Investigating the Mechanisms of Learning from a Constrained Preparation for Future Learning Activity

Stephanie A. Siler[1], David Klahr[1], and Norman Price[2]

[1]Carnegie Mellon University

[2]University of Massachusetts-Amherst

Author Note

Stephanie A. Siler, Psychology Department, Carnegie Mellon University,

5000 Forbes Avenue, Pittsburgh, PA 15213; siler@cmu.edu; 412-268-6136;

David Klahr, Psychology Department, Carnegie Mellon University,

5000 Forbes Avenue, Pittsburgh, PA 15213;

Norman Price, School of Education and Scientific Reasoning Research Institute,

University of Massachusetts-Amherst, Amherst, MA 01003.

Abstract

Many studies have shown benefits associated with engaging students in problem-solving activities prior to administering lessons. These problem-solving activities are assumed to activate relevant knowledge and allow students to develop some initial knowledge structures, which support understanding of the lesson. In this paper we report the results of two studies in which we investigated the underlying benefits of engaging in a preparatory activity—setting up experiments without running them or receiving feedback—prior to an interactive computerized lesson on experimental design compared to only engaging in the interactive lesson. We predicted that the seventh-grade participants who demonstrated some initial knowledge of the topic—experimental design—would benefit more from spending the whole time engaged in instructional activities. However, we expected students who did not demonstrate initial knowledge would benefit more from engaging in the preparatory activity, which would allow them to activate or develop initial knowledge that would aid their understanding of the subsequent instruction. The predicted condition by initial knowledge interaction was found in both studies. In Study 1, the benefit of only engaging in the instruction was found only for the lowest-knowledge of students who demonstrated initial knowledge. For students who did not demonstrate some initial knowledge, the benefit of completing the preparatory activity appeared to be due to the development of an understanding of the general goal of the activity rather than of specific knowledge of experimental design. Based on this finding, in Study 2, we investigated an initial *goal* by condition interaction. In fact, students who did not express an understanding of the task goal on the pretest benefited from engaging in the preparatory activity and students who did benefited more from the instruction. Again, this benefit appeared to be due to students' development of an appropriate understanding of the task goal during the preparatory activity.

Keywords: Preparation for future learning; middle-school students; transfer, experimental design instruction.

Investigating the Mechanisms of Learning

from a Constrained Preparation for Future Learning Activity

Several recent studies have shown benefits associated with engaging students in problem-solving activities prior to administering lessons on a topic (e.g., Schwartz & Martin, 2004; Kapur, 2010; Lorch et al., 2010). The explanation for the effectiveness of these initial problem-solving activities is that they provide a framework to support understanding of the subsequent lesson by (a) activating relevant prior knowledge, and (b) facilitating the construction of preliminary knowledge structures that are relevant to the larger instructional goals.

The characteristics of the pre-lesson problem-solving activities that promote greater understanding of the lesson are worth investigating. Some studies (e.g., Schwartz & Martin, 2004) utilized invention activities *prior* to the problem-solving activities that were designed so as to support initial student knowledge development. These activities included use of contrasting cases to prompt students to consider different aspects of the problem. All students participated in these initial activities, after which they were either (a) tasked with working in groups to invent solutions to a novel problem but were given no feedback from their instructor or (b) told the solution and practiced applying it. Students in the invention condition out-performed those in the "tell and practice" condition, but only when a learning resource—a worked example—was available on the posttest. Thus, the effect of the invention activity appeared to be that it enabled students to develop a sufficient understanding of some essential aspects of the problem that allowed them to "recognize the value of a solution once it becomes available" (p. 162), whereas the tell and practice students were unable to do so. The characteristics of the problem-solving activity for this and other studies discussed in this paper are shown in Table 1.

Kapur (2010) found that initial supportive activities such as those employing contrasting cases prior to problem solving may not be necessary for preparing students to learn from a subsequent lesson. That is, attempting to solve complex problems alone may be sufficient for producing learning gains. In one condition, seventh-grade students worked in groups attempting to solve complex problems involving speed/distance/time relationships and then worked individually on “extension” problems. These students received no additional support or scaffolding and no feedback from the instructors during problem solving. Although these students failed to produce correct answers to these complex problems, their later posttest performance and performance using structured-response scaffolds to solve more advanced problems was better than that of students who experienced a more traditional lecture-then-practice format. Kapur proposed two related mechanisms responsible for these learning gains: generation of inter-connected knowledge structures and knowledge differentiation.

These two studies indicate that unguided complex problem-solving activities—even those that fail to produce solutions—can promote learning. However, more recently, Lorch et al. (2010) found a benefit of initial engagement in a less complex, more constrained group problem-solving activity. In their study, fourth-grade students from both higher- and lower-achieving schools learned a procedure for designing simple experimental contrasts called “the control of variables strategy” (CVS). The core idea in CVS is to vary *only* the variable being tested—the focal variable—while controlling all other variables. Lorch et al. used three instructional conditions: (a) instruction via an interactive classroom lecture (the “instruct” condition) in which the teacher led discussions of whether presented experimental designs were good (unconfounded) experiments and why they were or were not good, (b) instruction in which groups of students designed and executed experiments (the “manipulate” condition), (c) instruction in which students first worked in groups designing and running experiments and then participated in the interactive lecture (the “both” condition).

Students in the “both” condition out-performed students in the other two conditions. However, students from higher-achieving schools benefitted more from completing the manipulate activity prior to the lesson in the both condition than their counter-parts from lower-achieving schools. However, in both the higher- and lower-achieving schools, students in the “instruct” condition out-performed students in the “manipulate” condition, showing that students need some instructional support to learn these skills. Lorch et al. also found that different aspects of understanding experimental design tended to be supported in the manipulation and lecture phases. Specifically, the manipulate task supported development of the more intuitive idea of contrasting variables, whereas the lecture portion supported understanding of the more challenging aspect of CVS: why it is necessary to control all non-focal variables. One question that arises with respect to these results is whether students' development of an understanding of the need to compare and contrast at least the focal variable *aided* in their comprehension of the subsequent lesson.

However, one limitation of the Lorch et al. (2010) study they acknowledged was that the instruction procedure was not shortened in the "both" condition. Thus, it is possible that the advantage of the "both" condition over the "instruct" condition was due to increased time on task rather than what students had learned during the initial preparatory activity, per se.

In the current studies, we attempted to replicate the contrast between the "instruct" and "both" conditions used in the Lorch et al. (2010) study in the domain of CVS. However, rather than engaging in teacher-led instruction, students individually worked with a computer tutor, "TED" (for Training in Experimental Design), that provided interactive instruction similar to that given in Lorch et al. This instruction was based on the method developed by Klahr and colleagues (e.g., Chen & Klahr, 1999; Toth, Klahr, & Chen, 2000; Klahr & Nigam, 2004; Strand-Cary & Klahr, 2008) and found to promote elementary and middle-school children's transfer of CVS over long periods of time and to different domains.

To control for time on task in the current studies, while students in the "Both" condition designed an experiment for each of the ramps variables without receiving feedback, students in the other condition completed one round of instruction in which they evaluated experiments and subsequently received feedback and explanations. Though this introduced a confound of task (design vs. evaluate) along with feedback, we believed the more generative design task was better-suited for supporting the initial knowledge development we hypothesized would aid students in subsequent instruction. Thus, we considered this a more appropriate comparison condition. Therefore, the current studies address the differential effects of either completing a feedback-free generative activity or engaging in an evaluative instructional activity in which feedback was provided, before engaging in the interactive instruction.

Because student-level information was available in the current studies, we were able to determine students' initial CVS knowledge and trace the development of that knowledge during the preparatory activity and beyond. In addition, we were able to investigate aptitude- and knowledge-by-treatment interactions in order to better understand the mechanisms of learning from the preparatory activity. Such information is also helpful in informing the development of our computer tutor—in particular, in deciding how to adapt instruction on CVS based on student-level information. Such adaptation is especially important for lower-ability students, who do not benefit from this method of instruction as much as their higher-ability counter-parts (e.g., Klahr & Li, 2005).

Table 1. Summary of characteristics of preparation for future learning activities.

			Characteristics of pre-lesson preparatory activity							
Study	Student Grade	Domain	Students have some initial knowledge of domain?	*Pre*-preparatory supportive activity?	Pair/Group or Individual p-solving?	Feedback from teacher?	Feedback from other students?	Exposed to other explanations?	Complex or Constrained problem?	Scaffolded p-solving?
Schwartz & Martin (2004)	9	Statistics	No	Y (contrasting cases)	Group	--	Maybe	Likely	Complex	--
Kapur (2009)	7	Math	Yes	--	Group, then Individual	--	Maybe	Likely	Complex	Yes
Lorch et al. (2010)	4	Science	(Likely some students do)	--	Group	--	Maybe	Likely	Constrained	Yes (filling out table)
Current studies	7	Science	(Some students do)	--	Individual	--	--	Y (multiple-choice responses)	Constrained	Yes (filling out table)

To determine what type of knowledge, developed during the preparatory activity, might benefit students' learning gains from instruction, we referred to Lorch et al. (2010), where students tended to gain an understanding of the need to vary the focal variable during the manipulate task. Therefore, we predicted a knowledge by treatment interaction, where students who did not have an initial understanding of this aspect of CVS (as demonstrated on a pretest) would benefit more from setting up experiments and explaining their designs prior to instruction. We predicted that the benefit of this activity would come from the development of some initial understanding of the nature of the task—that is, that the task is learning how to design experiments—and some initial concepts, in particular the idea of "comparing and contrasting" variables across conditions. Developing this knowledge prior to instruction may reduce the cognitive load students experience during the lesson (cf. Sweller, 1988), better enabling them to learn the primary point of the lesson—the underlying rationale for controlling variables.

