DRAFT of

Siler, S, A., Klahr, D., & Price, N. (2012) Investigating the mechanisms of learning from a constrained preparation for future learning activity. ***Instructional Science*****,** ***DOI: 10.1007/s11251-012-9224-7*****.**

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 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 spending the whole time engaged 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. In Study 2, the initial knowledge by condition interaction was found only for the higher-ability students.

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. In addition, lower-achieving students benefitted more from the manipulate activity prior to the lesson than their higher-achieving counter-parts. Furthermore, 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. 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 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 others 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. 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. (2010). 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), which was 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, 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 experimental evaluation instruction in which they evaluated experiments and afterwards 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 initial knowledge development we hypothesized would aid students in subsequent instruction and thus would be 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 is provided, before engaging in the interactive instruction.

Because student-level information was available in the current studies, we were able to look at students' initial CVS knowledge, and trace the development of that knowledge during the preparation 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).

Lorch et al. (2010) found that 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), thus better enabling them to learn the primary point of the lesson—the underlying rationale for controlling variables.

However, it is possible that 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. We predicted this because compared to students with no incoming knowledge or engineering goals, these students would experience reduced cognitive load and likely learn more from the instruction on the rationale for controlling variables. 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).

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?	Scaffolding during p-solving?
Schwartz & Martin (2004)	9	Statistics	No	Y (contrasting cases)	Group	No	Maybe	Likely	Complex	No
Kapur (2009)	7	Math	Yes	No	Group, then Individual	No	Maybe	Likely	Complex	Yes
Lorch et al. (2010)	4	Science	(Likely some students do)	No	Group	No	Maybe	Likely	Constrained	Yes (filling out table)
Current studies	7	Science	(Some students do)	No	Individual	No	No	Y (multiple-choice responses)	Constrained	Yes (filling out table)

In the current studies, we tested these predictions (summarized in Table 2) and investigated the mechanisms of learning from engaging in preparatory activities. 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. Fourteen students (seven in each 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. 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), $F(1, 123) = 21.66$, $p < .001$. Therefore, the students that comprised the analyzed data were lower-performing than the general student population at this school.

Materials and Procedure: As shown in Table 3, 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 both designed and evaluated an experiment for a given focal variable (See Figs 1 & 2 for a design and evaluate question, respectively). 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 are 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 in which they designed one experiment for each of the four variables. After setting up each experiment, 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 presented with 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 were different 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; this, and other open-ended responses were not analyzed in the version of the TED tutor used in this study,

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

The tutor then controlled all of the 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:

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

(g) 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,

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

Students in both of the conditions 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 other than 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 3. Study design and procedure.

	Condition	
Period	Both	Instruction-only
1	Story pretest	Story pretest
2	Ramps pretest	Evaluate maximally-contrastive experiment
	Evaluate non-confounded (CVS) experiment	
	Evaluate singly-confounded experiment	
	Story posttest	

Immediately after the instructional phase, all students completed the computerized story posttest, identical to the story pretest and scored the same way. This assessed students' ability to transfer what they learned to other domains.

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, $F(1, 46) = 1.59$, $p = .21$.

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 so it was not included as a covariate in subsequent analyses.

Standardized science scores: Because the 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), $F(1, 39) = 8.25$, $p = .007$, and MCAS science scores were highly significantly related to story posttest score ($r(39) = +.45$, $p = .003$), they 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.

