Promoting the Learning and Transfer of Basic Experimental Design Skills: Individual Differences in the Effects of Instructional Framing

Stephanie Ann Siler & David Klahr

Carnegie Mellon University

This research was supported in part by grants from the Institute of Education Sciences (R305A100404 and R305B090023), and in part by a grant from the National Science Foundation (SBE-0836012). We thank Kevin Willows and Cressida Magaro for their invaluable contributions on many aspects of this project. We also thank the participating school and especially the teachers who allowed us to conduct research in their classrooms.

Correspondence concerning this article should be addressed to Stephanie Siler, Psychology Department, Carnegie Mellon University, Pittsburgh, PA 15213.

E-mail: siler@cmu.edu

THIS IS A DRAFT; DO NOT CITE WITHOUT PERMISSION FROM AUTHORS.

Abstract

Although the ability to reason scientifically is critical for developing valid knowledge about the world, elementary and middle school students often have poor understandings of aspects of scientific reasoning such as designing controlled experiments. Providing explicit instruction focused on the rationale for controlling variables has been found to promote late-elementary and middle-school students' learning and transfer of the control of variables strategy (CVS) to novel domains; however, outcomes of younger, lower-ability, and lower-SES students are much poorer. In this study, we investigated whether alternative "framings" of the CVS instruction would better support outcomes among different student populations (low-SES fifth-grade vs. middle-SES sixth- and seventh-grade students), initial knowledge, and ability levels. We compared a typical "Science Experiment" framing to instruction framed as either making "fair comparisons" or "solving brain teasers." The Brain Teaser framing better supported transfer among younger/low-SES students but was associated with poorer outcomes among higher-ability students in the older/higher-SES population. The Fair Comparison framing induced misinterpretations among the younger/lower-SES population; however, when contrasting the focal variable was emphasized in a follow-up study, benefits of the Fair Comparison framing were found. These results suggest that lower-SES students in particular may benefit from instruction using such alternative terminologies.

Keywords: instructional framing, transfer, science inquiry

Promoting the Learning and Transfer of Basic Experimental Design Skills: Individual Differences in the Effects of Instructional Framing

The ability to reason scientifically is critical to developing valid knowledge about the world, and the design of "good" experiments is one of the fundamental components of scientific reasoning. The importance of experimental design is reflected in a wide range of science policy documents and assessments. For example, the National Science Education Standards (NSES, 1996) state that there is a "need for students to participate actively in designing experiments, selecting tools, and constructing apparatus, all of which are critical to the development of an understanding of inquiry" (p. 45), and "students should be able to review data from a simple experiment, summarize the data, and form a logical argument about the cause and effect relationships in the experiment" (p. 145). With respect to high-stakes science assessments, the 2009 National Assessment of Educational Progress (NAEP) includes extended computer-based interactive tasks in which students must design, execute, and interpret the outcomes of experiments. More recently, in the Next Generation Science Standards (2013), six of the eight science and engineering practices students are expected to engage in, beginning in kindergarten, are directly related to activities involved in formulating, executing, and evaluating experiment-based science projects. In general, the ability to reason about experiments and what they can tell us about the world is important for a well-informed populace.

One aspect of experimentation is designing controlled experiments using the Control of Variables Strategy (CVS). When applying CVS, one controls variables other than the one under investigation (the focal variable). Although CVS is a seemingly simple procedure, elementary and middle school students often perform poorly on tasks requiring them to design controlled experiments and evaluate the validity of experimental comparisons (e.g., Author, 1999; Bullock & Ziegler, 1999; Kuhn, Garcia-Mila, Zohar, & Andersen, 1995; Schauble, 1996).

We investigate whether the contextualization—or framing—of instruction on experimental design affects learning and transfer outcomes, and whether certain ways of framing instruction are better suited to students from different populations, initial knowledge, and ability levels. Our overarching *theoretical* question was which of two general instructional approaches—activating knowledge schemas or prompting more effortful problem-solving—better supports students' learning and transfer. Our practical goal was to find ways to improve learning and transfer outcomes of lower-ability students, lower-knowledge students, and/or students from lower-SES populations. To do this, we compared instruction on CVS using a typical "Science Experiment" framing to "Brain Teaser" and "Fair Comparison" framings. We hypothesized that the Brain Teaser framing would prompt more effortful problem-solving that would support understanding (H1) and that the Fair Comparison framing would support understanding primarily by eliciting relevant knowledge schemas (H2).

In what follows, we first discuss individual differences in learning CVS from instruction, then students' difficulties understanding different aspects of CVS. Then we discuss the specific framings used in the present study and present evidence from a survey we administered to inform our predictions of individual differences in the effect of framing condition.

Differential Learning of CVS

In schools serving primarily middle- to upper-SES student populations, providing explicit instruction in which students evaluate experimental comparisons then hear explanations for why the given design was or was not informative leads to high rates of transferring experimental design skills to other domains (Author, 1999; 2000; 2004; 2008; 2010). However, CVS learning and transfer rates differ across student SES level. For example, in Author (2005), low-SES students' post-instruction mastery rates for designing experiments in novel domains were only about one-fifth those of higher-SES students. In Lorch et al. (2010), fourth-grade students from lower-achieving schools serving predominately lower-SES student populations who were taught CVS in interactive classroom discussions performed significantly worse on a posttest requiring evaluation of experiments across domains than their peers in higher-achieving schools serving predominately higher-SES student populations. Students from lower-SES schools answered fewer than 60% of the evaluation questions correctly (chance was 50%). Moreover, unlike students from higher-SES schools, having the opportunity to design experiments in small groups before instruction did not improve posttest performance among lower-SES populations.

Sources of Difficulties Learning CVS

Knowledge schemas. Our analyses of student explanations further revealed frequent misunderstandings of the goal of the experimental design tasks (Author, 2013). As found previously (e.g., Schauble et al., 1991; Schauble, 1996; Tschirgi, 1980), students often misinterpreted the instructional goal as to *engineer* a particular outcome, often an optimal outcome (e.g., to make balls roll as far as possible),[1] rather than as learning to design experiments that will allow pursuit of a "science goal" of *finding out about* a particular variable. Such engineering goals are more commonly expressed in younger and lower-SES populations (Author, 2013). Thus, younger and lower-SES students may have weaker associations between experimentation and pursuing science goals.

[1] Even young children appear able to apply science goals in certain scenarios. For example, Sodian, Zaitchik, and Carey (1991) found that first- and second-graders can differentiate between engineering and science goal applications in simple tasks. Thus, it is likely that late-elementary and middle-school students understand and can apply science goals (Authors, 2013)—but have difficulty realizing the relevance of science goals in a given context.

Reasoning. Once students understand the goal of the task as setting up informative experiments, they generally simultaneously demonstrated an understanding of the need to contrast values of the variable(s) they were comparing (Author, 2012). This is consistent with Ford (2005), who commented on students' surprise that contrasting values of the focal variable "would ever be an issue" (p. 459) in designing experiments. Only after expressing knowledge of the task goal and contrasting the focal variable(s) did (some) students start controlling other variables in their designs. Students' explicit statements of the rationale for controlling variables generally occurred only after they first set up a controlled experiment. That is, students demonstrated procedural knowledge of controlling variables before a conceptual understanding of the rationale for controlling. Thus, understanding why controlling variables is necessary appears to be the most difficult aspect of CVS.

This may be expected, since understanding the rationale for CVS entails: (a) considering that other (non-focal) variables may affect the outcome, (b) realizing the indeterminacy of the cause(s) of a hypothetical effect if other variables are contrasted, and (c) concluding that, because of this indeterminacy, the other variables must be controlled. Students must be able to construct or at least follow this logical argument in order to understand the rationale for controlling variables. In Author (2010), students' deductive reasoning performance was more predictive of whether they expressed this rationale during instruction than other measures, including reading comprehension and science achievement. In turn, expressions of the rationale for CVS strongly predicted transfer.

Metacognition. Another possible source of difficulty appears to be metacognitive (see Author, 2013). Following in-class interventions, students often report that they already knew the instructional material. For example, among sixth-grade students who demonstrated no initial knowledge of CVS on a pretest, 67% later claimed that they knew the material prior to instruction (all students who showed CVS understanding on the pretest said they already knew the material). If students do not realize they do not understand, they may fail to exert the effort necessary for learning.

