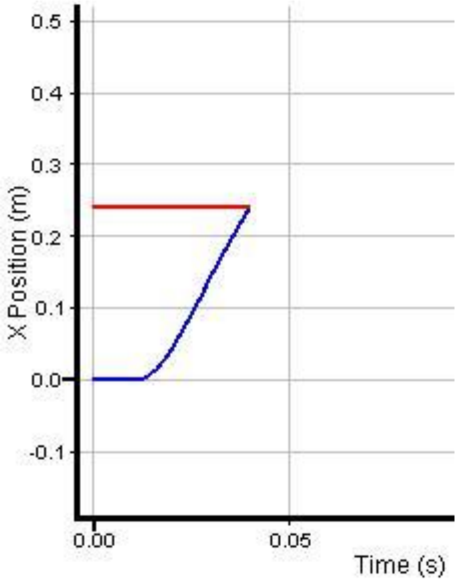
[INTERPRETATIONS_1] Describe what happened between the driver an airbag in this crash. Was the driver injured by the airbag? Explain based on the graph.



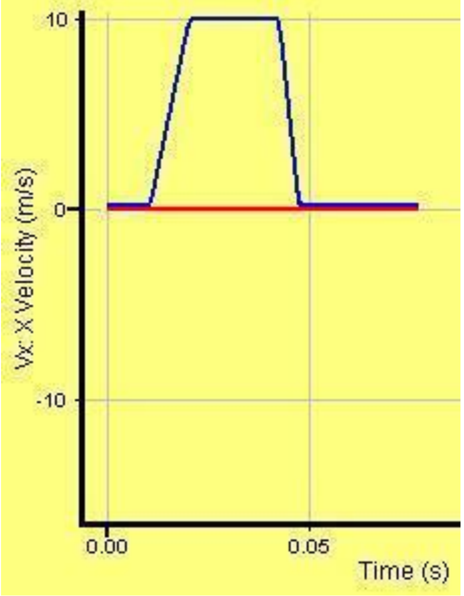
[INTERPRETATIONS_2] Describe what happened between the driver an airbag in this crash. Was the driver injured by the airbag? Explain based on the graph.



[INTERPRETATIONS_3] Describe what happened between the driver an airbag in this crash. Was the driver injured by the airbag? Explain based on the graph.

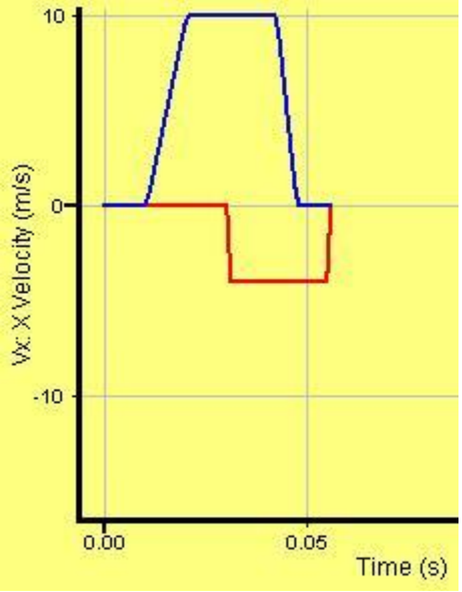


[INTERPRETATIONS_4] Describe what happened between the driver an airbag in this crash. Was the driver injured by the airbag? Explain based on the graph.



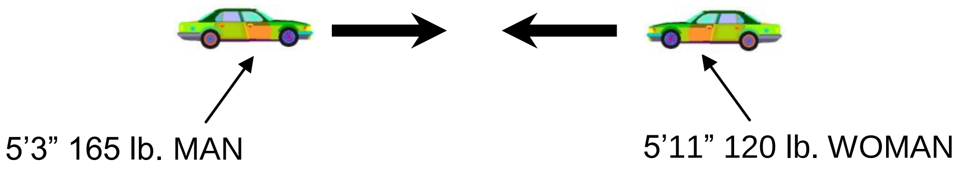
[INTERPRETATIONS_5] Describe what happened between the driver an airbag in this crash. Was the driver injured by the airbag? Explain based on the graph.