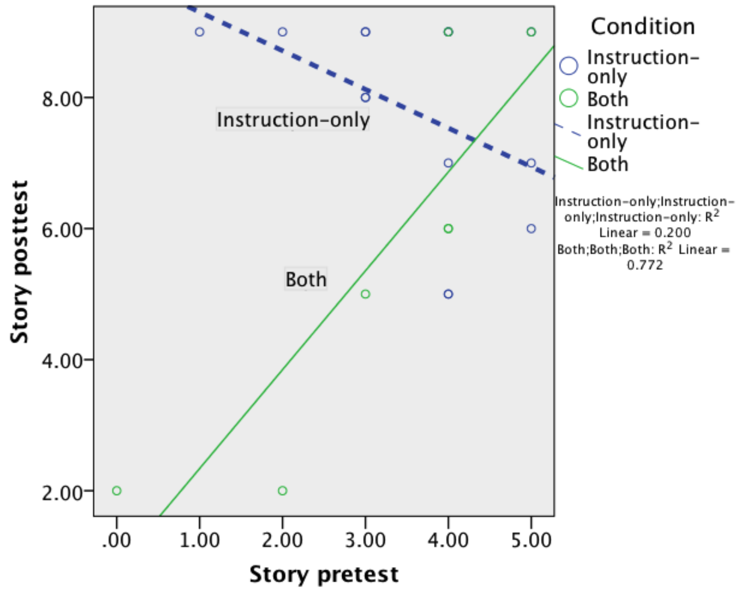
Initial knowledge of CVS: The predicted initial knowledge by treatment interaction was assessed by coding students' responses to the final item on the story pretest. This item required students to evaluate a non-contrastive set-up (i.e., the value for each variable was the same across conditions) and 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 the pretest. Student responses were classified as indicating "knowledge of contrasting focal variable" if they (a) responded that the experiment was *not* a good way to find out about the focal variable, (b) expressed the need to vary at least the focal variable, and (c) did so in order to "fix" the experiment.

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") or 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 post 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 initial knowledge of varying focal variable: We had 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 4, 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. 4 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 for only 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 is that students in the Both condition who scored low on the story pretest may have "solidified" their (perhaps recently-developed) understanding of the need to vary (at minimum) the focal variable during the ramps pretest. This is tentatively supported by analysis of students' story pretest and ramps pretest set-ups: although lower- and higher-story pretest students tended to over-apply the "compare and contrast" rule in their experimental designs on the story pretest, lowest-pretest 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, on average only 0.67 (SD = 1.15) of the higher-pretest students' ramps pretest set-ups were confounded, a significant difference, $F(1, 4) = 8.00$, $p < .05$. However, this result is tentative due to the low sample size. 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 of these students 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 of contrasting the focal variable: 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, $r(13) = +.38$, $p = .18$.

Then what *was* predictive of Both condition students' learning gains? In other words, 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, standardized science score, standardized 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 ($r(14) = -.66$, $p = .01$; $r(14) = +.51$, $p = .06$, respectively). In a backward regression with these two goal responses as independent variables, only the number of engineering goal responses remained in the model[1]. This result suggests that students' misinterpretations of the instructional goal during the ramps pretest—or failure to realize the science goal of the task—may have hindered their subsequent learning from 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 4 shows the frequency distribution for the number of engineering responses students selected during the ramps pretest. The majority of students (69%) selected at most one engineering response.

Table 4. 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
Number of students selecting X engineering responses	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 no more than one

[1] 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 goal response on the ramps pretest (to allow for one accidental engineering response)[2], 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), (11/16; 69%) $F(1, 18) = 3.28$, $p = .087$ (refer to Fig. 5). Thus, the predicted advantage of completing the ramps pretest prior to instruction is supported for students who did *not* misinterpret the goal of instruction more than once. In contrast, Both condition students with no initial CVS knowledge who selected an engineering response on *at least* two (5/16; 31%) 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, for students who misinterpreted the goal of the pretest task more often or who failed to realize the nature of the task, there may be an advantage of an additional round of instruction over completing the ramps pretest.