However, pre-instructional problem solving without feedback may actually be detrimental to learning in some circumstances. For example, in the domain of experimental design, we have found that students may misinterpret the goal of the instruction (Siler & Klahr, in press) as being something other than learning how to identify causal variables. One such misinterpretation is viewing the task within the framework of an engineering goal (Schauble et al., 1991; Siler & Klahr, in press) rather than designing experiments that allow one to find out about the effect of a particular variable. With an engineering goal, students attempt to set up experiments in order to produce a desired outcome. Typically, students attempt to produce a maximal outcome such as designing ramps that make the balls roll fastest.

In contrast, we expected that students whose pretest responses indicated that they already understood the need to contrast the focal variable would benefit more from the extra instruction, which focuses on the rationale for controlling the other variables, than from setting up experiments on the ramps pretest. This is because such students should experience reduced cognitive load and consequently have more available resources with which to learn the rationale for controlling variables during instruction. Furthermore, expression of an understanding of the underlying logic of CVS has been found to promote far-transfer performance (e.g., Siler et al., 2010).

In the current studies, we tested these predictions (summarized in Table 2) and investigated the mechanisms of learning from engaging in preparatory activities by determining what was most predictive of their posttest scores. Students completed both the preparatory activity and interactive instruction using a computer tutor that provides instruction in experimental design, the TED tutor. In order to minimize the potential negative effects of poor reading skills, instruction in the TED tutor includes audio voice-overs. Throughout all phases of the study, student actions were recorded and saved in log files, which were later analyzed.

Table 2. Predicted relative performance under the two instructional conditions

Knowledge of contrasting the focal variable?	Predicted performance
Yes	Instruction-only > Both
No	Both > Instruction-only

Study 1 Methods

Participants

Participants were 142 seventh-grade students from ten science classes taught by three teachers in a suburban middle school in Massachusetts. In this school, 24% of students qualified for free or reduced-price lunch. Thus, the majority of students in this school were from middle to high-SES backgrounds. As assessed by the Massachusetts Comprehensive Assessment System (MCAS), participants' English Language Arts proficiency rate (59%) was similar to the overall state average of 61%. However, participants' Math proficiency rate (47%) was somewhat higher than the state average of 39%. MCAS science proficiency rates were not available.

Students were randomly assigned to condition within each class. More than half of the students (n = 78, or 55%) demonstrated initial mastery of CVS on a pretest (i.e., they scored at least six out of nine) and were excluded from further analyses. The high rate of pre-instructional mastery strongly suggests that students in this sample had some experience in science inquiry and/or instruction in experimental design. Students who did not master the story pretest had significantly lower standardized math scores than those who did (M = 244.35, SD = 18.50; M = 257.21, SD = 14.01, respectively), $F(1, 127) = 20.07$, $p < .001$, as well as significantly lower standardized science scores (M = 239.25, SD = 16.96; M = 251.86, SD = 13.34, respectively), $F(1, 123) = 21.66$, $p < .001$.[1]

Therefore, the students who did not show CVS mastery on the story pretest were lower-performing than the general student population at this school. These students who did not show mastery on the story pretest were on average at the "proficient" level in math (See Table 3)[2], defined by the Massachusetts Department of Elementary and Secondary Education (MDESE) (2011) as students who "demonstrate a solid understanding of challenging subject matter and solve a wide variety of problems." These students performed at a level below "proficient" but above "needs improvement" in science, where students at the "needs improvement" level are defined by MDESE as those who "demonstrate a partial understanding of subject matter and solve some simple problems."

[1] There were no significant differences in standardized test scores between students of different teachers.

[2] From Massachusetts Department of Elementary & Secondary Education's "Spring 2011 MCAS Tests: Summary of State Results."

Table 3: MCAS Scaled Score Ranges[2]

Scaled Score Range	Performance Level
260–280	*Advanced*
240–258	*Proficient*
220–238	*Needs Improvement*
200–218	*Warning / Failing*

Fourteen students (seven per condition) had technical difficulties while using the computer tutor (e.g., they accidentally quit the program, requiring restarting and repeating instruction or missing segments of the instruction), and two Both condition students did not take the posttest. Data from the remaining 48 students (27 in the Instruction-only and 21 in the Both condition) were included in the following analyses. Of just these students, the mean MCAS science and math scores were at the "Needs improvement" and "Proficient" levels, respectively (M = 237.35, SD = 15.43; M = 244.21, SD = 18.26, respectively).

Materials and Procedure

As shown in Table 4, all students first completed the computerized "story" pretest, which served as the general measure of incoming CVS knowledge. This pretest consisted of six questions in three different domains (i.e., drink sales, rocket design, and cookie baking); for each domain, students first designed then evaluated an experiment for given focal variables (see Figs 1 & 2 for a design and evaluate question, respectively). In each domain, three variables each with two levels were given in the scenario description prior to the design item (e.g., in the drink sales items, the given variables were age of seller, type of drink, and the time of day). For each design item, students selected a value for all three variables in both conditions, then typed a response to "Explain why you set up the experiment the way you did." For each evaluate item, students first selected a response for whether the experiment was a good or bad way to find out about the focal variable. The evaluate items were for maximally contrastive (drink sales), unconfounded (rocket design), and non-contrastive (cookie baking) experiments. Students were asked to fix any experiments they evaluated as "bad."

Students were given one point for each unconfounded set-up they designed, one point for each of the three experimental set-ups they correctly evaluated (as "good" or "bad"), one point for correcting the maximally confounded and non-contrastive set-ups (in which all and none of the experimental variables were contrasted, respectively), and one point for responding that the unconfounded set-up was "a good way" to find out about the focal variable. Thus, the maximum story pretest score was nine.

Within two days, students began the next phase of the intervention in which they were first introduced to the virtual ramps apparatus and its four dichotomous variables (slope, surface, brand of ball, and starting position) on the TED tutor (Fig 3).

Students in the "Both" condition were then presented with a computerized ramps pretest (Fig 4) in which they designed one experiment for each of the four variables. For each variable, students heard and saw as on-screen text: "Design an experiment to test whether [focal variable] affects how far balls roll." After setting up each experiment by selecting a value for each variable on both ramps, they were asked to select their general goal in setting up the ramps from a drop-down menu (e.g., "I'm comparing the two ramps, or parts of them"), and then a more specific rationale (e.g., "To have only one part of the ramps different") from a subsequent drop-down menu. These menu responses were based on common open-ended responses given by students in earlier evaluations. Because they provide additional information to students, both correct and incorrect, the drop-down menus provide a form of scaffolding. However, students were not provided with any feedback either on their experimental set-ups or their responses to the probes in the drop-down menus.

Students in the "Instruction-only" condition were not given the ramps pretest, but rather began the interactive instructional phase of the lesson. They first evaluated a maximally-contrastive set-up where all four variables' values differed between the two ramps. Students in the Instruction-only condition then:

(a) indicated whether they thought the experiment was a good or bad way to find out about the focal variable (the ball),

(b) typed an open-ended response explaining their evaluation response,,

(c) responded "yes" or "no" to: "Imagine that the balls rolled different distances. Could you tell for sure that the ball caused the difference?"

(d) typed an open-ended explanation for their response,

(e) were given feedback on their response, then

(f) heard an explanation for why the set-up was not a good way to find out about the focal variable.

The tutor then simulated controlling all experimental confounds as the voice-over explained its actions, yielding one that was unconfounded with respect to the focal variable. The interaction then continued as follows:

(g) students were asked to explain why the experiment was a good way to find out about the focal variable,

(h) they were asked whether or not they "would know for sure" that the focal variable caused a hypothetical different outcome and to explain their response,

(i) they heard an explanation for why the unconfounded experiment was a good way to find out about the focal variable.

All students then evaluated an unconfounded experiment. The instructional interactions were the same as previously described through (e), where students heard an explanation for why the experiment was instead a good way to find out about the focal variable. Students in each condition then evaluated a singly confounded experiment (where one variable in addition to the focal variable was contrasted). Instruction for this singly-confounded experiment included all of the same interactions as described for the initial maximally-contrastive experiment.

Table 4. Study design and procedure.

<table>
<tr><th></th><th colspan="2">Condition</th></tr>
<tr><th>Period</th><th>Both</th><th>Instruction-only</th></tr>
<tr><td>1</td><td>Story pretest</td><td>Story pretest</td></tr>
<tr><td rowspan="4">2</td><td>Ramps pretest</td><td>Evaluate maximally-contrastive experiment</td></tr>
<tr><td colspan="2">Evaluate non-confounded (CVS) experiment</td></tr>
<tr><td colspan="2">Evaluate singly-confounded experiment</td></tr>
<tr><td colspan="2">Story posttest</td></tr>
</table>

Immediately after the instructional phase, all students completed the computerized story posttest, which consisted of the same items as the story pretest and was scored in the same manner. This assessed students' ability to transfer what they learned to domains beyond the instructional domain.

Results

Story pretest

Students in the Instruction-only condition had somewhat higher story pretest scores than Both condition students (M = 2.81, SD = 1.55; M = 2.24, SD = 1.61, respectively, out of 9), but this difference was not significant, $F(1, 46) = 1.59$, $p = .21$. There were no between-teacher differences in pretest scores among students who completed the study and whose data are included in the analyses, $F(2, 44) = 0.20$, $p = .82$.[3]

[3] Students of different teachers had similar standardized math and science scores; thus, the students in this study were similar across teachers. Furthermore, there were no interactions with teacher for outcome measures.