[INTERPRETATIONS_6] Describe what happened between the driver an airbag in this crash. Was the driver injured by the airbag? Explain based on the graph.



Study 2 & 3: INTERPRETATIONS

[HEADON_1] 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.



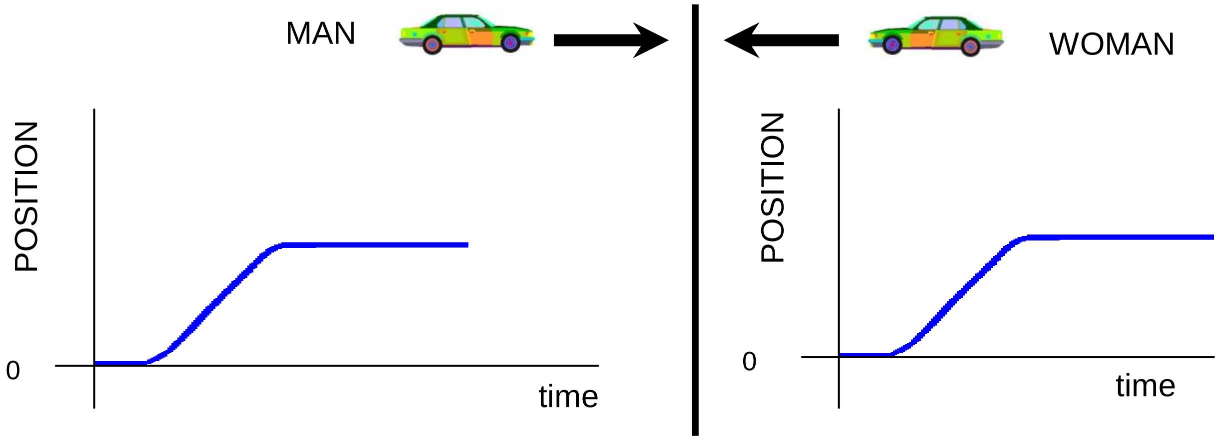
Which person do you think is **more likely to be injured** by an **airbag deploying**?
(Circle one.)

The MAN on the left

The WOMAN on the right

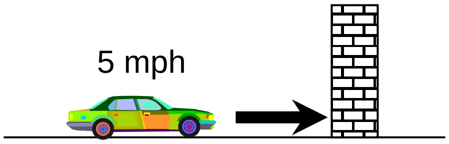
Explain your answer:

[HEADON_2] Draw the position graphs of the **driver** of each car to illustrate your choice. The airbag’s graph is provided.



Explain how your graphs illustrate your choice:

Two **identical** cars crash into a solid barrier, one at 5 mph and the other at 15 mph. Airbags in both cars deploy. The drivers in each car are men of the same height and build.



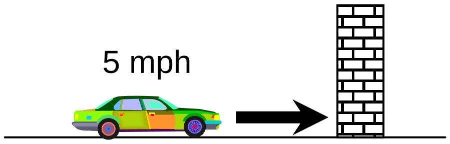
[WALL_1] Which person do you think is **<u>more likely to be injured</u>** by an **<u>airbag deploying</u>**? **(Circle <u>one</u>.)**

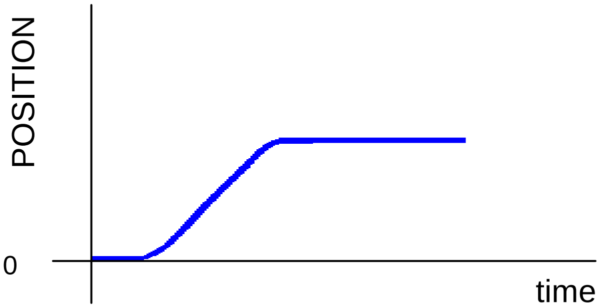
The driver on the left

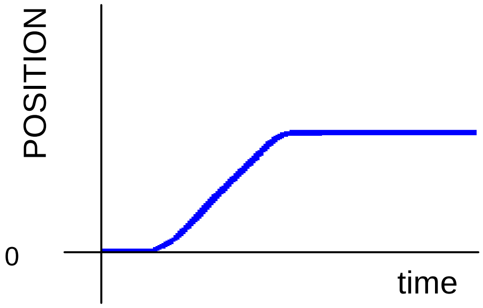
The driver on the right

Explain your answer:

[WALL_2] Draw the position graphs of the **driver** of each car to illustrate your choice. The airbag's graph is provided.