Instructional Framing

Prior research has shown that framing instruction in familiar terms that elicit learners' relevant conceptions—including knowledge schemas—can help compensate for knowledge or reasoning-related weaknesses (e.g., Anderson & Pearson, 1984; Bransford & Johnson, 1972; Cheng & Holyoak, 1985; Cheng & Holyoak, 1989; Dijksterhuis & Van Knippenberg, 1998; Marshall, 1995; Tang & Baumeister, 1984; Tse et al., 2007; Webster & Martocchio, 1993). Schemas are defined as "mental structures" that "summarize that which is common to a large number of… situations" (Anderson, Spiro, & Anderson, 1978, p. 434). They include specific knowledge and can be conceptualized as having "slots or placeholders that can be instantiated with certain particular cases" (p. 434). For example, "dining-at-a-fancy-restaurant" schemas commonly involve a "slot" for ordering an entrée, X, the entrée ordered in a specific situation.

Furthermore, framing instruction on experimentation may affect learning outcomes (Ford, 2005; McElhaney & Linn, 2010). Ford (2005) compared instructional activities focused on CVS (similar to the explicit instruction described previously) to instructional activities focused on measuring experimental outcomes in classrooms of primarily Caucasian sixth-graders at a suburban middle school. Students in the CVS condition performed better on a posttest assessing experimental design given the day after the unit ended; however, students in the Measurement condition performed better on other aspects of experimentation, including quantifying experimental outcomes and conducting multiple experimental trials. In McElhaney and Linn (2010), the wording of the research questions high school students investigated in a simulated environment was manipulated in ways that increased the scope of the investigation (e.g., "Does the driver's height make the biggest difference in whether the driver is injured?" vs. "Are tall or short drivers more likely to be injured by a deploying airbag?"). Among the highest-knowledge students, those given the more general question form (the first above) adopted more sophisticated investigation strategies and drew more nuanced conclusions. In these studies, the framing affected student outcomes, and in the latter study, the effect of framing depended on students' prior knowledge.

Framings compared in the present study

The instructional framings in the present study were chosen in part for their potential to compensate for weaknesses in initial knowledge (or goals), reasoning, and/or metacognitive skills. The baseline was a "Science Experiment" framing, typically used in science textbooks, in which the goal of the lesson was described as "learning to design good experiments." This "Science Experiment" (SE) framing and associated terminology was used in our prior studies (Authors, 1999, 2000, 2004, 2008, 2010). We compared this framing to "Fair Comparison" (FC) and "Brain Teaser" (BT) framings.

Fair Comparison. In the "Fair Comparison" framing, the goal of the lesson was described as "learning to make comparisons that are fair to what you are comparing." This framing was in part motivated by frequent usage of the term "fair" in many science textbooks (e.g., Scott Foresman Science; FOSS Variables, 2000) and elsewhere (e.g., in the Next Generation Science Standards (2013), controlled experiments are referred to as "fair tests").

Students are likely familiar with a wide range of situations involving "fair comparisons" that are analogous to controlled experiments. These include competitive scenarios (sports or games), in which all competitors abide by the same rules. For example, in a race to a finish line, it would not be "fair" if one runner started closer to the finish line or had any other notable advantages. Thus, we expected that students would be familiar with this schema.

In addition, students are likely to separately understand "fair" and "comparisons." Previous research has found that young children (three and four-year-olds) exhibit an understanding of fairness (Olson & Spelke, 2008), and even babies (between 19-21 months) appear to expect resources (e.g., toys or food) to be divided equally between people (Sloane, Baillargeon, & Premack, 2012). By age nine, children understand "fair" as "numerically identical sharing of goods" among different individuals (p. 6, Castelli et al., 2013). Students likely also generally understand the idea of comparing different entities or ideas (e.g., they are often asked to "compare and contrast" various entities or concepts in schoolwork).

We also chose this framing for its potential to support students' understanding of the lesson goal—which they may otherwise interpret within an engineering framework. That students' knowledge schemas related to "fair comparisons" could be leveraged to support their understanding of experimental design is suggested by research showing that eliciting schemas that include knowledge components related to the critical—and often more logically complex—features of the task can support performance. For example, in the seminal research of Cheng and Holyoak (1985; 1989), participants asked to verify a given abstract rule of the form "If A, then B" in a Wason card selection task correctly attempt to confirm the rule by selecting cases of "A" (on the visible side of a card) and verifying correspondences with "B" (on the reverse side of the card). However, they generally fail to check for "rule breakers," or "not Bs," to ensure they do not correspond with A. When this problem is reframed in terms of an intuitive "permission" schema (e.g., “If a person is drinking alcohol, they must be at least 21”), people are much more likely to check for rule breakers (i.e., those under 21 who are drinking).

As permission schemas support the most difficult aspect of the Wason card task, *fair comparison* schemas may support the most difficult procedural aspect of CVS—controlling variables—by prompting the checking of factors that may give one actor in the comparison an advantage over the other. Although we expected fair comparison schemas would provide a knowledge framework that could support students’ understanding of CVS, the success of this framing hinged on students interpreting instruction within this schema. If students’ “fair comparison” schemas are elicited during instruction, then following the logical argument underlying the rationale for controlling variables may not be necessary to understand the rationale for CVS. Thus, lower-ability students may benefit from the FC framing.

Brain Teaser. In the “Brain Teaser” framing, the goal of the lesson was described as “learning to solve some brain teasers.” Because brain teasers encompass a wide range of problem types (e.g., riddles, mysteries, visual puzzles, logic problems), unlike for Fair Comparisons, we did not expect students to hold schematic knowledge of brain teasers, including specific solution procedures that would support initial comprehension and reasoning.

However, as discussed previously, students often failed to realize that they did not understand the instruction, suggesting that metacognitive processes were at least one underlying cause of learning failures. Previous research suggests that activating certain conceptions may improve performance outcomes by eliciting metacognitive processes (e.g., Adam & Galinsky, 2012; Dijksterhuis & Van Knippenberg, 1998). For example, participants asked to think about traits of a professor out-performed participants asked to think of traits of a “football hooligan” on a test of general knowledge (Dijksterhuis & Van Knippenberg, 1998). In another study (Adam &

Galinsky, 2012), participants either wore a coat described as a “doctor’s white lab coat,” the same coat described as a “painters’ white coat,” or could see (but did not wear) the same “doctor’s” coat. Participants who wore the “doctor’s lab coat” performed better than participants in the other conditions on both a Stroop task that required selective attention and on a comparative visual search task that required sustained attentional focus. Thus, activating knowledge associated with “doctors” apparently affected metacognition by improving aspects of attentional control. We chose a “brain teaser” framing in an attempt to elicit expectations of a difficult task, which would prompt more effortful problem-solving behaviors that appear necessary for developing an understanding of the rationale for controlling variables.

To our knowledge, there is little research on the effects of using brain-teaser-like framings. Solving brain teasers also frames instruction as a game-like task. Although there has been a recent reemergence in incorporating game-like elements within educational software, there is little research comparing learning from game-like framings to more standard instructional framings (Rai & Beck, 2012). Furthermore, because of limited research on students’ conceptions of brain teasers (and “fair comparisons”), we administered a survey to verify our assumption that students consider solving brain teasers to be difficult and to gain further insight into students’ beliefs regarding brain teasers (as well as science experiments and fair comparisons). Research in social psychology (cf. Adam & Galinsky, 2012; Dijksterhuis & Van Knippenberg, 1998) suggests that if students do associate solving brain teasers with more effortful engagement—which, as discussed in the next section, is suggested by the framing survey results—then framing instruction as solving brain teasers may prime related behaviors.

Significant findings from survey

Thirty-one sixth-graders (18 from lower-SES and 13 from higher-SES families) at a local science and technology magnet school serving grades 6-12 completed the survey (shown in Appendix A), consisting of initial open-ended (constructed-response) items and Likert ranking items. Students had transferred from various schools into this magnet school at the start of sixth grade. The science teacher's ratings of individual students were used as the measure of students' general ability.

Open-ended responses. The majority of students (90%) indicated that the purpose of doing an experiment was to find out about something (e.g., "To find the answer" "See what happens and how"). However, higher-SES students were marginally more likely to indicate science goal associations than lower-SES students (100% vs. 83%, respectively) (p = .06). Most students (58%) indicated that the purpose of doing a brain teaser was to promote thinking (e.g., "Challenge your mind" "Exercise your brain" "Make you think"). The next most common response type (23%) was a science goal (e.g., "To find the answer and to do the problem"). However, there was no relationship between SES and response type (p = .28). There was a more diverse set of responses defining fair comparisons, including responses indicating a science goal, the idea of controlling, or a full CVS response indicating that only the variable one is testing should be contrasted (Table 1).