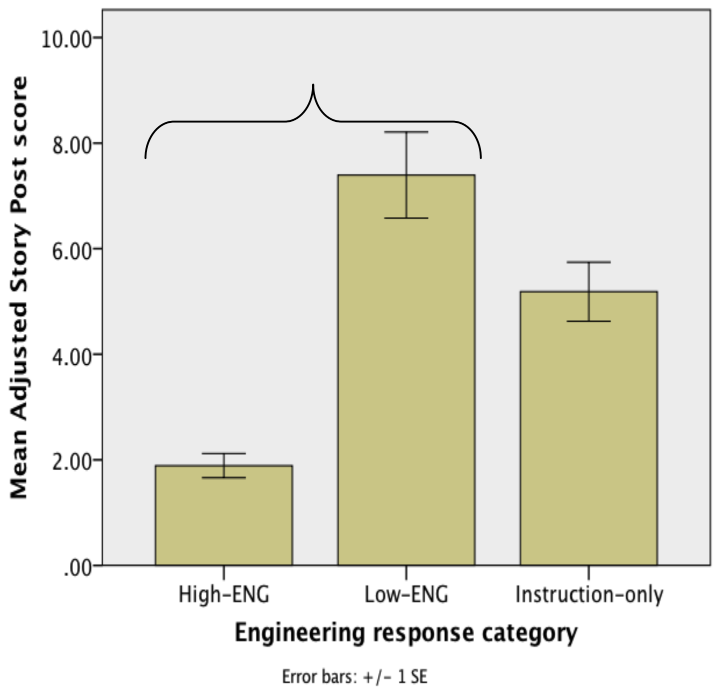


Fig. 5 Mean adjusted story posttest score by condition and response type.

The comparisons between the two Both groups and the Instruction-only group are suspect because we do not know 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 other words, are students in the Instruction-only condition more comparable to the "High-Engineering" or "Low-Engineering" students in the Both condition? The

[2] 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, $F(1, 13) = 3.49$, $p = .085$.

Instruction-only group was more similar to the "Low-Engineering" students in the Both condition in terms of story pretest score (*M* = 2.00, (SD = 1.53); *M* = 2.10 (1.52); *M* = 1.00 (0.82), respectively), and in terms of science achievement scores (*M* = 239.67 (SD = 16.55); *M* = 228.20 (12.49); *M* = 224.00 (8.72), respectively). Thus, these students likely 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 suggests there might have been an advantage of completing the ramps pretest prior to entering the instructional phase for students who correctly interpreted the goal of the activity.

Table 5. Study 1 results summary: Outcome by student characteristics.

Initial knowledge state	Student characteristic	Outcome
No initial knowledge	High-Engineering ramp pretest	Instruction Only > Both
	Low-Engineering ramp pretest	Both > Instruction Only
Initial knowledge of contrasting target variable	Higher-story pretest	Instruction Only = Both
	Low-story pretest	Instruction Only > Both

In summary, these results suggest that 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. However as shown in Table 5, of these "no-initial-knowledge" students, those who less often misinterpreted the goal of the ramps pretest tended to benefit more from the preparatory task than those students who misinterpreted the goal of the task. And there appears to be an advantage of first completing the initial ramps pretest activity over instruction-only for students who were less likely to misinterpret—or who realized—the goal of the activity. For students who showed 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. Lowest-knowledge students benefited more when they did *not* complete the preparatory activity prior to instruction.

Study 2 Methods

Because of the small sample sizes for some analyses in Study 1, we replicated it the following year in the same school. Study 2 procedure replicated that of Study 1 (refer to Table 3 for the procedure), 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 this study, there was more time between the 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.

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 low-performing relative to this general student population: 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 (M = 46.69, SD = 28.65; M = 84.93, SD = 16.32 for reading; M = 53.90, SD = 26.87; M = 86.91, SD = 11.56 for math). On average, these students were about average among all students who completed the MAP testing (but lower-performing among their school peers).

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). Because MAP standardized math percentiles were the most recent standardized test data available and because they were most highly correlated with story pre- and posttest scores, they were included as covariates (as a general ability measure) in subsequent analyses.

Time on task: Students in the Instruction-only condition took 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 so 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, $F(1,63) = 0.03$, $p = .88$). 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 post with covariates: In an ANCOVA including students' initial knowledge (i.e., whether or not they expressed an understanding of contrasting at least the focal variable on the final story pretest item) and condition as fixed variables and covarying for MAP math percentile scores and story pretest scores, the predicted condition by initial knowledge interaction could not be assessed because there was a significant condition by initial knowledge by math score interaction, $F(1, 34) = 6.65$, $p = .01$. Because of this 3-way interaction, the predicted condition by initial knowledge interactions were investigated separately for higher- and lower-ability students.