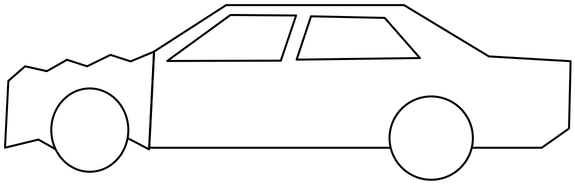




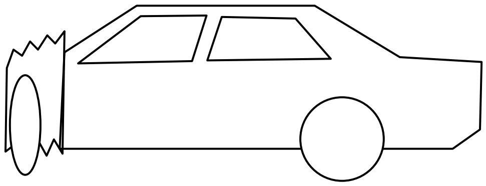
Explain how your graphs illustrate your choice:

A car manufacturing company tests two different designs for the front of a new car. Both designs are crashed head-on into a barrier at 30 mph, deploying the airbags. **The <u>front sections</u> of the two cars crumpled by <u>different amounts</u>, but the <u>passenger cabins</u> of both cars <u>remain undamaged.</u>**

[CRUMPLE_1] In which car is the **AIRBAG <u>more likely to injure the driver</u>** in a head-on crash? **(Circle <u>one</u>.)**



The car that crumpled LESS

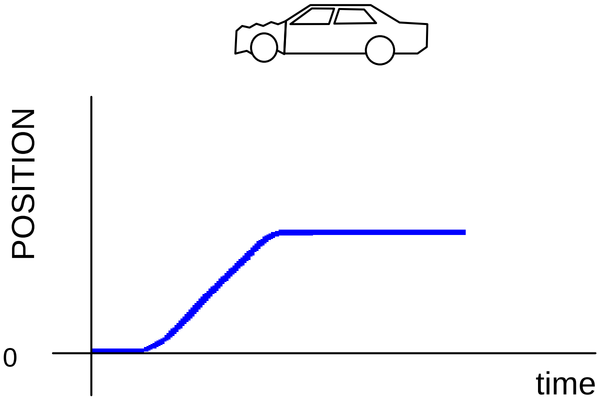


The car that crumpled MORE

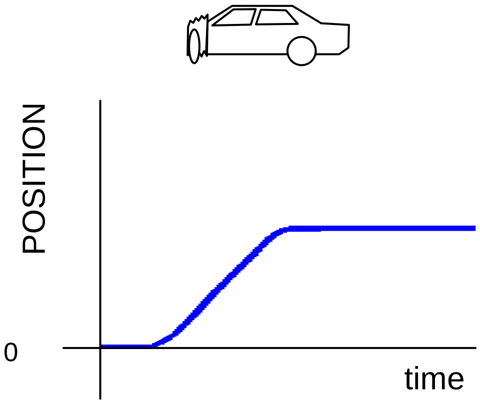
Explain your answer:

[CRUMPLE_2] Draw the position graphs of the **driver** of each car to illustrate your choice. The airbag's graph is provided.

crumpled LESS



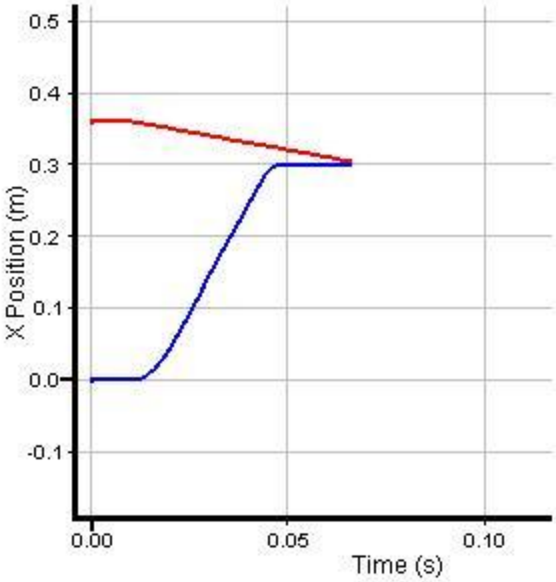
crumpled MORE



Explain how your graphs illustrate your choice:

Study 3: INTERPRETATIONS

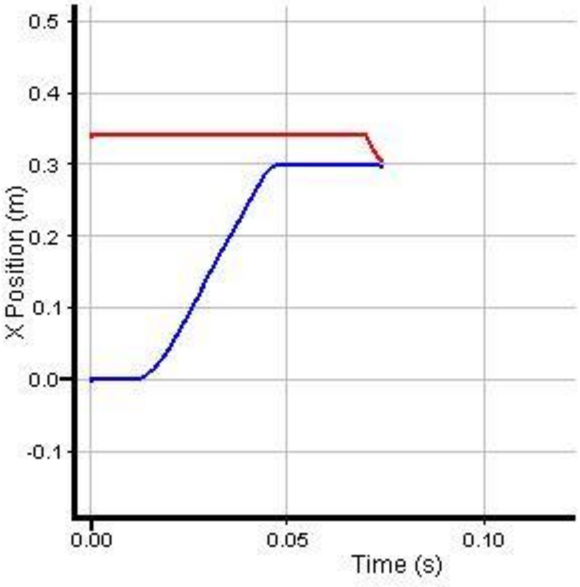
A "black box" data recorder produced this position-time graph of the passenger and airbag in a crash.



[INTERPRETATIONS_7] Was this crash SAFE or UNSAFE? Explain your choice.

[INTERPRETATIONS_8] Was the DRIVER HEIGHT, COLLISION SPEED, or CAR CRUMPLING most responsible for the result? Explain your choice.

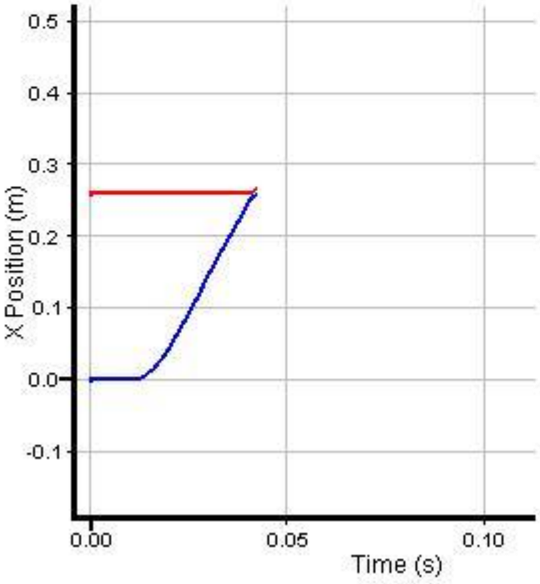
A "black box" data recorder produced this position-time graph of the passenger and airbag in a crash.



[INTERPRETATIONS_9] Was this crash SAFE or UNSAFE? Explain your choice.

[INTERPRETATIONS_10] Was the DRIVER HEIGHT, COLLISION SPEED, or CAR CRUMPLING most responsible for the result? Explain your choice.

A "black box" data recorder produced this position-time graph of the passenger and airbag in a crash.



[INTERPRETATIONS_11] Was this crash SAFE or UNSAFE? Explain your choice.

[INTERPRETATIONS_12] Was the DRIVER HEIGHT, COLLISION SPEED, or CAR CRUMPLING most responsible for the result? Explain your choice.