Time on task

Students in the Instruction-only condition took about 10% longer to complete the three rounds of evaluation instruction than students in the Both condition took to complete the ramps pretest and two rounds of evaluation instruction ($M = 23.17$ minutes, $SD = 5.91$; $M = 21.02$ minutes, $SD = 3.95$, respectively). However, this difference was not significant, $F(1, 46) = 2.06$, $p = .16$. Time on task was not correlated with learning outcome and was not included as a covariate in subsequent analyses.

Standardized science scores

Instruction-only students' standardized MCAS science scores were significantly higher than the Both condition students' ($M = 242.52$, $SD = 16.45$; $M = 229.56$, $SD = 11.05$), $F(1, 39) = 8.25$, $p = .007$.[4] Furthermore, MCAS science scores were highly correlated with story posttest score ($r(39) = +.45$, $p = .003$) and were included as covariates in subsequent analyses.

Story posttest

Students in the Instruction-only condition had significantly higher posttest scores than students in the Both condition ($M = 7.00$, $SD = 2.10$; $M = 5.38$, $SD = 3.25$, respectively, out of 9), $F(1, 45) = 4.27$, $p = .045$. However, when science scores were covaried, there was no overall difference between conditions, $F(1, 37) = 0.10$, $p = .76$ (adjusted $M = 6.38$, $SE = 0.57$; $M = 6.10$, $SE = 0.64$, respectively). Thus, there was no main effect of condition on transfer performance.

Initial knowledge of CVS

The predicted initial knowledge by treatment interaction was assessed by coding students' responses to the final item on the story pretest. This item required students to evaluate a non-contrastive set-up (i.e., the value for each variable was the same across conditions) and to correct it by varying only the focal variable. Thus, it directly assessed students' understanding of the need to vary the focal variable. Moreover, because this was the final item on the story pretest, it could reveal knowledge of contrasting variables students may have developed while completing

[4] Instruction-only students had marginally significantly higher MCAS math scores, $F(1, 36) = 3.37$, $p = .08$.

the pretest. Student responses were classified as indicating "knowledge of contrasting the focal variable" if they indicated that the experiment was *not* a good way to find out about the focal variable and expressed the need to vary at least the focal variable (e.g., students said the experiment was "Bad" because "They didn't put in different sweeteners" or "They're the exact same cookie!"), and they contrasted at least the focal variable.

Student responses that did not express a need to vary at least the focal variable were coded as "no [CVS] knowledge" responses. These responses included indications that the student was unable to explain their evaluation response ("I don't know" or "I just guessed") and off-target responses indicative of engineering goals (cf. Schauble et al., 1991; Siler & Klahr, in press). With engineering goals, students attempt to produce some desired outcome (e.g., the fastest rocket or the best-tasting cookies) or they may evaluate an experiment they perceive to meet a desired outcome as "good" (e.g., by indicating that the non-contrastive cookies experiment, in which both cookies had sugar, was "good" because "people like sugar in their cookies"). (Note that it is possible that students who were unable to produce explanations for their evaluation responses actually held engineering goals.)

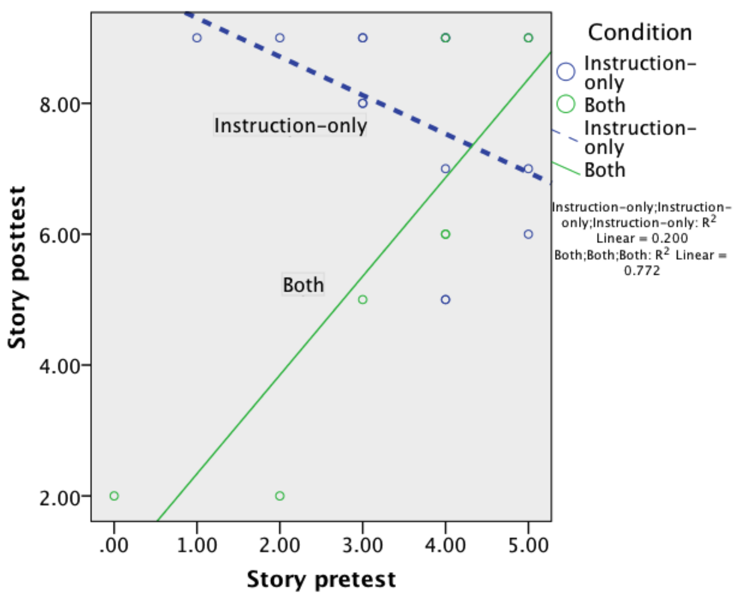
Story posttest with covariates

In an ANCOVA covarying for standardized science and story pretest scores and including students' initial knowledge and condition as fixed variables, the predicted condition by initial knowledge interaction was significant, $F(1, 32) = 5.61$, $p = .02$. (There were also significant condition by science score and condition by story pretest interactions, $p = .044$ and $p = .041$, respectively.) Because of this interaction, the effect of condition for students who expressed knowledge of comparing/contrasting the focal variable is investigated separately from the effect of condition for students who did not express this understanding.

Students with some initial knowledge. We predicted that the Instruction-only condition would be more effective than the Both condition for those students who demonstrated knowledge of varying the focal variable (as assessed on the final question of the story pretest). To test this, an

ANCOVA with condition (Instruction-only or Both) as the independent variable, story pretest score and science achievement scores as covariates, and story posttest score as the dependent variable was run. However, for just these students, the effect of condition could not be determined because there was a significant condition by story pretest interaction, $F(1, 11) = 11.95$, $p = .005$. As shown in Figure 5, there was a significant positive relationship between story pretest and posttest score for students in the Both condition, $r(6) = +.88$, $p = .009$, but no correlation for students in the Instruction-only condition, $r(13) = -.45$, $p = .11$.

Fig. 5 Condition by pretest interaction for students with an understanding of varying the focal variable



Thus, the predicted advantage of the Instruction-only condition was found, but only for the lowest-knowledge students. Why did only the lowest-pretest students in the Both condition perform so poorly on the story posttest relative to the Instruction-only students (rather than all students, regardless of initial knowledge, as predicted)? One possibility consistent with a micro-genetic analysis of students' story and ramps pretest responses is that students in the Both condition who scored low on the story pretest “solidified” their (perhaps recently-developed) understanding of the need to vary at least the focal variable during the ramps pretest. Although both lower- and higher-knowledge students tended to over-apply the “compare and contrast” rule in their experimental designs on the story pretest, the lowest-knowledge students were more likely to continue to over-apply this rule on the ramps pretest, where on average, 3.33 (SD = 1.15) of their set-ups were confounded. In contrast, the higher-pretest students appeared to learn to control variables during the pretest activity; on average, only 0.67 (SD = 1.15) of the higher-pretest students’ ramps pretest set-ups were confounded. These higher-knowledge students set up significantly more unconfounded experiments, $F(1, 4) = 8.00$, $p < .05$; however, this result is tentative due to the low sample size. The lowest-pretest students also tended to continue to focus on “comparing and contrasting” throughout the instructional phase, at the expense of learning to control the other variables. For example, one student initially responded that the first (an unconfounded) experiment was “good” because “they are different balls and ramps [surfaces]” (although the balls were actually the same). This student persisted in only referencing the contrasted—and not the controlled—variables throughout the instruction. On the final question, this student responded that the (unconfounded) experiment was good “because [the balls] are starting at different places.” In contrast, all of the higher-knowledge students showed at least some understanding of the need to control variables on the ramps pretest by designing unconfounded experiments and selecting explanations that expressed the need to control variables (e.g., “So only one thing is different”). All of these students continued into the instructional phase expressing the need to control all but the focal variable as well as the rationale for controlling in their open-ended responses.

Students with no initial knowledge. For students who did not even display an understanding of the need to vary at least the focal variable on the final story pretest item, we predicted an advantage for the Both condition. However, there was no significant effect of condition, $F(1, 21) = 0.57$, $p = .46$. If the development of an understanding of varying at least the focal variable benefits students' learning from the instructional phase, then the number of ramps pretest responses indicating at minimum knowledge of contrasting at least the focal variable would be expected to correlate with story posttest. However, this relationship was not significant, $r(12) = +.38$, $p = .18$.

Then what *was* predictive of Both condition students' learning gains? Specifically, what preparatory activity actions were associated with better transfer performance? In an exploratory analysis, pairwise correlations of story posttest score as the dependent variable were performed on a number of independent variables (story pretest score, MCAS science score, MCAS math score, the number of "engineering goal" responses given on the ramps pretest, the number of "science goal" responses given on the ramps pretest, the number of unconfounded experiments students set up on the ramps pretest, and the number of explanations indicating knowledge of both varying the focal variable and controlling all others students selected on the ramps pretest). Only the number of engineering and science goal responses were significantly correlated with story posttest score ($r(12) = -.66$, $p = .01$; $r(12) = +.51$, $p = .06$, respectively). In a backward regression with these two goal responses as independent variables, only the number of engineering goal responses remained in the model[5]. This result suggests that students' misinterpretations of the instructional goal during the ramps pretest—or failure to realize the science goal underlying the task—may have hindered their subsequent learning during instruction. For just the Both condition students who selected at most one engineering response, there was still no significant relationship between the number of responses indicating at least an understanding of contrasting variables and posttest score. This null result again runs counter to the hypothesis that students' development of this understanding accounted for any advantage of completing the preparatory activity. Table 5 shows the frequency distribution for the number of

[5] Note that the number of science goal and the number of engineering goal responses were not perfectly inversely related because students could also select "I don't know" or "I don't really have a goal."

engineering responses students selected during the ramps pretest. The majority of students (69%) selected fewer than two engineering responses.

Table 5. Engineering response frequencies (of students with no initial knowledge of varying the focal variable).