Higher-ability students. For students with higher math percentile scores (i.e., who scored at or above the 60^{th} percentile), there was a significant condition by initial knowledge interaction, $F(1, 14) = 5.35$, $p < .05$. This interaction was in the expected direction (Fig. 6). Students with some initial knowledge had significantly higher scores in the Instruction-only than Both condition (M = 8.86, SD = 0.38; M = 6.71, SD = 1.98, respectively), $F(1, 12) = 6.93$, $p = .02$[3]. Perhaps due to the small number of students who were higher-ability but nonetheless did not express this basic knowledge of contrasting the focal variable, though the trend was in the expected direction, there was no significant advantage of the Both condition, $F(1, 4) = 0.46$, $p = .54$. Because of this small sample and because no students in the Both condition selected more than two engineering responses on the ramps pretest (and thus there were no "high-engineering" student in this sample), it was not possible to assess the relationship between engineering goal response frequency (high/low) and story posttest performance, as was done in Study 1. Students with no initial knowledge of contrasting the focal variable in the Instruction-only condition had similar standardized math and science scores as Both condition students. Thus, they likely would have also been "low-engineering" students had they been in the Both condition. These results are consistent with the Study 1 results for low-engineering students, who benefited more in the Both condition.

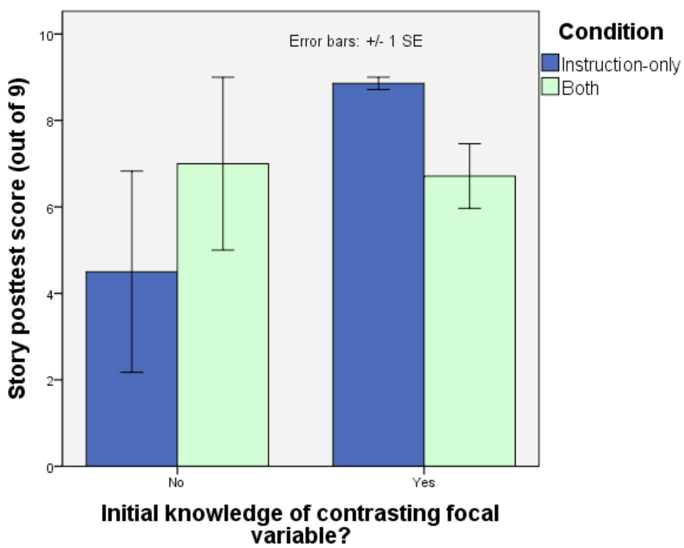


Fig. 6 Story posttest score by condition and initial knowledge level (higher-ability students).

Lower-ability students: There was no significant condition by initial knowledge interaction for the lower-ability students, $F(1, 23) = 1.35$, $p = .26$, and no main effect of condition (Instruction-only: M = 4.43, SD = 2.50; Both: M = 3.85, SD = 2.41), $F(1, 24) = 0.15$, $p = .70$. There was, however, a main effect of initial knowledge, $F(1, 24) = 5.66$, $p = .03$. Students who expressed an understanding of varying at least the focal variable on

[3] Unlike in Study 1, there was no condition by story pretest interaction ($p = .98$). Nor was there a significant effect of story pretest ($p = .75$). These results may be due to the smaller range of story pretest scores for these higher-ability students in Study 2, where no students scored less than two.

the story pretest scored significantly higher on the story posttest than those who did not (M = 5.33, SD = 2.50; M = 3.20, SD = 1.97, respectively).

But, unlike Study 1, for just those students who did not express an understanding of varying the focal variable on the story pretest, there was no evidence that the development of a science goal perspective during the ramps pretest task was associated with learning. That is, there was no significant relationship between the frequency of students' engineering or science goal selections on the ramps pretest and story posttest score. Nor were there significant relationships between posttest score and other plausible factors (including the number of unconfounded experiments students designed on the ramps pretest, the number of complete explanations students selected, and the number of responses indicating at least the need to compare variables across conditions). Even the number of correct open-ended responses these lower-ability students gave during the interactive instruction – which was inversely correlated with the number of engineering responses students gave on the ramps pretest – was not related to story posttest score.

However, the best predictor of the number of correct open-ended responses students gave during the instruction was the frequency of engineering and no-goal responses – or, inversely, the number of science goal responses – students selected during the ramps pretest ($r(7) = -.84$, $p = .02$)[4]. MAP math percentile score was also significantly related ($r(7) = +.81$, $p = .03$). When these factors were entered into a multiple regression model, both factors remained significantly related to the number of correct responses students gave during the instruction. Thus, analogous to Study 1, where understanding the ramps pretest task goal was the best predictor of story posttest performance, understanding the science goal nature of the task appeared to have the strongest influence on instructional performance in Study 2.