Table 1
Frequency of "Fair comparison" response type by SES

	Response type					
SES	Compare/ Science goal	Control	CVS	Social	Don't know	Other
Lower	3	3	1	3	3	5
Higher	3	3	2	0	1	4

Overall, 29% of students expressed the idea of controlling variables, and there was no difference across SES level. However, there were no references to controlling in *any* explanations of science experiments or brain teasers. This supports that the fair comparison framing will best support the most difficult procedural aspect of CVS—controlling variables.

Likert ratings. Mean ratings per item are shown in Table 2 (see Appendix A for the survey). Because in the present study we compared the Science Experiment with a Brain Teaser framing and the Science Experiment with a Fair Comparison framing, we also consider how students respond to analogous items across framings (e.g., Doing science experiments is fun" vs. "Solving brain teasers is fun") to further inform our predictions. To investigate SES by framing interactions, we conducted repeated measures analyses of covariance (ANCOVAs); students' ratings were the dependent variable, SES a between-subjects factor, and student ability was covaried. We also investigated within-SES differences across analogous framing items and between-SES differences for each item (however, only findings that included cross-framing differences—and thus a baseline (SE) framing for comparison—informed our predictions).

Table 2

Mean Likert rating (1-strongly disagree to 5-strongly agree) per survey item by student SES

		SES level	
Survey item	All students	Lower	Higher
4. Solving BT hard	2.98 (0.76)	3.19 (0.75)[*]	2.64 (0.63)[*]
5. Doing science experiments hard	2.44 (0.81)	2.39 (0.85)	2.43 (0.76)
6. Fair comparisons hard	2.87 (1.12)	3.12 (1.05)	2.46 (1.13)
7. Think solving BT	3.68 (1.00)	3.81 (1.02)	3.50 (1.02)
8. Think in science class	3.66 (0.90)	3.78 (0.88)	3.50 (0.94)
9. Concentrate in science class	4.12 (0.78)	3.94 (0.80)	4.29 (0.73)
10. Concentrate solving BT	**4.18 (0.77)** [a]	4.00 (0.69)	4.36 (0.84)
11. BT fun	**3.64 (1.19)** [a]	3.33 (1.19)	**3.93 (1.14)** [a]
12. Science experiments fun	4.09 (1.21)	4.22 (1.17)[b]	3.93 (1.33)
13. Learning science fun	3.82 (1.01)	3.61 (1.15)	4.00 (0.78)
14. Science experiments find out	4.42 (0.75)	4.39 (0.78)	4.43 (0.76)
16. BT find solution	3.91 (0.96)	4.00 (0.94)	3.71 (0.99)
18. Science experiments exciting (no corresponding BT item)	3.53 (1.08)	4.00 (0.94)[**]	3.07 (1.00)[**]
19. BT require reasoning	3.94 (0.86)	3.94 (0.87)	3.93 (0.92)
20. Learning science reasoning	4.12 (0.78)	4.17 (0.79)	4.07 (0.83)
21. Designing experiments reasoning	4.09 (0.88)	4.17 (0.86)	4.00 (0.96)

$^{*}p < .05$ $^{**}p < .01$ $^{***}p = .001$ [a] Significant positive correlation with ability.
[b] Significant negative correlation with ability. Within-SES squares indicate significant differences in means across framing items. Cross-SES squares indicate differences across SES-level. Cross-SES/framing item squares indicate significant SES-level by framing interactions.

Significant outcomes are reported below; those that also informed our predictions are italicized.

- Lower-SES students rated "solving brain teasers" as more difficult than higher-SES students (p = .04); however, there was no cross-SES difference in ratings of the difficulty of "doing science experiments" (p = .99). *The SES by framing (BT vs. SE) interaction was significant, p = .045.*
- *Lower-SES students rated solving brain teasers as more difficult than doing science experiments (p < .001),* but higher-SES students did not.

- *Lower-SES students also rated "doing science experiments" as more fun than "solving brain teasers" (p = .003);* however, there was no difference among higher-SES students.
  The SES by framing interaction was marginally significant (p = .06).
- *Lower-SES students rated "making fair comparisons" as more difficult than doing science experiments (p = .02);* higher-SES students did not, and the SES-level by framing interaction was not significant, $p = .14$.
- *Higher-SES students more strongly agreed that the goal of science experiments was to find out about something than solving brain teaser was (p = .03);* however, there was no difference among lower-SES students, and the interaction was not significant.
- Lower-SES students more strongly agreed that "Doing a science experiment is about trying to make an interesting/exciting result" than higher-SES students ($p = .009$). This suggests that lower-SES students may more strongly associate engineering goals and experimentation than higher-SES students (consistent with Author, 2013).
- Finally, considering within-SES results when correlations with ability existed, among lower-SES students, agreement with "doing science experiments is fun" was negatively related to ability. However, agreement that "solving brain teasers is fun" was unrelated. This framing by ability interaction was significant ($p = .02$), where *lower-ability, low-SES students tended to rate doing science experiments as more fun than solving brain teasers.*
- Although there were no differences relating to "thinking hard" while solving brain teasers, "thinking hard" was positively related to difficulty ratings for solving brain teasers among lower-SES students ($r = +.42$, $p = .08$) but not among higher-SES students ($r = +.06$, $p = .84$).

Predictions

SE vs. BT. Lower-SES—but not higher-SES—students rated solving brain teasers as both more difficult and less fun than doing science experiments. Although this may seem to suggest that lower-SES students would benefit more in the SE than BT framing, enjoyment of a lesson and learning outcomes are often inversely related (Clark, 1982). This may stem in part from students' preference for instruction that requires less effort to achieve success (Clark, 1982). Similarly, positive affective states are associated with poorer performance on "tasks that demand careful, effortful processing" (Friedman, Forster, & Denzel, 2007, p. 142) including logical reasoning tasks (Fiedler, 1988).

If students' expectations of task difficulty positively and enjoyment negatively influence effort exertion, lower-SES students may exert more effort in a BT than SE framing than higher-SES students. Thus, the benefit of the BT over SE framing would be greater for lower-SES than higher-SES students (P1). Thus, we also predicted that the BT framing would better promote CVS understanding than the SE framing among lower-SES students (P2). Because lower-ability lower-SES students considered doing experiments more "fun" than solving brain teasers, they may exert more effort in a BT than SE framing and thus learn more (P3). Predictions are summarized in Table 3. The only framing-related difference among higher-SES students was a stronger association of science goals with experimentation than solving brain teasers. This suggests the possibility that higher-SES students may benefit more in the SE than BT framing. Although not a formal prediction in the present study, if we do find this pattern, this is a plausible explanation.

Table 3
Summary of predictions

Framing	Mechanism	Expected outcome vs. Science Experiment framing
Brain Teaser	H1: Elicit effortful problem-solving	Better understanding of CVS • P1: BT vs. SE better among lower- than higher-SES students • P2: BT > SE among lower-SES students • P3: BT > SE lower-ability students (lower-SES)
Fair Comparison	H2: Elicit "fair comparison" schemas	Support full CVS development • P4: FC > SE among lower-knowledge (within each population) • P5: FC > SE among lower-ability students (within each population) • P6: FC > SE among lower-SES students;

SE vs. FC. We expected that instruction framed in terms of fair comparisons would support students' understanding of both the comparison *and* controlling aspects of CVS. Thus, we predicted that the Fair Comparison (FC) framing would benefit the lowest-knowledge students, or students with weak or no knowledge associated with experimental design (P4). Because understanding the rationale for controlling variables in the Science Experiment framing appears dependent on deductive reasoning skills, we expected the FC framing would better support lower-ability students' complete CVS understanding than the SE framing (P5). However, because higher-ability students are likely better able to follow the logical argument for controlling variables presented in the Science Experiment framing (and repeated throughout instruction), we did not expect them to necessarily perform better in the FC than SE framing. Finally, because lower-SES students likely have weaker knowledge of experimental design, including science goal orientations (Author, 2013), we expected them to benefit more from the FC than SE framing (P6).