	Number of Engineering responses selected (out of 4)			
	0	1	2	3
Frequency	6	5	4	1

Furthermore, comparing Instruction-only students with no initial CVS knowledge to Both condition students with no initial CVS knowledge who selected at most one engineering goal response on the ramps pretest (to allow for one accidental engineering response)[6], there was a marginally significant advantage of the Both condition over the Instruction-only condition (unadjusted: $M = 6.80$, $SD = 2.90$; $M = 6.18$, $SD = 2.44$, respectively; adjusted for science score: $M = 7.54$, $SE = 0.77$; $M = 5.51$, $SE = 0.73$, respectively), $F(1, 18) = 3.28$, $p = .087$ (shown in Fig. 6). Thus, the predicted advantage of completing the ramps pretest prior to instruction is supported for students who did *not* misinterpret the goal of instruction on more than one item. In contrast, Both condition students with no initial CVS knowledge who selected engineering responses on *at least* two ramps pretest items scored significantly lower on the story posttest than students with no initial CVS knowledge in the Instruction-only condition (unadjusted: $M = 6.18$, $SD = 2.44$; $M = 1.00$, $SD = 1.00$; adjusted for standardized science score: $M = 5.79$, $SE = 0.55$; $M = 2.44$, $SE = 1.14$, respectively), $F(1, 11) = 6.50$, $p = .03$. Thus, students who misinterpreted the goal or failed

[6] There was still a marginally significant advantage of the Both over the Instruction-only condition when just the Both condition students who selected *no* engineering responses were included in the analysis, $F(1, 13) = 3.49$, $p = .085$.

to realize the nature of the pretest task may benefit more from an additional round of instruction than completing the ramps pretest.

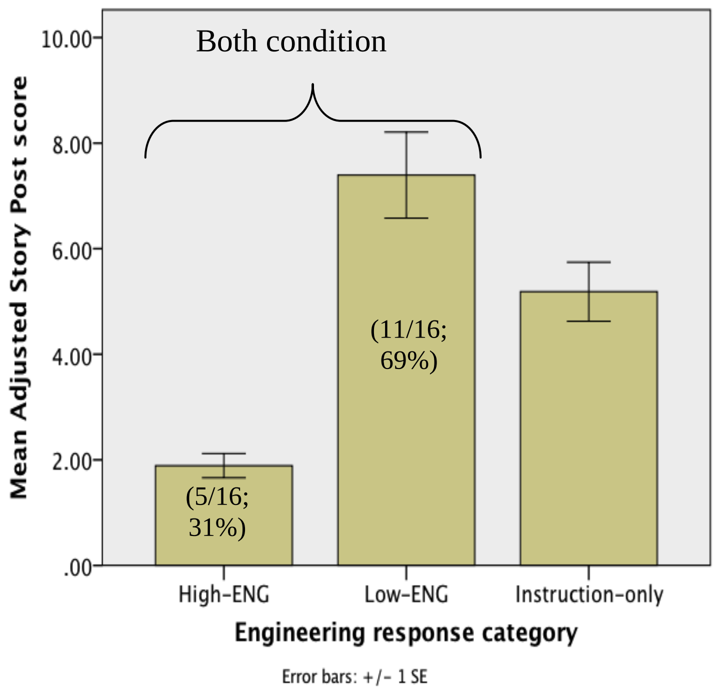


Fig. 6 Mean adjusted story posttest score by condition and response type.

However, the comparisons between the two Both groups and the Instruction-only group should be met with caution because the percentage of students in the Instruction-only group who would have fallen into the high- and low-engineering categories on the ramps pretest in unknown. In other words, are students in the Instruction-only condition more comparable to the "High-Engineering" or "Low-Engineering" students in the Both condition? The Instruction-only group was more similar to the Low-Engineering than High-Engineering students in the Both condition in terms of story pretest score (M = 2.00, SD = 1.53; M = 2.10, SD = 1.52; M = 1.00; SD = 0.82, respectively), and in terms of science achievement scores (M = 239.67, SD = 16.55; M = 228.20, SD = 12.49; M = 224.00; SD = 8.72, respectively). Because of their similarities on these measures, it is plausible that these Instruction-only students would generally have correctly interpreted the goal during the ramps pretest had they completed the ramps pretest rather than gone directly into the instruction. This supports an advantage of completing the ramps pretest prior to entering the instructional phase for students who correctly interpreted the goal of the activity.

Table 6. Study 1 results summary: Outcome by initial knowledge state and student characteristic.

Initial knowledge state	Student characteristic	Outcome
No initial knowledge	High-Engineering ramp pretest	(Instruction-only > Both?)
	Low-Engineering ramp pretest	Both > Instruction-only
Some initial knowledge (contrasting focal variable)	Higher-story pretest	Instruction-only = Both
	Low-story pretest	Instruction-only > Both

In summary, we predicted that students with some initial knowledge of experimental design would benefit more from the instructional activity and students who did not express initial knowledge of contrasting the focal variable would benefit more from first completing the pretest activity. And, as predicted, there was a significant initial knowledge by condition interaction. However, overall, students with no initial CVS knowledge did *not* benefit more from first setting up and explaining ramps experiments compared to going directly into the instruction. As shown in Table 6, only those “no initial knowledge” students who correctly interpreted the goal of the ramps pretest benefited more from the preparatory task than from instruction. Though students who more often misinterpreted the nature of the ramps pretest task performed worse on the story posttest, whether they would have performed significantly worse than a similar group of students who only received instruction needs to be addressed in future research. For students who expressed initial knowledge of contrasting at least the focal variable, although there was no overall difference between conditions, there was a condition by story pretest score interaction. The lowest-knowledge students benefited more when they did *not* complete the preparatory activity prior to instruction.

Initial knowledge state: Science vs. Non-science orientation

Students' correct interpretation of the task (i.e., in terms of a science goal) during the ramps pretest was highly correlated with posttest performance. If the preparatory ramps pretest task was successful primarily because it helped orient students' framing of the instructional task, then we would expect students who came into the instruction with non-science goals in particular to benefit from the ramps pretest task, and students who came into the instruction with science goals to benefit more from instruction. This suggests considering students' task interpretation expressed on the story pretest—rather than simply whether they contrasted at least the focal variable—as the "initial knowledge state." That is, on the final story pretest item, students may have expressed some knowledge of contrasting (at least) the focal variable but did so to serve an engineering goal rather than a science goal. In Siler and Klahr (in press), we identified many cases where students contrasted variables *not* because they were comparing them, but rather, because they held engineering goals of producing different results across conditions[7].

Thus, when students expressed a desire to contrast one or more variables but did not explicitly provide a rationale consistent with a science goal of *trying to find out about* one or more variables (e.g., "I want to see if the sugar is better than honey"), we examined their previous responses on the story pretest to aid in disambiguation. If students expressed a science goal on at least one earlier item, we credited their final item response as "Science goal." If students did not express a science goal on a previous item (i.e., they expressed engineering goals and/or were unable to explain their goals), we coded students' final responses as "No science goal." Students

[7] This engineering goal may be related to students' notion that a "good" experiment is one that shows that one or more variables have an effect on the outcome (Siler & Klahr, in press).

who were unable to explain their evaluation decision or who expressed engineering goals on the final pretest item were (as in the original coding) classified as “No science goal” students. Finally, student responses that did not express the need to contrast variables were coded for whether they expressed a science goal on the final question (e.g., “I’m trying to find out about the sweetener”).

Recoded initial knowledge. When student responses were re-categorized in this way, only two students who were categorized as expressing “Some initial knowledge” were considered “No science goal.” The remaining “Some initial knowledge” students were categorized as “Science goal” and all of the “No initial knowledge” students were classified as “No science goal.” This latter finding is consistent with that reported in Siler and Klahr (in press): students who expressed science goals on the story tests almost always contrasted one or more variables. Thus, when students did not express the necessity of contrasting variables, they generally did not hold science goals. Consequently, all result patterns discussed for Study 1 students held when initial knowledge was re-categorized in terms of goal expression on the final story pretest item.

Study 2 Methods

Study 1 was replicated the following year in the same school. The Study 2 procedure was the same as that of Study 1 (shown in Table 4), with the major exceptions that: (a) because computer labs were not available, the story pretest and posttests were given on paper rather than on the computer, and (b) the story posttests were given the next day rather than immediately after students completed the instruction. Thus, in Study 2, there was more time between instruction and posttest than in Study 1.

Participants

Participants were seventh-grade students from eight science classes taught by two teachers in the same (primarily middle-high-SES) suburban middle school as in Study 1. Unlike the previous year, where students were randomly assigned to teacher, higher-ability students were clustered with one teacher (“Teacher C”) and the lower-ability students with Teacher B.

Students were randomly assigned to condition within each class. Of the 139 students who completed the story pretest, 63 (45%) demonstrated initial mastery of CVS (i.e., they scored at least six out of nine) and were excluded from further analyses. Of the 76 remaining students, eight experienced technical difficulties (five Instruction-only and three Both-condition students) and three did not complete the story posttest. Data from the remaining 65 students (35 in the Instruction-only and 30 in the Both condition) were included in the following analyses.

Students comprising the analyzed data were lower-performing than the general student population at this school: their standardized reading and math percentile scores on the MAP (Measures of Academic Progress) tests were significantly lower than those of students who mastered CVS on the story pretest (reading: $M = 46.69$, $SD = 28.65$; $M = 84.93$, $SD = 16.32$, respectively; math: $M = 53.90$, $SD = 26.87$; $M = 86.91$, $SD = 11.56$, respectively). Though lower-performing than their school peers, these students were about average among all seventh-grade students in Massachusetts who completed MAP testing.