Appendix E

Embedded assessment rubrics

INTERPRETATIONS_1

Score	KI level	Description
Blank	0	No response
None	1	Off task answer
Isolated/ Invalid/ Irrelevant	2	description of outcome only (safe/unsafe)
Partial	3	correct description of process only (completion of airbag deployment)
Basic	4	connection between outcome (safe) and completion of airbag deployment
Complex	5	connection among outcome (safe), completion of airbag deployment, and driver motion attributes (distance, velocity, or time)

INTERPRETATIONS_2

Score	KI level	Description
Blank	0	No response
None	1	Off task answer
Isolated/ Invalid/ Irrelevant	2	description of outcome only (safe/unsafe)
Partial	3	correct description of process only (incomplete airbag deployment)
Basic	4	connection between outcome (unsafe) and incomplete airbag deployment
Complex	5	connection among outcome (unsafe), incomplete airbag deployment, and driver motion attributes (distance, velocity, or time)

INTERPRETATIONS_3

Score	**KI level**	**Description**
Blank	0	No response
None	1	Off task answer
Isolated/ Invalid/ Irrelevant	2	description of outcome only (safe/unsafe)
Partial	3	correct description of process only (incomplete airbag deployment)
Basic	4	connection between outcome (unsafe) and incomplete airbag deployment
Complex	5	connection among outcome (unsafe), incomplete airbag deployment, and driver motion attributes (distance)

INTERPRETATIONS_4

Score	**KI level**	**Description**
Blank	0	No response
None	1	Off task answer
Isolated/ Invalid/ Irrelevant	2	description of outcome only (safe/unsafe)
Partial	3	correct description of process only (airbag deployment)
Basic	4	connection between outcome (safe/unsafe) and incomplete airbag deployment
Complex	5	connection among outcome (safe/unsafe), airbag deployment, and driver motion attributes (did not move)

INTERPRETATIONS_5

Score	KI level	Description
Blank	0	No response
None	1	Off task answer
Isolated/ Invalid/ Irrelevant	2	description of outcome only (unsafe)
Partial	3	correct description of process only (incomplete airbag deployment)
Basic	4	connection between outcome (unsafe) and incomplete airbag deployment
Complex	5	connection among outcome (unsafe), incomplete airbag deployment, and driver motion attributes (driver and airbag stopped simultaneously)

INTERPRETATIONS_6

Score	KI level	Description
Blank	0	No response
None	1	Off task answer
Isolated/ Invalid/ Irrelevant	2	description of outcome only (unsafe)
Partial	3	correct description of process only (complete airbag deployment)
Basic	4	connection between outcome (safe) and complete airbag deployment
Complex	5	connection among outcome (safe), complete airbag deployment, and driver motion attributes (driver stopped after airbag finished deploying)

HEADON_1

Score	KI level	Description
0	Blank	No response
1	None	Off task answer No scientific ideas
2	Invalid/ Irrelevant/ Isolated	Man has stronger build
3	Partial	*Incomplete connection between height of person and risk for injury* One of the following - The woman is taller/man is shorter - Woman sits further/man sits closer to the steering wheel - Man encounters airbag before fully inflating
4	Basic	Connection between height of drivers and distance seated from the steering wheel OR Connection between distance seated from steering wheel and being in the path of the airbag as it deploys
5	Complex	Connection between height of drivers and distance seated from the steering wheel AND Connection between distance seated from steering wheel and being in the path of the airbag as it deploys

HEADON_2

Score	KI level	Description
0	Blank	No response
1	None	incomplete graphs
2	Invalid/ Irrelevant/ Isolated	no correct distinctions between graphs
3	Partial	*Incomplete distinction between graphs/cases* - woman's graph is safe, man's graph is unsafe OR - woman's starting position is clearly indicated higher OR - explanation addresses man's closer starting position/higher graph
4	Basic	*Complete distinction between graphs/cases* - woman's graph is safe, man's graph is unsafe AND - woman's starting position is clearly indicated higher OR - explanation addresses man's closer starting position/higher graph
5	Complex	*Complete distinction between graphs/cases* - woman's graph is safe, man's graph is unsafe AND - woman's starting position is clearly indicated higher AND - explanation addresses man's closer starting position/higher graph