Method (Study 1)

Participants

Because the effect of a particular framing may differ for students from different populations, abilities (e.g., comprehension, metacognitive, and reasoning skills), and initial knowledge levels, the study was conducted with two diverse student populations. One population (L5) included predominately low-SES fifth-grade students. The other (M67) included predominately middle-SES students in sixth and seventh grades. At the time of the study (Spring semester), all students from both populations were taking science.

Population-L5. Eighty fifth-grade participants from three urban schools serving primarily low-SES African-American students completed the pretest. Schools A and B were private Catholic schools, and School C was a charter school. Fourteen students were from School A, 23 from School B, and 43 from School C. Three students (3.8%)—one from School A and two from School C—scored above mastery level (67% correct) on the pretest and were excluded from analyses. Thus, a total of 77 L5 students were included in analyses of framing effects.

Population-M67. Fifty-six students (22 sixth-graders and 34 seventh-graders) from a suburban middle school that served primarily middle-SES Caucasian students (School D) completed the pretest. Seven of the 22 sixth-grade students (32%) and 10 of the 34 seventh-grade students (29%) demonstrated mastery on the pretest and were excluded from analyses of framing effects, leaving 39 M67 students (15 sixth-grade and 24 seventh-grade students).

M67 students were significantly more likely than L5 students to demonstrate mastery on the pretest, $\chi^2(2, N = 141) = 19.96$, $p < .001$, and M67 students scored significantly higher on the pretest than L5 students ($M = 3.38$, $SD = 3.06$; $M = 1.44$, $SD = 1.60$, respectively, $F(1, 139) = 24.19$, $p < .001$, $\eta_p^2 = .15$). However, there were no differences in total pretest score between

Schools A, B, and C for L5 students (p = .23), or between grades six and seven for M67 students (p = .58). Similarly, there were no differences in mastery rates within either population (i.e., across schools for L5 or across grades for M67). Thus, each population may be considered homogenous with respect to initial CVS knowledge.

Design

Stratified random assignment based on reading or ability level was used to assign students to condition within each classroom. Standardized (Pennsylvania System of School Assessment, or PSSA) reading levels were available only at School C. In Schools A, B, and D, science teachers' ratings of each student's "general ability" level on a 5-point Likert scale (1 = well below average; 2 = below average; 3 = average; 4 = above average; 5 = well above average) were used as the ability measure. Within each population, students rated above the mean teacher rating (or PSSA reading level) were considered higher-ability and those rated below the mean were considered lower-ability.

Instructional activities were the same across conditions; only the specific wordings used in instruction differed. For example, in the SE framing, instruction referenced "science experiments"; in the FC framing, the instruction referenced "fair comparisons"; in the BT framing, instruction referenced "brain teasers." In the latter conditions, there were no references to "science experiments" until mid-instruction (see Table 4).

Procedure

The study took place during students' regular 40-minute science class periods. In each participating class, each student worked individually on a computer in their school's computer laboratory for the entirety of the procedure. As summarized in Table 4, the entire study—including the story pre- and posttests—occurred across four class periods (one per day), for a total intervention time of about two hours. Instruction occurred in two class periods on consecutive days (approximately one hour of instructional time). As summarized in Table 4, all students first completed a computerized pretest that consisted of six story problems, each in one of three different contexts (e.g., selling drinks, as shown in Figure 1).

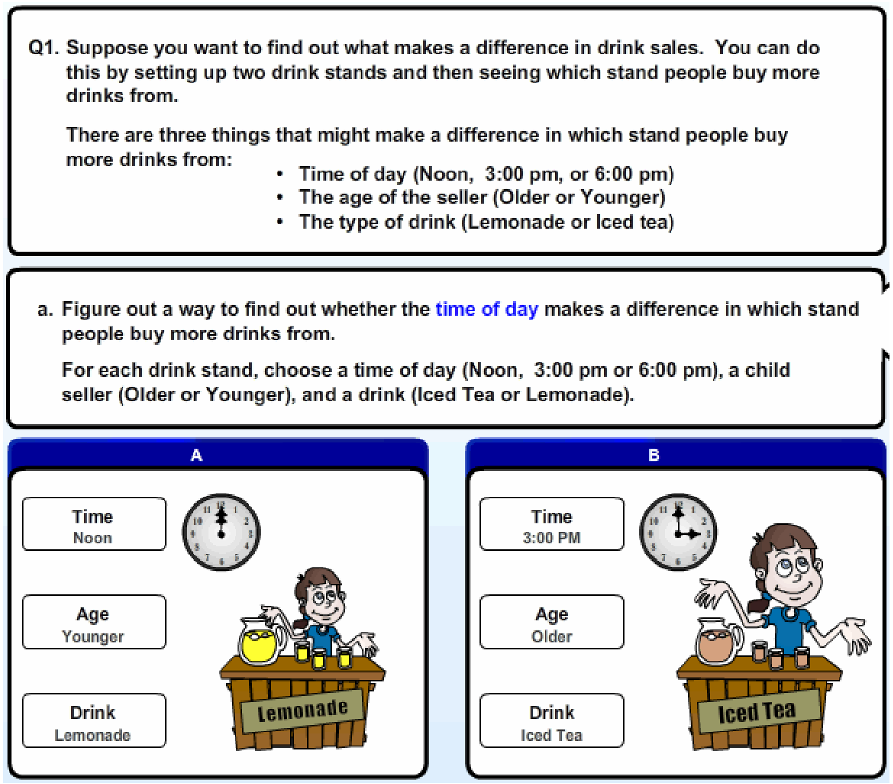

Figure 1. Example design item from story test.

In each problem context, students first designed an experiment for a given focal variable then evaluated another experiment for a different focal variable. For each design item, students were first asked to “figure out a way to find out whether the [focal variable] makes a difference in [outcome],” and explain why they set up [the scenario] the way they did. For each evaluate item, students were shown a set-up that a hypothetical student designed to “find out about [a particular variable]” and asked to evaluate it as a “good” or “bad” way to find out whether [the focal variable] makes a difference. After evaluating the set-up, students typed explanations for their responses. If they indicated it was a bad way to find out about the focal variable, they were prompted to change the set-up to make it a good way. The three set-ups students evaluated on the story pretest were (a) a maximally-contrastive experiment, in which all three variables were contrasted across the two conditions, (b) an experiment in which only one variable—but not the given focal variable—was contrasted, and (c) a non-contrastive experiment in which none of the three variables were contrasted (the two conditions were identical). The story pretest did not include terminology specific to any particular framing, such as “experiment,” “fair,” and “brain teaser.” Students could return to earlier problems (e.g., to check or revise previous responses).

Table 4

Study procedure

Day 1	**Story pretest**
Day 2	**Instruction:** Lesson introduction (framing-specific) Framing-specific example (maximally confounded) Introduction to ramps apparatus/variables Evaluate non-confounded design (framing-specific) *****Bridge to science framing***** Evaluate two confounded designs (science framing) **Ramps posttest**
Day 3	**Story posttest** (new domains)

The next day, students worked through computerized instruction on experimental design, delivered via a computer tutor, the “TED” tutor (e.g., Author, 2010), where they were introduced to the goal of the lesson in their respective framing condition (framing-specific wording of this and subsequent instruction is shown in on-line supplemental materials). During instruction, students did the following:

1. Viewed a confounded comparison (of two laundry detergents) and were told why the comparison was not good (again, in the language of their respective framing condition).

2. Were introduced to a virtual ramps apparatus—including its four variables (the type of ball used to roll down the ramp, the type of surface on the ramp, the slope of the ramp, and the starting position of the ball when it starts rolling).
3. Were shown a non-confounded comparison, and asked whether it was a "good" or "bad" experiment ("fair" or "unfair" comparison, or "good" or "bad" solution, depending on framing condition), then explained their evaluation response.
4. Received feedback on their multiple-choice responses, still in the framing-specific language of their condition.
5. Students in the BT and FC conditions (only) were given an explanation intended to bridge the initial framing to the subsequent science experiment framing language used for the remainder of instruction. This explanation included: "designing science experiments is just like designing these [fair comparisons/solving these brain teasers]."
6. Students in all three conditions then evaluated and received feedback on their responses to two confounded experiments within a Science Experiment framing.
7. Students completed a ramps posttest, in which they set up one experiment for each of the ramp's four variables then explained their design.