Study 2 Results

Standardized science scores

There were no significant differences between conditions on any of the available standardized test scores (MAP math, MAP reading, MCAS math, or MCAS science). MAP standardized math percentiles were the most recent standardized test data available and they were most highly correlated with story pre- and posttest scores; thus, they were included as covariates (as a general ability measure) in subsequent analyses.

Time on task

As in Study 1, students in the Instruction-only condition took slightly (1.36 minutes or 6%) longer to complete the three rounds of evaluation instruction than students in the Both condition took to complete the ramps pretest and two rounds of evaluation instruction ($M = 22.82$ minutes, $SD = 4.14$; $M = 21.46$ minutes, $SD = 3.23$, respectively). This difference was not significant, $F(1, 61) = 2.05$, $p = .16$. As with Study 1, time on task was not correlated with learning outcome and was not included as a covariate in subsequent analyses.

Story posttest

There was no difference between the Instruction-only and Ramps conditions on the nine-point story posttest ($M = 5.69$, $SD = 3.02$; $M = 5.80$, $SD = 2.78$, respectively, $F(1,63) = 0.03$, $p = .88$). This difference remained non-significant when story pretest and MAP math percentile scores were included in the ANCOVA, $F(1, 47) = 0.07$, $p = .80$. Thus, as in Study 1, there was no main effect of condition on transfer performance.

Story posttest with covariates

Because there was a significant teacher by condition by initial knowledge (Science goal vs. No science goal) interaction, $F(1, 35) = 4.29$, $p < .05$, the condition by initial knowledge interaction was examined separately for students of each teacher. To gain insight into the possible causes of this interaction, we examined the characteristics of students of the two teachers (Teacher B and Teacher C). Overall, students of Teacher C scored significantly higher on the story pretest than students of Teacher B ($M = 5.38$, $SD = 2.69$; $M = 4.40$, $SD = 3.08$, respectively), $F(1, 139) = 4.04$, $p < .05$. Likewise, the CVS mastery rate on the story pretest was marginally higher for Teacher C's students than Teacher B's students (51% and 37%, respectively), $\chi^2(1, 142)$. = 3.06, $p = .08$. Furthermore, of students who did not show pretest mastery (i.e., who scored less than 6 out of 9 points), Teacher C's students again scored significantly higher on the story pretest than Teacher B's ($M = 3.06$, $SD = 1.60$; $M = 2.35$, $SD = 1.45$, respectively), $F(1, 77) = 4.24$, $p = .04$. In addition, students of Teacher C also had significantly higher standardized reading comprehension percentile scores than students of Teacher B ($M = 72.78$, $SD = 17.53$; $M = 41.49$, $SD = 25.70$, respectively), $F(1, 60) = 29.42$, $p < .001$.

To determine whether the differences between students of Teachers B and C could be explained by ability differences, reading comprehension scores were covaried in an ANCOVA.

However, there was no significant relationship between reading comprehension and story pretest score, and Teacher C's students pretest scores remained marginally higher, $F(1, 60) = 3.89$, $p = .05$. Outcomes for science achievement scores exactly paralleled those of reading comprehension. Thus, reading ability and general science knowledge did not explain the advantage of students in Teacher C's classes. Consequently, it is feasible that Teacher C's students experienced more CVS-related instruction than Teacher B's.

Table 7. Knowledge and ability characteristics by study.

	All students	Only students included in analyses		
	Mastery rate	Story pretest	MCAS Science score level	Math score/ percentile level
Study 1	55%	2.57 (1.60)	237.35 (15.43) "Needs improvement"	244.21 (18.26)[a] "Proficient"
Study 2: Teacher B	37%	2.49 (1.50)	229.67 (13.40) "Needs improvement"	41.71 (26.43)[b]
Study 2: Teacher C	51%	3.17 (1.63)	241.83 (11.82) "Proficient"	73.05 (16.97)[b]

[a] MCAS Math score.

[b] MAP math percentile scores.

Of just Study 2 students whose data were included in the following analysis, Teacher C's students' story pretest scores were marginally higher than Teacher B's (shown in Table 7), $F(1, 62) = 3.07$, $p = .085$. In addition, students of Teacher C had significantly higher MCAS science scores, $F(1, 52) = 12.18$, $p = .001$, significantly higher standardized MAP math percentile scores, $F(1, 48) = 23.32$, $p < .001$, and significantly higher reading comprehension percentile scores, $F(1, 49) = 17.10$, $p < .001$, than students in Teacher B's classrooms. Thus, for students whose data were included in these analyses, Teacher C's were higher-knowledge, higher-ability, and may have had more prior CVS-related instruction than Teacher B's.

Furthermore (as shown in Table 7), the mean story pretest score of Study 1 students whose data were included in the analyses were closer to those of Teacher B's students. In addition, Study 1 students and Teacher B's students were on average both at the "Needs Improvement" performance level on MCAS science achievement scores, whereas Teacher C's students were on average "proficient." Thus, Teacher B students whose data are included in the following analyses are more similar to Study 1 students than Teacher B's students in terms of story pretest scores and standardized science achievement scores. The Study 1 students and Teacher B's students appear to be somewhat below average among seventh-grade students who participated in the standardized testing in Massachusetts.

Teacher B. For students in Teacher B's classes, there was no main effect of condition on posttest score (Instruction-only: $M = 4.00$; $SD = 3.06$; Both: $M = 4.94$, $SD = 2.90$, $F(1, 32) = 0.85$, $p = .37$). This main effect remained non-significant when story pretest and standardized math scores were included as covariates, $F(1, 23) = 0.80$, $p = .38$.

However, the predicted condition by initial knowledge interaction was significant, $F(1, 20) = 5.13$, $p = .04$, and in the expected direction. That is, "No science goal" students gained more in the Both condition and "Science goal" students gained more in the Instruction-only condition. For "No science goal" students, there was a marginally significant time (pre to post) by condition interaction, $F(1, 23) = 3.54$, $p = .07$, where students in the Both condition gained more than those in the Instruction-only condition. Furthermore, though students in the Both condition showed pre-to-post learning gains, $F(1, 10) = 8.45$, $p = .02$, students in the Instruction-only condition did not, $F(1, 10) = 0.07$, $p = .80$. There was a significant condition by time interaction for "Science goal" students, $F(1, 3) = 30.94$, $p = .01$, where students in the Instruction-only condition gained more than students in the Both condition (though there were marginally significant gains in each condition). However, this latter result is tentative due to the small sample size.

For "No science goal" students in the Both condition, there were significant relationships between posttest score and (a) the total number of science goal responses (or, inversely, the number of engineering and "I don't know" responses), $r(12) = +.56$, $p = .06$, (b) story pretest score, $r(12) = +.51$, $p = .09$, and (c) standardized science achievement scores, $r(11) = +.56$, $p = .07$.[8] When these three variables were entered into a backward regression model, only the number of science goal responses remained (though the partial correlations for science achievement and number of science goal responses were similar). Thus, as in Study 1, goal interpretation during the ramps pretest was most strongly related to transfer performance for students who did not initially show understanding of the task goal.

Teacher C. As with Study 1, there was no main effect of condition (Instruction-only: $M = 7.24$, $SD = 2.02$; Ramps: $M = 7.33$, $SD = 1.78$), $F(1, 27) = 0.02$, $p = .89$. The main effect of condition remained non-significant when story pretest and standardized math scores were included as covariates, $F(1, 19) = 0.001$, $p = .98$.

However, unlike in Study 1 and with Teacher B's students, there was no significant condition by initial goal interaction, $F(1, 14) = 0.86$, $p = .37$. Only students' standardized math scores were significantly related to posttest score, $F(1, 18) = 8.98$, $p = .008$. Based on the high posttest scores (where one standard deviation above the means is greater than the nine-point maximum possible score in each condition), this result may be due to a ceiling effect for these higher-ability and higher initial CVS knowledge students. The results of both Studies 1 and 2 are summarized in Table 8.

---

[8] Variables not significantly related to story posttest score include: the number of unconfounded experiments students designed on the ramps pretest, the number of responses indicating at least understanding of contrasting variables, standardized math scores, and standardized reading scores.

Table 8. Summary of outcomes by initial knowledge state and student characteristic by study.

Study 1			Study2		
Initial knowledge state	Student characteristic	Outcome	Initial knowledge state	Teacher B	Teacher C
No initial knowledge/ No science goal	High-Engineering ramp pretest	(Instruction-only > Both?)	No Science goal	Both > Instruction-only (goal selection on ramps pre related to post)	(only MCAS math scores related to posttest)
	Low-Engineering ramp pretest	Both > Instruction-only			
Some initial knowledge/ Science goal	Low-story pretest	Instruction-only > Both	Science goal	Instruction-only > Both[a]	
	Higher-story pretest	*Instruction-only = Both*			

[a] This result is tentative due to the low sample size.

Discussion

The goals of these studies were (a) to uncover the conditions in which completing an initial feedback-free preparatory activity was beneficial to transfer performance, and (b) to determine the mechanisms responsible for any benefits of the activity. To do this, we looked into individual differences in initial knowledge that affected whether completing this activity prior to an interactive instruction was more beneficial to learning than if this time had been spent engaging in instructional activities. Participants were seventh-grade students from a primarily middle-to-high SES population who likely had some prior instruction in experimental design. In both studies, no overall advantage of one condition was found. This result differs from those of Lorch et al. (2010), who found that—for fourth-grade students from higher-achieving schools in particular—there was an advantage of completing a ramps activity in which students designed experiments prior to engaging in instruction compared to only engaging in the instruction. In the current studies, the benefit of condition appears to depend on the student's initial knowledge state.