WALL_1

Score	KI level	Description
0	Blank	No response
1	None	Off task answer No scientific ideas
2	Invalid/ Irrelevant/ Isolated	Car on the left is slower and has greater risk
3	Partial	*Incomplete connection between driving speed and risk for injury* One of the following - Car on the right is going faster - Driver on the right travels faster with respect to airbag - Driver on right encounters airbag before fully inflating
4	Basic	Connection between driving speed and driver speed with respect to airbag OR Connection between driver speed with respect to airbag and being in the path of the airbag as it deploys
5	Complex	Connection between driving speed and driver speed with respect to airbag AND Connection between driver speed with respect to airbag and being in the path of the airbag as it deploys

WALL_2

Score	KI level	Description
0	Blank	No response
1	None	incomplete graphs
2	Invalid/ Irrelevant/ Isolated	no correct distinctions between graphs
3	Partial	*Incomplete distinction between graphs/cases* - slow graph is safe, fast is unsafe OR - right graph is steeper OR - explanation addresses driver's faster speed with respect to airbag
4	Basic	*Complete distinction between graphs/cases* - slow graph is safe, fast graph is unsafe AND - slow slope is flatter OR - explanation addresses graph slope or driver velocity with respect to airbag
5	Complex	*Complete distinction between graphs/cases* - slow graph is safe, fast graph is unsafe AND - slow slope is flatter AND - explanation addresses graph slope or driver velocity with respect to airbag

CRUMPLE_1

Score	**KI level**	**Description**
0	Blank	No response
1	None	Off task answer No scientific ideas
2	Invalid/ Irrelevant/ Isolated	Less crumpled car protects the driver better airbag comes out faster in the crumpled car there is more impact/damage to the crumpled car Crumpling absorbs the force of impact
3	Partial	*Incomplete connection between crumpling and airbag/driver collision* One of the following - Crumpling absorbs the force of impact - The airbag has more time to fully inflate if the car crumples more - The airbag can injure drivers who are near the steering wheel during deployment - Less crumpled car stops more rapidly
4	Basic	*Full connection between crumpling and airbag/driver collision* Connection between car crumpling and delay time OR Connection between delay time and being in the path of the airbag as it deploys
5	Complex	*Complex connection between crumpling and airbag/driver collision* Connection between car crumpling and delay time OR Connection between delay time and being in the path of the airbag as it deploys

CRUMPLE_2

Score	KI level	Description
0	Blank	No response
1	None	incomplete graphs
2	Invalid/ Irrelevant/ Isolated	no correct distinctions between graphs
3	Partial	*Incomplete distinction between graphs/cases* - more crumpled graph is safe, less crumpled graph is unsafe OR - more crumpled has longer delay time OR - explanation addresses delay time
4	Basic	*Complete distinction between graphs/cases* - more crumpled graph is safe, less crumpled graph is unsafe AND - more crumpled has longer delay time OR - explanation addresses delay time
5	Complex	*Complex distinction between graphs/cases* - more crumpled graph is safe, less crumpled graph is unsafe - more crumpled has longer delay time AND - explanation addresses delay time

INTERPRETATIONS_7

Score	KI level	Description
0	Blank	No response
1	None	Off task answer
2	Isolated/ Invalid/ Irrelevant	description of outcome only (safe/unsafe)
3	Partial	Partial connection between outcome and graph - Driver was safe because the airbag deployed fully/airbag was stationary at time of collision
4	Basic	Exactly one connection between two of the following ideas - complete deployment of airbag - airbag was stationary at time of collision - time indices for collision events - low driver velocity with respect to steering wheel/low car velocity prior to impact - sufficient driver height/initial distance to steering wheel - sufficient crumpling/delay time
5	Complex	At least two connections among the following ideas - complete deployment of airbag - airbag was stationary at time of collision - time indices for collision events - low driver velocity with respect to steering wheel/low car velocity prior to impact - sufficient driver height/initial distance to steering wheel - sufficient crumpling/delay time