The next day, students completed the computerized story posttest, which included items isomorphic to those on the story pretest but in different contexts. As with the story pretest, this assessment did not contain any framing-specific terminology.

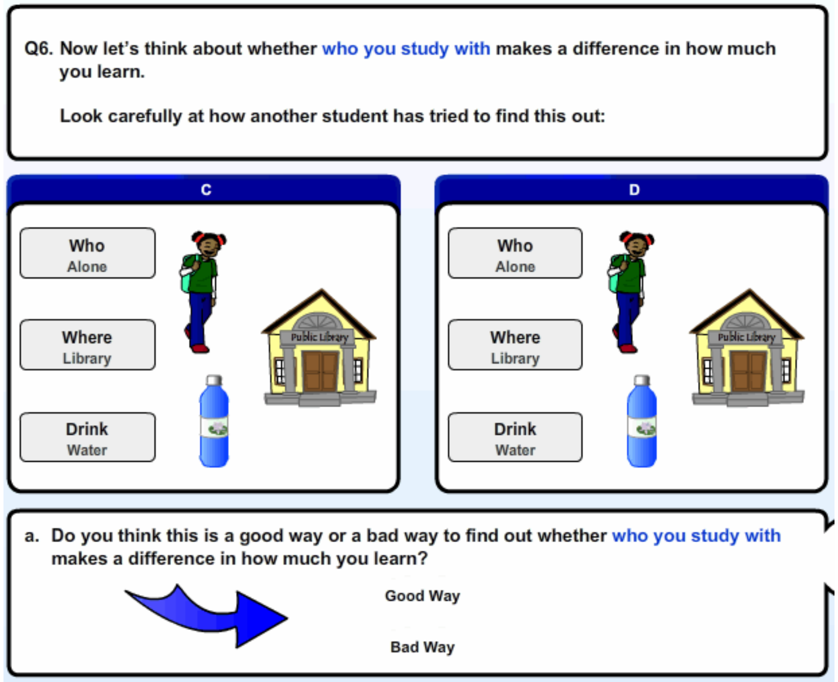

Figure 2. Sample evaluate item of story test.

Scoring

Ramps and story tests. On the ramps posttest, we gave one point for each unconfounded experiment students designed (out of 4 points). On the story tests, we gave one point for (a) each unconfounded experiment students designed (Figure 1 shows a confounded design) (out of 3 points), (b) each design they correctly evaluated as a "good way" or "bad way" (see Figure 2) (out of 3 points), and (c) each experiment they correctly modified (e.g., by contrasting studying alone with studying with a friend in Figure 2) (out of 3 points), for a maximum possible 9 points.

Initial knowledge measure. To investigate predicted initial knowledge by condition interactions, we coded students' responses to the final story pretest item (shown in Figure 2). Students who indicated that the set-up was not a good way to find out about the focal variable because (at least) the focal variable should be contrasted were coded as having "some" CVS knowledge (specifically, all but controlling variables). In contrast, responses indicating engineering goals in which "best" variable settings were used (e.g., indicating the comparison was a “good” way because “being by yourself is better”) or that their evaluation response was a guess were coded as “no knowledge” (of CVS, including its purpose). No students expressed a complete CVS understanding on the final story pretest question.

Consistent with prior findings (e.g., Author, 2013), among students who did not show mastery on the story pretest and were thus included in analyses, M67 students were significantly more likely to express “some” CVS knowledge on this final item than L5 students (55% vs. 31%, respectively), $\chi^2(1, N = 115) = 6.20$, $p = .01$. Thus, M67 students included in the following analyses were generally higher-knowledge (and more likely associate science goals with experimentation) than L5 students.

Analyses

Quantitative analyses. To assess our predictions (Table 5), we conducted both cross-population analyses (to assess predicted population by condition interactions) and within-population analyses (to assess predicted condition by ability/knowledge interactions). For each dependent variable, we first tested whether the effects of framing differed by student population (i.e., population by condition interactions) (model type (1) in Table 5); initial knowledge was included as a fixed factor.

Table 5

Summary of correspondences between analyses and predictions

Prediction	Analysis to test prediction For both Ramps and Story posttest
Cross-population:	**L5 vs. M67**:
BT > SE for lower-SES than higher-SES (P1)	(1) Population x condition interaction
Lower-SES population:	**L5:**
BT & FC > SE (P2; P6)	(2) L5: main effect condition
BT & FC > SE among lower- ability (P3; P5)	(3) L5: condition x ability interaction
FC > SE among lower- knowledge (P4)	& condition x knowledge interaction
Higher-SES population:	**M67**:
(no predictions)	(2) M67: main effect condition
FC > SE among lower-ability (P5)	(3) M67: condition x ability interaction
FC > SE among lower-knowledge (P4)	& condition x knowledge interaction

We also investigated *within-population* differences.[2] In all within-population analyses, school was included as a fixed factor in Population-L5 and grade was included as a fixed factor in Population-M67. (Unless otherwise noted, interactions with school and grade were not significant.) Within each population, we then conducted tests with only condition and school/grade included in analysis of variance (ANOVA) in order to test for main effects of condition (model type (2) in Table 5). To investigate possible interactions with ability and initial knowledge level (and possible higher-order interactions), we conducted within-population analyses with condition, school/grade, ability, initial knowledge level, and interaction terms included in ANCOVA (model type (3) in Table 5). Although we did not predict main effects of condition for M67 students or condition by initial knowledge interactions due to BT vs. SE framings, all main effects and interactions were investigated within each population to allow us to uncover any unexpected results. Table 5 summarizes the order in which the analyses are presented and their correspondence to predictions for the BT and FC framings.

Table 6
Wording of evaluation question for a "good" (unconfounded) experiment by condition

Condition		
Science Experiment	Fair Comparison	Brain Teaser
"Think about Amal's problem, and the *experiment* that he designed. Is this a *good* way to *design this experiment*?"	"Think about Amal's problem, and the *comparison* that he designed. Is this a *fair* way to *set up the comparison*?"	"Think about Amal's problem, and the *solution* that he designed. Is this a *good* way to *solve the problem*?"
"Why is it [not] a *good* way?"	"Why is it [not] a *fair* way?"	"Why is it [not] a *good* way?"

Qualitative analysis. To better understand how the different framings affected students' interpretations of instruction, we coded the responses students gave to the first question asked

[2] Because of the potential for differences in ability rankings across populations (and across grades/schools), we investigated main effects and/or interactions for each population separately.

during instruction, which referenced an unconfounded experiment with surface as the focal variable. Framing-specific versions of this question are shown in Table 6.

Student responses were categorized as follows:

- “Complete CVS:” explanations that the set-up was good because the focal variable was contrasted and the other variables were controlled (e.g., “[it’s good] because there is only one difference”).
- “Partial CVS:” explanations that the set-up was good because it followed at least one (but not all) of the CVS rules (e.g., “[it’s good] because it has two different ramps and it will show which one’s better”).
- “Contrastive:” explanations that the set-up was not good because some (non-focal) variables were not contrasted, but did not indicate a particular goal for contrasting (e.g., “[it’s not good] because they are at the same angle”).
- “Engineering: Best outcome:” explanations that the design was good because the chosen variable levels would produce good outcomes or the design was bad because the chosen variable levels would not produce good outcomes (e.g., “[it’s not good] because he used the wrong type of ramp so that will cause it not to be on an even surface and cause it to not work” or “[it’s not good] because it should not be in the middle it should be on the top so it can go faster”).
- “Same set-ups:” explanations indicating the design was bad because the focal variable (surface) differed across ramps or the design was good because the focal variable did not matter (e.g., “[it’s not good] because the surface of the ramp is different” or “[it’s not good] because it’s not exactly the same”).

Results

Ramps Posttest

Cross-population/Population-L5. There was no population by condition interaction for ramps posttest score, $F(2, 109) = 0.95$, $p = .39$, $\eta_p^2 = .02$. Within Population-L5, means scores (Table 7) were similar across conditions, $F(2, 73) = 0.38$, $p = .69$, $\eta_p^2 = .01$. There were no interactions between condition and either initial knowledge or ability. The effect of condition remained non-significant with initial knowledge and ability included as covariates ($p = .34$). Thus, among L5 students, none of the predicted effects were found for immediate learning as shown on the ramps posttest.