Discussion

The goals of these studies were (a) to uncover the conditions in which completing an initial preparatory activity in which no feedback was provided 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 that affected whether completing this activity prior to an interactive instruction was more beneficial to learning than if this time had been spent in instructional activities. Participants were seventh-grade students in a primarily middle-to-high SES population who likely had some prior instruction in experimental design. Because Lorch et al. (2010) found that students completing an initial ramps design activity developed the idea of contrasting the focal variable across ramps—an idea that was relevant to, but not the focus of the later instruction— we predicted a condition by initial knowledge interaction. Specifically, we predicted that students with no initial understanding of the need to "compare and contrast" at least the focal variable would benefit more from first setting up experiments on a "ramps pretest," provided this led to the application of this intuitive understanding to the task. In contrast, we expected that students who already had this understanding would

---

[4] The number of engineering and no-goal responses on the ramps pretest was a better predictor of instructional answer quality than: (a) story pretest score, (b) the number of unconfounded experiments students designed on the ramps pretest, (c) the number of responses indicating at least knowledge of contrasting variable values, and (d) the number of complete CVS explanations students selected.

benefit more from the interactive instruction, which emphasized the underlying logic of CVS – an understanding of which has been found to correlate with transfer (Siler et al., 2010). Consistent with our prediction, there was a significant interaction of condition by knowledge of "comparing and contrasting" at least the focal variable for all students in Study 1 and for the higher-ability students in Study 2.

However, in both studies, no overall advantage to completing the preparatory activity was found among students with no initial knowledge of varying the focal variable or other aspects of CVS (many of whom may have held misconceptions about the goal of the task). Moreover, in Study 1, 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 is not consistent with our hypothesis that the development of initial knowledge components of CVS was the underlying mechanism of learning from the preparatory activity. Rather, in an exploratory analysis, the number of "engineering" goal responses on the ramps pretest was found to be the only one among several plausible factors considered that correlated with learning outcome, and this correlation was negative. This suggests the possibility that it is the development of a more general framing of the activity rather than specific knowledge structures that accounted for the benefit of the preparatory activity for the primarily middle-to-high-SES seventh-grade student participants. Likewise, in Study 2, the frequency of science goal selections during the ramps pretest was the best predictor of students' response accuracy during the interactive instruction[5]. However, further support for this possibility is necessary due to the small sample size of these post-hoc analyses. If found, providing some form of instructional support to help students understand the nature of the task (and subsequent lesson) may make completing the ramps pretest task more beneficial for lower-knowledge students.

In spite of these results, it is still possible that the fourth-grade students in the both condition of the Lorch et al. (2010) study benefited from the development of an initial understanding of the need to contrast at least the focal variable in conjunction with an understanding of the goal of the task. One factor that may account for the different results of Study 1 and the Lorch et al. study is the ages of students; 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)[6]. 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. Alternatively, consistent with the results

---

[5] However, there was no significant relationship between goal responses or other plausible predictors and posttest performance. It is possible that the knowledge structures developed by these lower-ability and low-knowledge students during the ramps pretest and the relatively brief instruction were unstable over the one-day interval between instruction and the posttest. That is, noise due to subsequent learning and forgetting may account for these non-significant correlations.

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

of the current studies, it is possible that the lower-ability students in the Lorch et al. study, who benefited more from the initial ramps group activity than their higher-ability counter-parts, also did so because they learned or realized the science goal of the activity during the task (without necessarily having learned to compare and contrast the focal variable). And the higher-ability students may have benefited less from this activity because they already understood the nature of the task.

For students who *did* show an understanding of the need to contrast variables across the experimental conditions (and thus applied science goals), there was a condition by story pretest score interaction in Study 1. In Study 2, there was a significant advantage of the Instruction-only condition (for the higher-ability students) but no condition by story pretest score interaction[7].