Furthermore, Lorch et al., (2010) found that students developed the procedural knowledge of contrasting the focal variable primarily from completing an initial ramps design activity —an idea that was relevant to, but not the primary focus of the instruction. If students' development of this knowledge benefited their subsequent learning from instruction, we predicted there would be a condition by initial knowledge interaction. Specifically, we predicted that students with no initial understanding of the need to contrast at least the focal variable would benefit more from first setting up ramps experiments, provided doing so led to the development of this procedural understanding. In contrast, we expected that students who already understood the need to contrast would benefit more from the interactive instruction, which emphasized the underlying rationale for controlling other variables, an understanding of which has been found to correlate with transfer performance (Siler et al., 2010).

Consistent with our prediction, in Study 1, there was a significant interaction between condition and knowledge of "comparing and contrasting" at least the focal variable. However, among students who expressed an understanding of the need to contrast variables across the experimental conditions, there was a condition by story pretest score interaction. Study 1 students in the Instruction-only condition tended to do well on the story posttest, regardless of their initial CVS knowledge as measured by pretest score. In contrast, the lowest-knowledge students in the Both condition performed worse than their counter-parts in the Instruction-only condition. Microgenetic analyses revealed that the lowest-knowledge students seemed to reinforce their knowledge of contrasting variables and over-generalized this knowledge to variables other than the focal variable during the ramps pretest. During instruction, this knowledge seemed to have framed students' interpretation of the questions. For example, when asked: "Is this a good way to find out if the type of surface affects how far balls rolled?" these students gave responses such as: "Yes, because the surfaces are different," indicating consideration of the contrasting aspect of CVS but not the controlling aspect. Higher-pretest students with some initial knowledge (or understanding of the task goal) performed similarly in the Instruction-only and Both conditions. This appeared to be due to students' development of a complete procedural understanding of CVS during the ramps pretest. This understanding may have freed cognitive resources and allowed students to focus on the rationale for controlling variables, which was previously found to be predictive of transfer performance.

However, for Study 1 students with no initial knowledge of varying the focal variable or other aspects of CVS (and who did not express understanding of the science goal nature of the task), no overall advantage to completing the preparatory activity was found. Moreover, there was no significant correlation between learning outcome and the frequency of drop-down menu selections that indicated at least an understanding of the need to contrast variables. This finding was not consistent with the hypothesis that the development of initial procedural knowledge of contrasting was the underlying mechanism of learning from the preparatory activity. Rather, in an exploratory analysis, the number of "engineering" and the number of science goal responses students selected on the ramps pretest were found to be the only ones among several plausible factors considered that correlated with learning outcome in Study 1. This suggests that the development of a more general framing of the preparatory activity accounted for the benefit of the activity for the primarily middle-to-high-SES seventh-grade student participants.

Because there was a significant relationship between goal selection during the ramps pretest and learning outcome, we looked at whether students' understanding of the task goal predicted which condition was more beneficial. When Study 1 students' initial goals (Science goal vs. No science goal), as expressed on the final story pretest item, were coded, the same pattern of results was found as for students' initial knowledge of contrasting variables.[9]

In Study 2, we again examined the interaction between students' expressed goal (Science goal or No science goal) and condition. If students' development of a science goal orientation

[9] This is likely because only two Study 1 students switched categories; these students, who were initially categorized as "Some initial knowledge [of contrasting variables]," were classified as "No science goal".

during the ramps pretest were responsible for the benefit of the preparatory activity, then students without this orientation should benefit from the ramps pretest provided they develop this goal orientation, and students who expressed this goal orientation should benefit more from directly entering instruction. This predicted interaction was found for the lower-knowledge and -ability students of one of the participating science teachers, who were similar to Study 1 students in terms of initial knowledge and science achievement scores. As in Study 1, students who did not express understanding of the story pretest task goal benefited more from first completing the ramps pretest, but students who expressed this understanding benefited more from the interactive instruction. Furthermore, as in Study 1, the frequency of students' selections of the appropriate (science) goals during the ramps pretest was most strongly correlated with posttest performance. This adds support to the hypothesis that the development of a correct understanding of the task goal underlies the benefit of engaging in the ramps pretest for students with no initial such knowledge. However, this interaction was not found for the higher-ability and higher-knowledge students of the other Study 2 teacher, who appeared to have reached ceiling on the story posttest.

Some study results also justify the concerns based on some findings (e.g., Hiebert & Wearne, 1996; Tuovinen & Sweller, 1999) that lower-ability students in particular may require more guidance during the preparatory activity (cf. Schwartz, Chase, Oppezzo, & Chin, in press) to benefit from it. In Studies 1 and 2, students' goal selections during the ramps pretest were predictive of their posttest performance, and students who more frequently continued to express engineering goals during the ramps pretest performed significantly worse than students who generally did not. These "engineering" students may have performed worse on the posttest than their counter-parts who were only given instruction; however, given possible differences between these two groups, further research is necessary to confirm this. If confirmed, this result not only suggests that some students may not benefit from the preparatory activity, but that the activity may actually be detrimental to their subsequent learning[10].

In addition, the results of Study 1 low-pretest students who expressed knowledge of the task goal on the story pretest yet performed poorly on the posttest, bring up the possibility that—at least in some cases—students may not benefit from the preparatory activity. These cases may include students who were just developing an intuitive understanding of the concept, whose simplified conceptions of experimental design appeared to become robust through practice. However, further research is also necessary to determine the extents to which the development of over-simplified conceptions during preparatory activities is an obstacle to subsequent learning during instruction. If supported in future research, in cases where students failed to benefit from the preparatory activity, it may be better to either provide additional support to students during the activity (such as feedback on their responses). Alternatively, it may be better to forgo this activity and provide instruction immediately.

These results extend those of Lorch et al. (2010), who found that fourth-grade students developed an initial understanding of the need to contrast the focal variable primarily during the preparatory task (and developed knowledge of controlling variables primarily from the instruction). The present results suggest that the more important factor—which accounted for the benefit of the preparatory activity for students from higher-achieving schools—may have been the development of a perhaps more deep-rooted science-goal orientation to the task.

The fourth-grade students from lower-achieving schools did *not* significantly benefit from only manipulating the ramps, and manipulating the ramps prior to instruction was only marginally better than only engaging in instruction (Lorch et al., 2010), resulting in a less than a

[10] Because the number of engineering responses students selected during the ramps pretest was not correlated with any achievement score or pretest score, identifying the students who are likely to continue misinterpreting the goal during the preparatory task may be difficult.

one-half of one point difference between conditions on a delayed posttest. On the other hand, students from the higher-achieving schools benefited from simply manipulating the ramps and also significantly benefited when they did this prior to instruction compared to only engaging in instruction. Based on the current results, the poorer learning of students from lower-achieving schools during the manipulation task may have been due to their more frequently misunderstanding the nature of the task during the (feedback-free) ramps manipulation phase. And the benefits of this initial activity above instruction alone for students from higher-ability schools may have been due to their development of a science goal orientation. This possibility is supported by Siler, Klahr, and Matlen (under review), who found that students' science goal understanding of the story pretest task increased with both age and SES-level (the latter of which was related to achievement scores). Furthermore, fifth-grade students from higher-SES populations generally expressed both science and engineering goals on the story pretest, whereas the fifth-grade students from lower-SES backgrounds tended to only express engineering goals.

However, it is possible that the fourth-grade students in the both condition of the Lorch et al. (2010) study benefited from the development of a science goal understanding of the task *in conjunction with* an initial understanding of the need to contrast at least the focal variable. That is, students in the Lorch et al. study were three years younger than those in the current studies. According to Gathercole (1999), "complex working memory" in particular, which plays a role in such "complex cognitive activities as language comprehension, mental arithmetic, and reasoning," develops between fourth and seventh grades (approximately 9 to 12 years old).[11] Thus, because they are less equipped to handle an increased cognitive load, these younger students are likely to have more difficulty comprehending instruction with less prior knowledge to support their understanding and therefore require knowledge of contrasting variables in addition to holding a science orientation toward the lesson.

[11] Other types of short-term memory do not change as much in this age range.

It is important to reiterate that, although time on task was controlled in the current studies, because we wanted the preparatory activity to be maximally generative in order to promote construction (or elicitation) of the pre-instructional knowledge hypothesized to support subsequent learning, the procedure confounded problem task (i.e., design versus evaluation of experiments) and feedback provision (i.e., none vs. some). Thus, the effect of feedback, per se, on learning is not addressed by this study. That is, it *cannot* be concluded from this study whether completing a feedback-free preparatory activity prior to instruction would have been better or worse than completing the *same* activity with feedback and instructional explanations. More stringently-controlled studies are necessary to determine the effects of specific instructional factors (such as feedback) on subsequent learning and transfer of CVS and other skills. Such studies are also necessary to support that the development of specific initial knowledge during the preparatory activity—rather than a task-specific factor—provided the learning advantage of the Both condition for some students. What *is* addressed in the current studies are the differential effects of completing a feedback-free generative activity versus engaging in an evaluative instructional activity in which feedback is provided, before engaging in the interactive instruction.