Table 7
Posttest means (standard deviation) by population and framing condition

	L5			M67		
Posttest	SE	BT	FC	SE	BT	FC
Ramps[a]	1.88 (1.78)	2.28 (1.72)	2.00 (1.54)	2.46 (1.66)	3.08 (1.31)	3.46 (0.88)
Story[b]	2.80 (2.77)	3.56 (2.90)	2.96 (2.93)	6.00 (3.46)	4.38 (3.53)	5.77 (3.47)

[a] Out of 4 points. [b] Out of 9 points.

Population-M67. Within Population-M67, there was a marginally significant effect of condition, $F(2, 34) = 3.03$, $p = .06$, $\eta_p^2 = .15$ (see Table 7). Pair-wise comparisons revealed that students in the FC condition scored higher than students in the SE condition ($p = .02$) but not significantly higher than students in the BT condition ($p = .39$). Thus, for M67 students, there appeared to be an immediate benefit of the FC condition for learning. As with L5, there were no interactions between condition and initial knowledge or condition and ability. Thus, among M67 students, none of the *predicted* effects were found for immediate learning. However, there was an unexpected advantage of the FC over the SE condition among all M67 students (not just the lower-knowledge or lower-ability students, as predicted, P4 and P5).

Story Posttest

Cross-population. There was a marginally significant population by condition interaction for total story posttest score, $F(2, 103) = 2.64$, $p = .08$, $\eta_p^2 = .05$ (see Figure 3 and Table 7). This interaction was due to a cross-population interaction with the BT and SE conditions, $F(2, 69) = 4.13$, $p = .046$, $\eta_p^2 = .06$. As predicted (P1), the advantage of the BT over the SE framing was greater among L5 than M67 students. In the SE framing, M67 students out-performed L5 students, $F(1, 34) = 8.32$, $p = .007$, $\eta_p^2 = .20$. However, M67 and L5 students performed similarly in the BT condition, $F(1, 35) = 0.50$, $p = .49$, $\eta_p^2 = .01$.

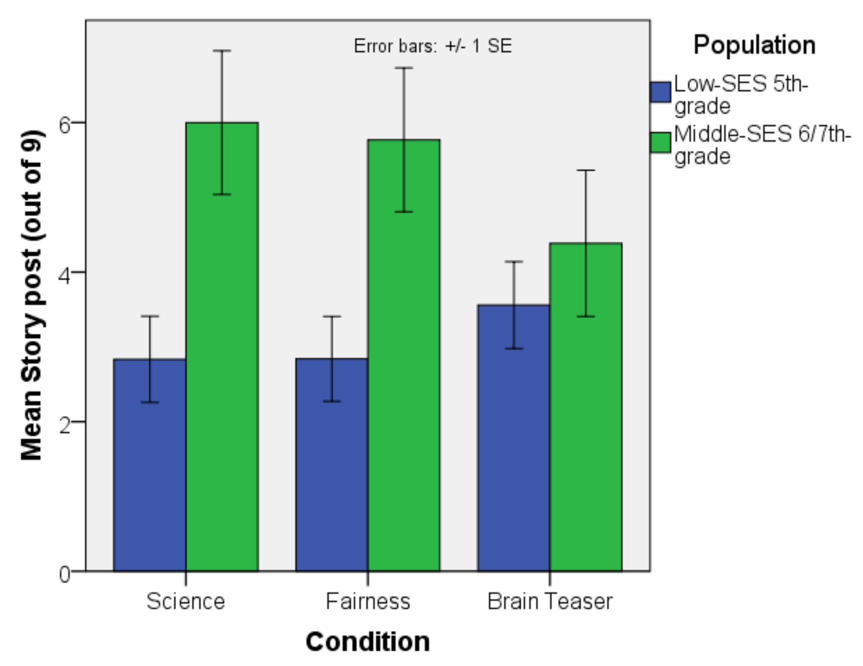

Figure 3. Condition by population interaction for transfer performance.

Population-L5. There was no main effect of framing condition for total posttest score (out of 9), $F(2, 71) = 0.41$, $p = .67$, $\eta_p^2 = .01$, counter to P2 and P6. With total posttest score as the dependent variable and initial knowledge level (“no” vs. “some” knowledge), ability, condition, and school included as independent variables, there was a significant condition by ability by knowledge interaction, $F(2, 52) = 8.36$, $p = .001$, $\eta_p^2 = .24$. There was no effect of condition among L5 students who expressed “no” knowledge on the final pretest item ($p = .54$), counter to P4. Among students who expressed “some” knowledge, the effect of condition depended on ability, $F(2, 14) = 4.94$, $p = .02$, $\eta_p^2 = .41$. There was no effect of condition for higher-ability students, ($p = .22$), but there was a significant effect for lower-ability students, $F(2, 10) = 4.02$, $p = .05$, $\eta_p^2 = .45$. Lower-ability L5 students who expressed at least a science goal scored higher in the BT than in the SE condition ($p = .03$) and FC ($p = .03$) condition. The advantage of the BT over SE condition among the lower-ability L5 students is partially consistent with P3, but this advantage was *only* among students who showed “some” CVS knowledge on the pretest. Counter to P5, there was no difference between the FC and SE conditions among these lower-ability students ($p = .92$). The BT framing brought (higher-knowledge) lower-ability students up to the level of the higher-ability students. Because the advantage of the BT framing was among higher-knowledge students (who expressed an understanding of contrasting the focal variable on the final pretest item), these results suggest that lower-ability L5 students primarily learned to *control variables* from instruction in the BT framing. This corresponded with a trend that students in the BT framing spent more time answering questions for the first instructional experiment than students in the SE framing, $t(28.29) = 1.53$, $p = .07$ (one-tailed).

Table 8

Summary of results per prediction

Prediction	Study 1		Study 2
	DV: Ramps posttest	DV: Story posttest	DV: Story posttest[a]
Cross-population:	**L5 vs. M67:**	**L5 vs. M67:**	
BT > SE: lower-SES (P1)	(n.s.)	BT vs. SE: Population x condition (P1)	(n/a)
Lower-SES population:	**L5:**	**L5:**	**L5:**
BT > SE (P2)	(n.s.)	(n.s.)	(n.s.)
FC > SE (P6)	(n.s.)	(n.s.)	(n.s.)
BT > SE: lower-ability (P3)	(n.s.)	BT > SE: Low-ability “some” knowledge	(n.s.)
FC > SE: lower-ability (P5)	(n.s.)	(n.s.)	(n.s.)
FC > SE: lower-knowledge (P4)	(n.s.)	(n.s.)	FC > SE: Lower-knowledge (P4) BT > SE: Higher-knowledge BT & FC > SE: Higher-ability
Higher-SES population:	**M67:**	**M67:**	
(no predictions)	FC > SE	(n.s)	(n/a)
FC > SE: lower-ability (P5')	(n.s.)	(n.s.)	
FC > SE: lower-knowledge (P4')	(n.s.)	(n.s.)	
		SE; FC > BT: High-ability “no” knowledge	

[a] There were no significant effects for ramps posttest.

Population-M67. Although FC students showed better performance on the ramps posttest, on the *story* posttest, there was no main effect of condition for Population-M67, $F(2, 35) = 0.55$, $p = .58$, $\eta_p^2 = .03$. This was due to a significant performance drop between the ramps and story posttest in the FC condition ($p = .01$) but no drop among SE students ($p = 1$). Including covariates in ANCOVA, as with L5, there was a significant condition by ability by knowledge interaction, $F(2, 20) = 3.74$, $p = .04$, $\eta_p^2 = .27$. Among students who expressed “some” initial knowledge, there was no significant effect of condition, $F(2, 16) = 0.57$, $p = .58$, $\eta_p^2 = .07$. However, among students who did *not* express even a science goal, there was a significant condition by ability interaction, $F(2, 11) = 5.39$, $p = .02$, $\eta_p^2 = .50$.[3] The effect of framing condition was not significant for lower-ability students ($p = .71$), counter to P5 (and P4). However, among higher-ability M67 students, there was a significant effect of condition, $F(2, 10) = 8.23$, $p = .008$, $\eta_p^2 = .62$, where students in the BT framing scored significantly *lower* than those in either the SE ($p = .004$) or FC framing ($p = .007$). Higher-ability (low-knowledge) students scored similarly in the SE and FC conditions ($p = .89$) (counter to P4). Results are summarized in Table 8.