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 story pretest score). In contrast, the lowest-knowledge students in the Both condition performed worse than their counter-parts in the Instruction-only condition. Evidence suggests that the lowest-knowledge students reinforced their knowledge of comparing and contrasting and applied this to variables other than the focal variable during the ramps pretest. During instruction, this understanding seemed to have framed how students interpreted 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," only considering the contrasting aspect of CVS and not the controlling aspect. This result brings up the possibility that—at least in some cases—students with weak developing knowledge may not benefit from an initial preparatory activity—or at least one that does not include explicit feedback. Instead, it may be better to either provide such students with feedback on their responses or to provide instruction immediately, without an initial preparatory activity, before an overly simplified conception can be made robust through practice. However, further research is necessary to determine the extent to which the development of over-simplified conceptions during preparatory activities is an obstacle to learning.

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 provided the learning advantage of the Both condition for some students (rather

[7] It is possible no interaction was found in Study 2 because there was a smaller range of story pretest scores for these higher-ability Study 2 students—none of whom scored less than two out of nine.

than a task-specific factor). 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.

In Study 2, the predicted condition by initial knowledge interaction held only for the higher-ability students. This suggests the possibility that outcomes for other student populations, such as younger students and lower-achieving students, may not show the predicted interaction pattern. As suggested by the results of Study 2 vis-à-vis Study 1, this predicted pattern may not hold for such student populations, especially when there are longer delays between instruction and the posttest[8]. Thus, future studies are necessary to determine whether the same patterns of results hold for different populations of students, for varying study procedures, and even interactions between these factors. By investigating questions such as these, we can continue to determine the boundary conditions in which engaging students in an initial preparatory activity 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. 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.

[8] Recall that in Study 1, the story posttest was given immediately following the instruction, whereas in Study 2 it was given the next day.

References

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.

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.

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.

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.