A final point to consider is whether these results would transfer to other student populations. In the current studies, the student population was primarily middle/high-SES, and the initial mastery rates on the pretest were around 50%, indicating that this seventh-grade student population had had some CVS-related instruction. However, the students who did not demonstrate initial mastery and whose data were included in the analyses scored significantly lower on all achievement scores examined than students who showed mastery on the pretest. Though about average among students in their state who completed this standardized testing, within their school, these students were relatively low-achieving. Because students may compare themselves to their peers and/or base their self-efficacy beliefs on such evidence as their school grades (cf. Marsh, 1986; Bong, 1998) or assigned classes (e.g., remedial/advanced, as Study 2 students were assigned to), this student population is likely to have relatively lower self-perceived abilities and less positive motivational beliefs than their peers who were excluded from analyses. One possible consequence of poorer self-concepts is a lower tolerance for frustration: if one believes s/he is unlikely to succeed, negative emotions are likely to arise when one experiences difficulties. And in fact, research has demonstrated that one' efficacy beliefs predict both task persistence (e.g., Vollmeyer and Rheinberg, 2000) and task performance (e.g., Wigfield & Guthrie, 1997; Randhava, Beamer & Lundberg, 1993; Zimmerman, Bandura & Martinez-Pons, 1992). Because more information is given during instruction, and this information is given in somewhat lengthy (and what some students had complained were conceptually dense) explanations, students may become confused. If students' self-perceptions of the abilities are negative, they are more likely to give up trying to understand the instructional content. In contrast, the informational load during the ramps pretest is minimal in comparison, so students are less likely to experience confusion than when completing instruction. This may result in a more pronounced difference in favor of the Both condition for such students compared to their more self-assured counter-parts.

Thus, future studies are necessary to determine whether the same patterns of results hold for different populations of students, for varying study procedures, as well as interactions between these factors. By investigating questions such as these, we can continue to determine the boundary conditions in which engaging students in initial preparatory activities benefits subsequent learning from instruction.

Acknowledgements

We wish to thank Bryan Matlen and Kevin Willows for their insightful comments leading up to and on this manuscript, the teachers at Amherst Regional Middle School for their cooperation in running this study, and Cressida Magaro for her assistance with various aspects of data collection and scoring. We would also like to thank the anonymous reviewers for their comments and suggestions. This research was supported by the Institute of Education Sciences, U.S. Department of Education, through Grants R305H060034 and R305A100404 to CMU. The opinions expressed are those of the authors and do not necessarily represent views of IES or the U.S. Department of Education.

References

Bong, M. (1998). Tests of the internal/external frames of reference model with subject-specific academic self-efficacy and frame-specific academic self-concepts. *Journal of Educational Psychology, 90*(1), 102-110.

Chen, Z., & Klahr, D. (1999). All other things being equal: Children's acquisition of the Control of Variables Strategy. *Child Development, 70*(5), 1098–1120.

Gathercole, S. E. (1999). Cognitive approaches to the development of short-term memory. *Trends in Cognitive Sciences, 3*(11), 410-418.

Hiebert, J. & Wearne, D. (1996). Instruction, understanding, and skill in multidigit addition and subtraction. *Cognition & Instruction, 14*(3), 251-283.

Kapur, M. (2009). Productive failure in mathematical problem solving. *Instructional Science, 38*(6), 523-550.

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

Lorch, R. Jr., Lorch, E. P., Calderhead, W. J., Dunlap, E. E., Hodell, E. C., & Freer, B. (2010). Learning the control of variables strategy in higher and lower achieving classrooms: Contributions of explicit instruction and experimentation. *Journal of Educational Psychology*, *102*(1), 90-101.

Marsh, H. W. (1986). Verbal and math self-concepts: An internal/external frame of reference model. *American Educational Research Journal*, *23*, 129-149.

Randhawa, B. S., Beamer, J. E., and Lundberg, 1. (1993). Role of mathematics self-efficacy in the structural model of mathematics achievement. *Journal of Educational Psychology, 85* (1), 41-48.

Schauble, L., Klopfer, L. E., & Raghavan, K. (1991). Students' transition from an engineering model to a science model of experimentation. *Journal of Research in Science Teaching, 28*(9), 859-882.

Schwartz, D. & Martin, T. (2004). Inventing to prepare for future learning: The hidden efficiency of encouraging original student production in statistical instruction. *Cognition and Instruction*, *22*(2), 129-184.

Schwartz, D. Chase, C., Oppezzo, M. & Chin, D. (In press). Practicing versus inventing with contrasting cases: The effects of telling first on learning and transfer. *Journal of Educational Psychology*.

Siler, S. A., & Klahr, D. (In press). Detecting, Classifying and Remediating Children's Explicit and Implicit Misconceptions about Experimental Design. In Proctor, R. W., & Capaldi, E. J. (Eds.), Psychology of Science: Implicit and Explicit Reasoning. New York: Oxford University Press.

Siler, S. A., Klahr, D., Magaro, C., Willows, K., & Mowery, D. (2010). Predictors of transfer of experimental design skills in elementary and middle school children. *Proceedings of the 10th ITS 2010 Conference. Lecture Notes in Computer Science, 6095,* 198-208.Strand-Cary, M. & Klahr, D. (2008). Developing elementary science skills: Instructional effectiveness and path independence, *Cognitive Development, 23*, 466-511.

Siler. S. A., Klahr, D., & Matlen, B. (Under review). Conceptual change when learning experimental design.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning, *Cognitive Science, 12*, 257-285.

Tuovinen, J. & Sweller, J. (1999). A comparison of cognitive load associated with discovery learning and worked examples. *Journal of Educational Psychology, 91*(2), 334-341.

Vollmeyer, R. & Rheinberg, F. (2000). Does motivation affect performance via persistence? *Learning and Instruction*, *10*(4), 293-309.

Wigfield, A., & Guthrie, J. T. (1997). Relations of children's motivation for reading to the amount and breadth of their reading. *Journal of Educational Psychology*, *89*, 420-433.

Zimmerman, B. J., Bandura, A. & Martinez-Pons, M. (1992). Self-motivation for academic attainment: The role of self-efficacy beliefs and personal goal setting, *American Educational Research Journal*, *29*(3), 663-676.