Initial Instructional Responses

We now discuss the responses students gave to the first question asked during instruction, which referenced an unconfounded experiment with surface as the focal variable. We were particularly interested in how the younger/lower-SES students in the FC condition—who did not

[3] Higher-ability no-knowledge M67 students performed worse in the BT framing on both story posttest design items, $F(2, 10) = 7.02$, $p = .01$, $\eta_p^2 = .58$ (BT < SE, $p = .008$), and evaluate items, $F(2, 10) = 4.19$, $p = .048$, $\eta_p^2 = .46$ (BT < SE, $p = .02$).

show *any* predicted benefits—responded to this question. Framing-specific versions of this question are shown in Table 6.

Cross-population. The goal response pattern for each population is shown in Table 9. M67 students tended to express a complete or partial CVS understanding more often in the FC than SE framing, but L5 students tended to express a complete or partial CVS understanding more often in the SE than FC condition. This 3-way interaction was significant, $\chi^2(1, N = 72) = 6.44$, $p = .01$.

Population-L5. For all L5 students, the most notable difference was that students in the FC condition were significantly more likely to respond that the experiment was *not* a fair way because the surfaces were different, $\chi^2(2, N = 78) = 16.05$, $p < .001$. This implies that a non-contrastive experiment *would* be a fair/good way. The majority of L5 students in the FC condition (59%) gave this type of response. Thus, the majority of students in this condition seemed to have developed an incorrect understanding of experimentation from the FC framing condition. The only other response type that differed by condition was best outcome engineering, which students in the BT condition were more likely to give, $\chi^2(2, N = 78) = 8.13$, $p = .02$.

Table 9
Proportion of responses to first instructional question by population and condition

Response type	L5			M67		
	SE	FC	BT	SE	FC	BT
Complete CVS	28%	15%	16%	54%	77%	38%
Partial CVS	20%	11%	24%	0%	8%	31%
Complete/partial	48%	26%	40%	54%*	85%*	69%
Contrastive	12%	4%	20%	0%	0%	8%
Best outcome	4%**	0%**	**20%****	0%	0%	16%
Same set-ups	20%***	**59%*****	12%***	46%*	15%*	8%*
(Unclear)	16%	7%	8%	0%	0%	0%
Don't know	0%	4%	0%	0%	0%	0%
Total responses:	25	27	25	13	13	13

$^{*}p < .10$ $^{**}p < .05$ $^{***}p < .01$ $^{****}p = .001$.

Population-M67. For M67 students, there was again a significant relationship between condition and response type, (p = .02). However, as shown in Table 9, in contrast to the younger and lower-SES population, students in the FC—rather than SE—condition tended to give more CVS responses, though this difference was not significant, $\chi^2(2, N = 39) = 3.96$, p = .14. However, M67 students in the FC condition were marginally more likely to give complete or partially-correct responses (p = .08) than in the SE framing (85% vs. 54%). This trend is consistent with the advantage of the FC condition for the older and higher-SES student population shown on the ramps posttest. Also in contrast to L5, M67 students in the SE—rather than FC—condition were more likely to express a "Same" view that either (a) the experiment was (not) good because it would (not) produce the same outcomes, or (b) the experiment was not good because the surfaces differed (46% vs. 15%, respectively, p = .05).

Discussion (Study 1)

We investigated whether the contextualization—or framing—of instruction on experimental design affects learning and transfer outcomes, and whether certain ways of framing instruction are better suited to students from different populations, initial knowledge, and ability levels. We hypothesized that a "brain teaser" framing would elicit expectations of task difficulty and thus prompt more effortful problem-solving—particularly among students from lower-SES populations (P1; P2), and lower-ability students within lower-SES populations (P3). We also hypothesized that a "fair comparison" framing would activate knowledge schemas that would support initial understanding and transfer—particularly among students from lower-SES populations (P6), lower-knowledge students (P4; P4'), and lower-ability students (P5; P5').

Cross-Population Interactions

M67 students showed better initial learning outcomes on the ramps posttest in the FC than SE framing. However, there were no other differences in the effect of framing on initial learning outcomes (on the ramps posttest) across student populations, ability levels, or initial knowledge levels. Consistent with prior findings that instructional framing interacts with learner characteristics such as knowledge level (McElhaney & Linn, 2010), motivational factors (e.g., Tang & Baumeister, 1984), and even mood (e.g., Friedman, Forster, & Denzler, 2007), there were interactions between framing and student factors for *transfer* performance. The BT framing differentially affected transfer performances across populations. Consistent with P1, L5 students' transfer performance was better in the Brain Teaser than Science Experiment condition, but the trend was reversed among older/higher-SES students.

Within-Population Interactions: Brain Teaser vs. Science Experiment Framing

Within each population, the effect of framing on transfer performance depended on at least some student-specific factors. In the L5 population, the effect of the BT framing depended only on student ability; lower-ability students who expressed initial knowledge of contrasting the focal variable on the story pretest showed better CVS transfer in the BT than other framing conditions. Because these students showed understanding of contrasting the variable (or variables) they were testing on the final pretest item, the BT framing best supported students' understanding of controlling variables during instruction. As discussed previously, understanding the need to control non-focal variables likely requires analytical modes of thinking (or reasoning). Thus, as hypothesized, framing instruction in terms of solving brain teasers appeared to elicit reasoning among the (higher-knowledge) lower-ability L5 students.

However, this outcome is only partially consistent with our prediction (P3) that lower-ability students within the L5 population would benefit more from the BT than SE framing because only higher-knowledge (rather than all) lower-ability students showed gains. The BT framing did not appear to elicit science goals among lower-knowledge students, who did not show any gains on the transfer posttest. In fact, students in the BT framing were more likely to express engineering goals on the initial instructional question than students in the other framing conditions. Similarly, lower-SES students tended to less strongly agree that the purpose of solving brain teasers was to find out about something (i.e., a science goal purpose) than doing science experiments.

In contrast to lower-ability L5 students with some initial CVS knowledge, among higher-ability M67 students who did *not* express any CVS understanding on the final pretest question, those given the BT framing performed *worse* than students in the other conditions. That is, among higher-ability M67 students who did not show an understanding of the need to contrast the focal variable on the final pretest question, there was actually a detrimental effect of the BT framing on transfer performance. This result is consistent with the survey result that a “science goal” of finding out about something was more strongly associated with doing science experiments than solving brain teasers among higher-SES students (see Table 2). Together, these results suggest that the BT framing did not support an understanding—or realization—of the science goal underlying the purpose of the lesson in *either* student population.

Fair Comparison Schema

We hypothesized that the fair comparison framing would elicit schematic knowledge that would support initial CVS understanding among the lower-SES population (P6), lower-knowledge students within either population (P4; P4'), and lower-ability students within either population (P5; P5'). However, neither population showed the predicted benefits of the FC framing on transfer performance.

However, there were indications that the FC framing elicited schematic knowledge of fair comparisons among the higher-SES students. M67 students actually showed better initial understanding of CVS (on the initial instructional question) in the FC than SE condition. Evidence of a quick uptake of a CVS understanding is consistent with a holistic adoption of a pre-existing "fair comparison" knowledge schema in the FC framing, rather than an incremental process of conceptual change (cf. Vosniadou, Vamvakoussi, & Skopeliti, 2008). This is also consistent with the framing survey results, where students expressed CVS—that is, that only one (focal) variable should differ between conditions—only in their explanations of "fair comparisons." Additionally, ideas about controlling variables were expressed only for fair comparisons; students did not mention controlling variables in their explanations of the purpose of doing experiments or solving brain teasers. M67 students in the FC framing continued to out-perform those in the SE framing condition on the ramps posttest.

In contrast, in their initial instructional responses, the majority of L5 students given the FC framing expressed a belief that the experimental conditions should be set up identically. Consequently, L5 students in the FC condition gave on-target responses at half the rate of L5 students in the SE and BT conditions. Thus, rather than supporting students' understanding, the FC framing induced quickly-adopted, schema-like, misconceptions that the set-ups should be identical.