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

Fig. 1 Screenshot of story pre/posttest design item

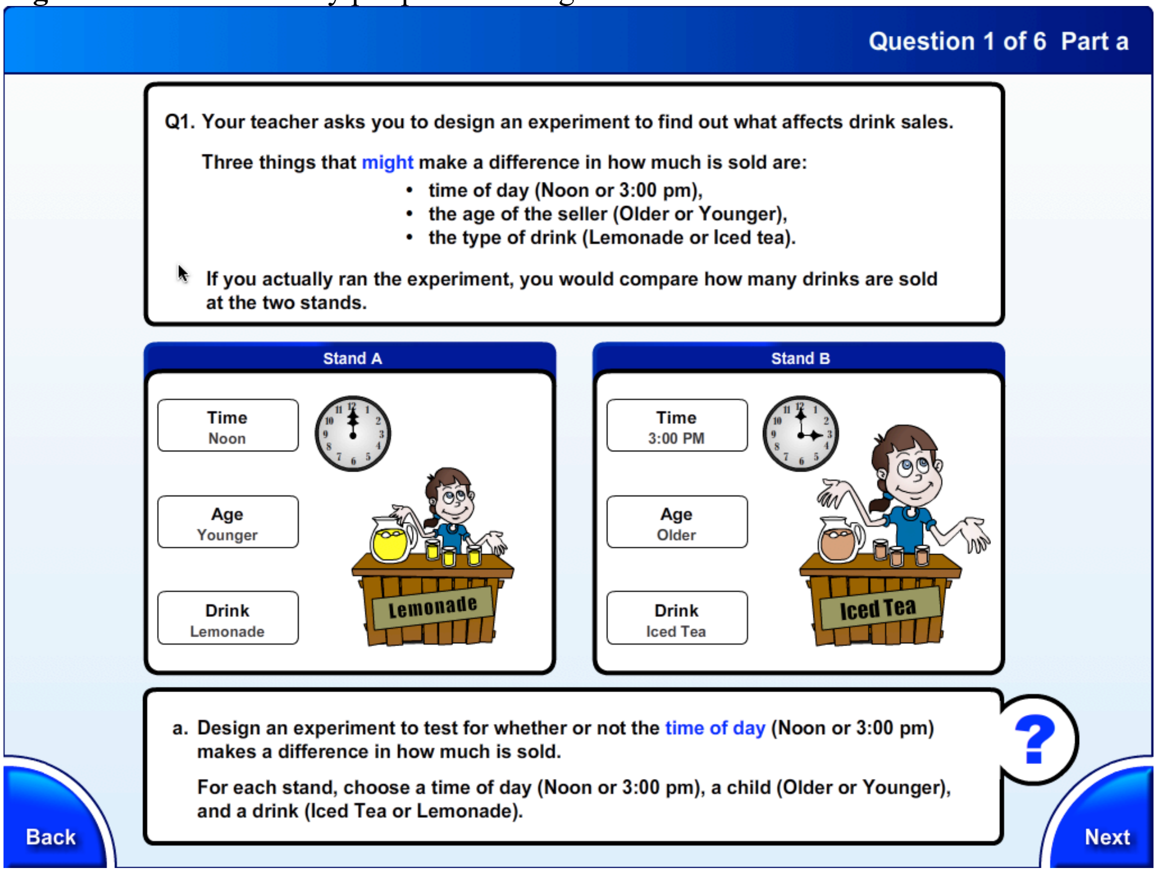


Fig. 2 Screenshot of story pre/posttest evaluate item

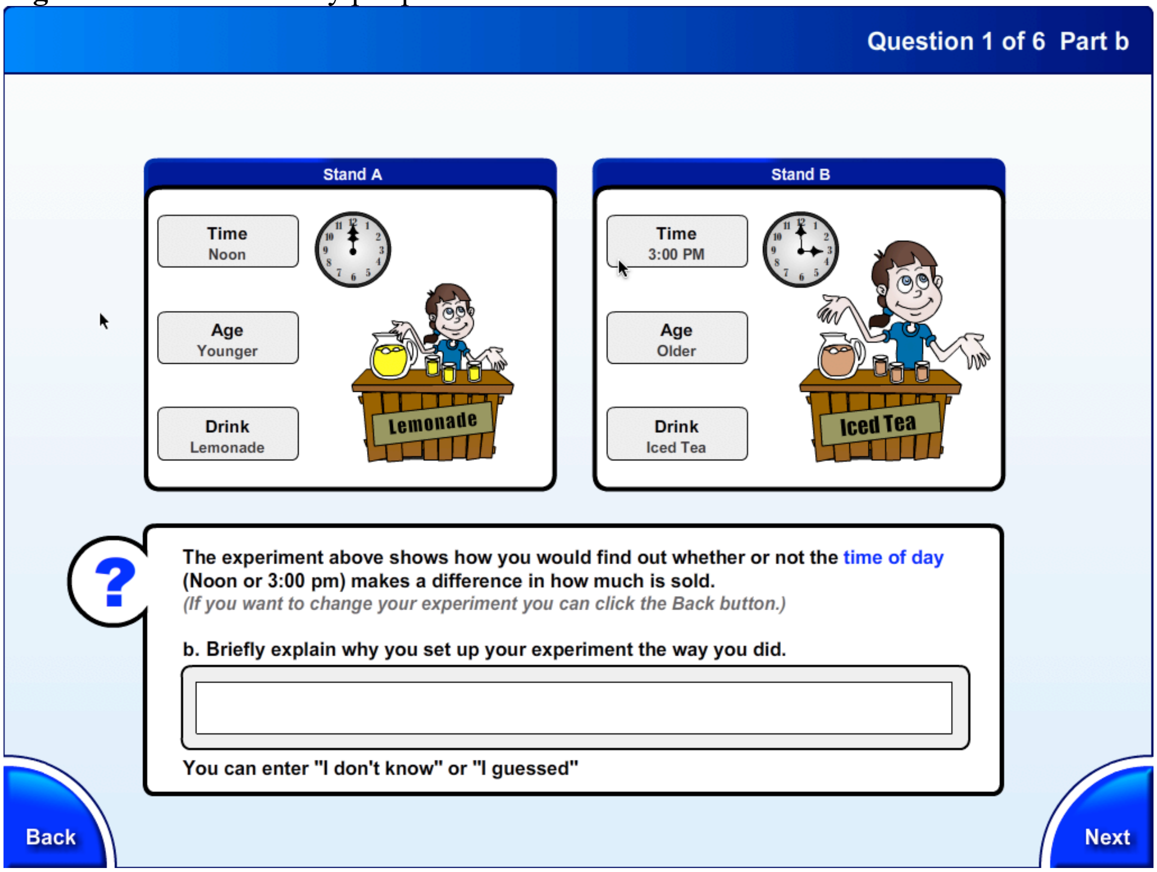


Fig. 3 Screenshot of introduction to ramps apparatus prior to instruction