Fig. 1 Screenshot of story pre/posttest design item

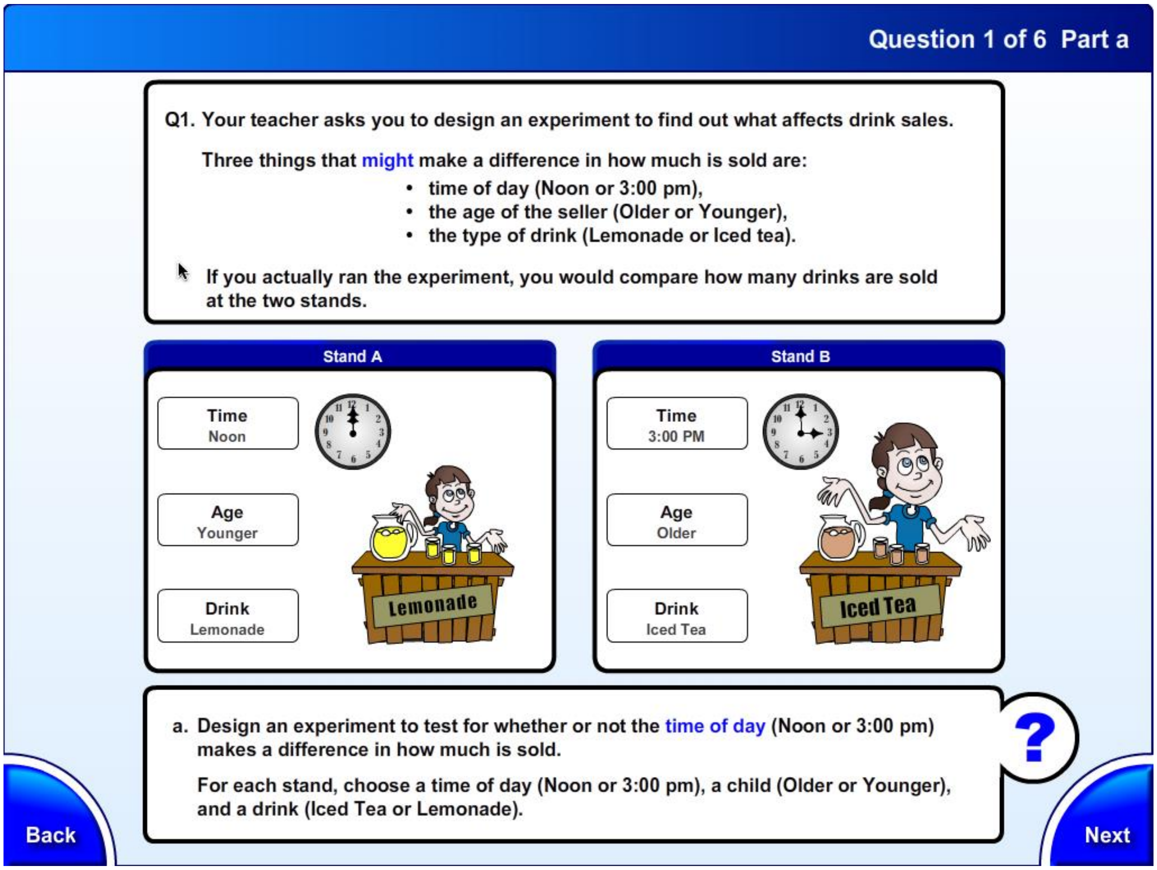


Fig. 2 Screenshot of story pre/posttest evaluate item

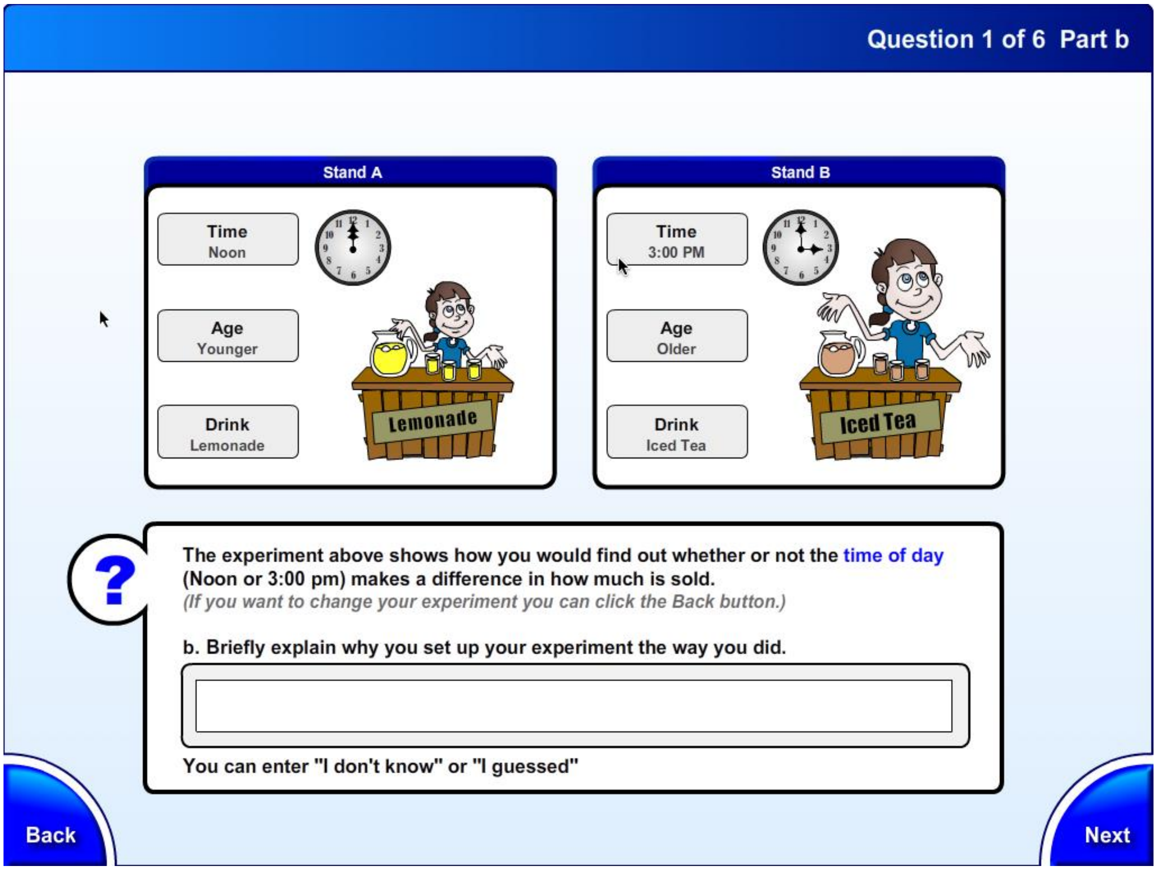


Fig. 3 Screenshot of introduction to ramps apparatus prior to instruction

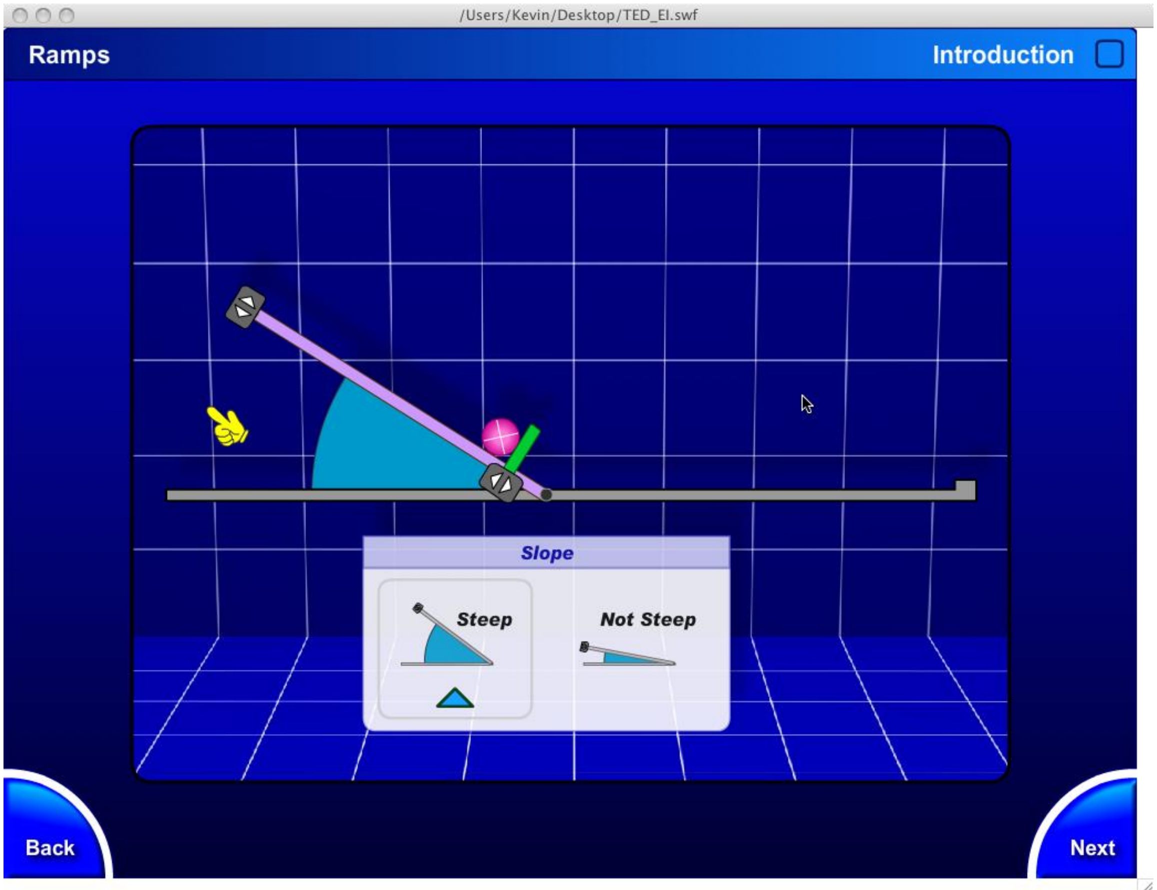
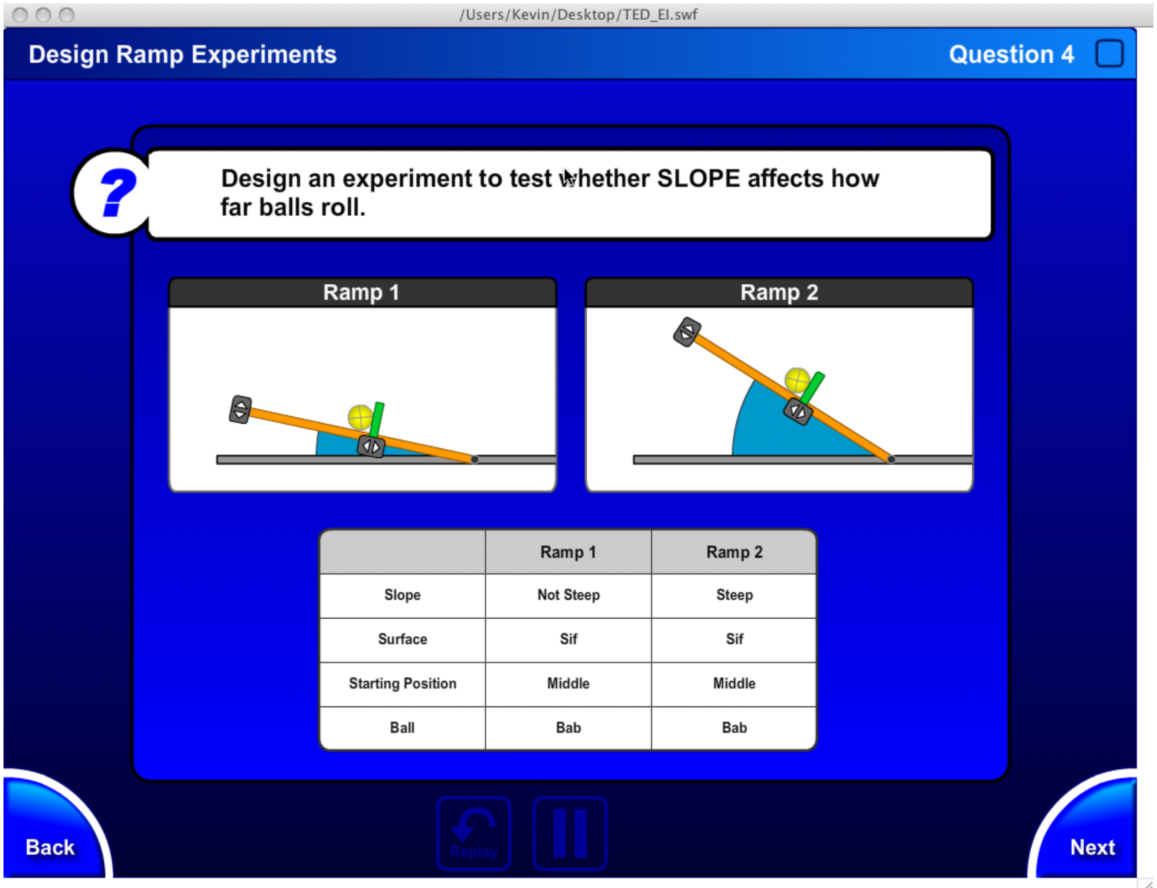


Fig. 3 Screenshot of a ramps pretest question



	Ramp 1	Ramp 2
Slope	Not Steep	Steep
Surface	Sif	Sif
Starting Position	Middle	Middle
Ball	Bab	Bab