INTERPRETATIONS_8

Score	**KI level**	**Description**
0	Blank	No response
1	None	Off task answer
2	Isolated/ Invalid/ Irrelevant	No relevant scientific ideas
3	Partial	One of the following ideas without connection to others - Airbag deployed fully - Driver was slow approaching the airbag - Collision speed was slow - Driver was sufficiently far from the steering wheel to allow airbag to deploy - Driver was not particularly tall - Short delay time - Little crumpling
4	Basic	One full connection between two of the following ideas: - Airbag deployed fully - Driver was slow approaching the airbag - Collision speed was slow - Driver was sufficiently far from the steering wheel to allow airbag to deploy - Driver was not particularly tall - Short delay time - Little crumpling
5	Complex	At least two full connections among the following ideas: - Airbag deployed fully - Driver was slow approaching the airbag - Collision speed was slow - Driver was sufficiently far from the steering wheel to allow airbag to deploy - Driver was not particularly tall - Short delay time - Little crumpling

INTERPRETATIONS_9

Score	KI level	Description
0	Blank	No response
1	None	Off task answer
2	Isolated/ Invalid/ Irrelevant	description of outcome only (safe/unsafe)
3	Partial	Partial connection between outcome and graph - Driver was safe because the airbag deployed fully
4	Basic	Exactly one connection between two of the following ideas - complete deployment of airbag - airbag was stationary at time of collision - time indices for collision events - long delay time/car crumpling - sufficient driver height/initial distance to steering wheel - sufficient crumpling/delay time
5	Complex	At least two connections among the following ideas - complete deployment of airbag - airbag was stationary at time of collision - time indices for collision events - long delay time/car crumpling - sufficient driver height/initial distance to steering wheel - sufficient crumpling/delay time

INTERPRETATIONS_10

Score	KI level	Description
0	Blank	No response
1	None	Off task answer
2	Isolated/ Invalid/ Irrelevant	No relevant scientific ideas
3	Partial	One of the following ideas without connection to others - Airbag deployed fully - Long delay time - Car crumpled a lot - Driver was sufficiently far from the steering wheel to allow airbag to deploy - Driver was not particularly tall - High velocity - High collision speed
4	Basic	One full connection between two of the following ideas: - Airbag deployed fully - Long delay time - Car crumpled a lot - Driver was sufficiently far from the steering wheel to allow airbag to deploy - Driver was not particularly tall - High velocity - High collision speed
5	Complex	At least two full connections among the following ideas: - Airbag deployed fully - Long delay time - Car crumpled a lot - Driver was sufficiently far from the steering wheel to allow airbag to deploy - Driver was not particularly tall - High velocity - High collision speed

INTERPRETATIONS_11

Score	KI level	Description
0	Blank	No response
1	None	Off task answer
2	Isolated/ Invalid/ Irrelevant	description of outcome only (safe/unsafe)
3	Partial	Partial connection between outcome and graph - Driver was unsafe because the airbag did not deploy fully
4	Basic	Exactly one connection between two of the following ideas - incomplete deployment of airbag - airbag was moving at time of collision - time indices for collision events - short driver/short initial distance between driver and steering wheel - delay time/collision speed unimportant to outcome
5	Complex	At least two connections among the following ideas - incomplete deployment of airbag - airbag was moving at time of collision - time indices for collision events - short driver/short initial distance between driver and steering wheel - delay time/collision speed unimportant to outcome

INTERPRETATIONS_12

Score	KI level	Description
0	Blank	No response
1	None	Off task answer
2	Isolated/ Invalid/ Irrelevant	No relevant scientific ideas
3	Partial	One of the following ideas without connection to others - Airbag did not deploy fully - Driver was insufficiently far from the steering wheel to allow airbag to deploy - Driver was too short - Delay time and collision speed were unimportant to the outcome
4	Basic	One full connection between two of the following ideas: - Airbag did not deploy fully - Driver was insufficiently far from the steering wheel to allow airbag to deploy - Driver was too short - Delay time and collision speed were unimportant to the outcome
5	Complex	At least two full connections among the following ideas: - Airbag did not deploy fully - Driver was insufficiently far from the steering wheel to allow airbag to deploy - Driver was too short - Delay time and collision speed were unimportant to the outcome