The poorer initial performance of L5 students in the FC framing compared to M67 students may be due to cross-population differences in prior knowledge or schemas associated with the fairness terminology. The survey results suggest SES as one factor that influenced students' interpretations of the FC framing: only lower-SES (who were also African-American) students interpreted "fair comparisons" within a social framework, as in being treated fairly (e.g., as “When you can be treated fairly” and “When you have equal rights”). Lower-SES students were more likely to express a social response than higher-SES students (p = .06). These expressions appear related to notions of “fair” as “equally-divided resources” found among young children (Olson & Spelke, 2008; Sloane, Baillargeon, & Premack, 2012). Such early conceptions of “fairness” may be reinforced more in students from lower-SES backgrounds, based on more frequent experiences with discrimination and distributing limited resources. In addition, the younger (L5) students simply have had less time to develop other—more complex—conceptions related to “fairness.”

Thus, consistent with our hypothesis, the FC framing appeared to support understanding via eliciting knowledge schemas. However, a “fair comparison” schema was only elicited in the older and higher-SES student population. Among younger, lower-SES students, a “fair as equal” schema appeared to be elicited instead. Given these findings, we examined whether presenting the fair comparison framing in a way that placed more emphasis on the contrasting aspect could support students’ learning above the “traditional” science experiment framing.

Method (Study 2)

Procedure

We conducted a follow-up study using the same materials and procedure used in Study 1 (see Table 4). However, we included a framing-specific “ramps pretest” after the framing-specific example and ramps introduction. The ramps pretest was identical to the ramps posttest described in Study 1, except that students provided explanations of their designs by selecting responses from drop-down menus. No feedback was given during the ramps pretest. However, immediately after completing the ramps pretest, students who did not contrast the focal variables in their designs were given remedial instruction that included practice contrasting focal variables and conceptual instruction on why contrasting the focal variable is necessary. The only other procedural difference from Study 1 was that students completed the instruction during their 40-minute science class on two consecutive days rather than one; instruction on the second day began with the bridge to the science framing.

Participants

Twenty-eight fifth-graders from an urban private Catholic school (School B in Study 1) serving primarily low-SES African-American students completed all phases of the study. No students demonstrated an initial understanding of CVS on the story pretest. Stratified random assignment to condition based on the teacher's ability ratings resulted in similar ability (p = .97) and story pretest performance (p = .87) across conditions.

Results and Discussion (Study 2)

Because only two students (one in the FC and one in the SE condition) expressed "Some" knowledge of CVS on the final question of the story pretest, we used story pretest score (out of nine) as the measure of initial knowledge. As in Study 1, there was no main effect of condition or interactions of initial knowledge or ability with condition on the ramps posttest (again, failing to support P2 and P6). However, for total story posttest score, there was an interaction between condition and student ability, $F(1, 19) = 3.69$, $p = .04$, $\eta_p^2 = .28$ (Figure 4a). This interaction was significant for FC vs. SE (p = .01) and marginally significant for BT vs. SE (p = .07). Counter to P3 and P5, higher-ability—rather than lower-ability—students in the alternative framing conditions were better able to transfer CVS to novel problems than higher-ability students in the SE framing.

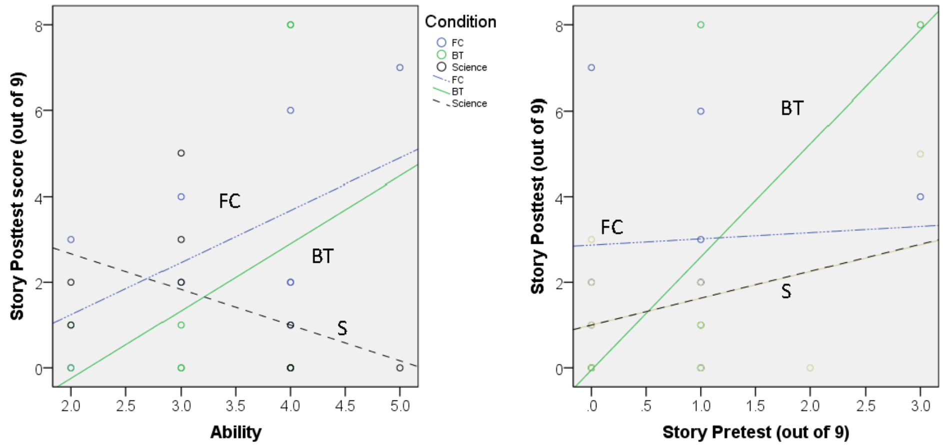

Figure 4. Relationship between transfer performance and condition by (a) student ability, and (b) story pretest.

There was also a significant condition by pretest interaction, $F(1, 19) = 4.54$, $p = .02$, $\eta_p^2 = .32$ (Figure 4b). This interaction was significant for BT vs. SE ($p = .03$). Like the L5 students with "Some" initial CVS knowledge in Study 1, higher-knowledge students in the BT condition tended to perform better than higher-knowledge students in the SE condition. Consistent with P4, among lower-knowledge students (who scored less than two on the story pretest), those in the FC condition scored marginally higher than those in the SE framing, $F(1, 15) = 3.97$, $p = .065$, $\eta_p^2 = .21$. Results are summarized in Table 7.

Thus, emphasizing the contrasting aspect of experimental design by including a ramps pretest and remedial instruction on contrasting focal variables improved students' transfer of CVS to the story posttest given the day after instruction. Unlike in Study 1, where the majority of L5 students gave "Same set-up" responses to the initial instructional question, none of the FC students gave this type of response. There was no difference between conditions in response type in this study ($p = .16$). Finally, M67 students in Study 1 showed initial benefits of the FC framing on the ramps posttest but not on the transfer posttest due to a significant drop in the percent of unconfounded experiments students set up across posttests in the FC but not SE condition. This suggests M67 students' learning in the FC framing was not as robust as in the SE framing. However, we found the reverse trend in this study, where the decrease in performance from the ramps to story posttest was marginally *less* among students in the FC than SE condition, $F(1, 16) = 4.0$, $p = .06$, $\eta_p^2 = .21$.

General Discussion

Although designing controlled experiments is a simple skill that involves only a few procedural rules (and an understanding of why controlling variables is necessary), younger students and students from lower-SES populations consistently show poorer learning and transfer outcomes following instruction on experimental design within a "science experiment" framing (Author, 2005; Dean & Kuhn, 2006; Lorch et al., 2010). Based on research showing that the contextualization—or framing—of instruction may affect task performance (e.g., Adam & Galinsky, 2012; Anderson & Pearson, 1984; Bransford & Johnson, 1972; Cheng & Holyoak, 1985; Cheng & Holyoak, 1989; Dijksterhuis & Van Knippenberg, 1998; Marshall, 1995; Tse et al., 2007), we investigated whether alternative instructional framings on experimental design—targeting different learning mechanisms—could better support student learning and transfer. The theoretical question underlying the framings chosen addresses the efficacy of supporting learning and transfer by attempting to compensate for knowledge- and reasoning-related weaknesses by activating analogous *knowledge schemas* or attempting to compensate for metacognitive weaknesses by *prompting effortful problem-solving* compared to baseline instruction. Specifically, we attempted to activate students' intuitive knowledge schemas related to "fair comparisons" in the "Fair Comparison" framing and prompt effortful problem-solving via a "Brain Teaser" framing.

We selected the Brain Teaser framing based on our prior findings that students who failed to learn CVS from instruction often did not realize their initial knowledge deficits and thus may have failed to exert the effort necessary for learning. Our survey results suggested that lower-SES—but not higher-SES—students may view solving brain teasers as more difficult than doing science experiments; thus, we predicted that they would exert more effort in a Brain Teaser than Science Experiment framing and consequently learn more. Further, because our survey results indicated that the lower-ability of the lower-SES students may view solving brain teasers as less “fun” than doing science experiments, the benefits of a BT framing via effort inducement may be even greater among lower-ability students. We predicted that younger/low-SES students would benefit more than a higher-SES population from the Brain Teaser framing (P1). Additionally, we predicted that the (younger and) lower-SES students in general, and lower-ability low-SES students in particular would show better learning and transfer from Brain Teaser framing than a Science Experiment framing (P2; P3).

We selected a Fair Comparison framing in part based on findings that tie knowledge (specifically, science goal orientations) and reasoning to CVS learning. Our survey results suggested that students are more likely to associate controlling variables with fair comparisons than science experiments, per se. We predicted that younger/low-SES students in general (P6), lower-knowledge students (P4; P4'), and lower-ability students (P5; P5') would learn CVS better within a Fair Comparison framing because this framing would prime pre-existing knowledge schemas related to fair comparisons that include science-goal orientations and can support the most difficult aspect of CVS—controlling variables.

Brain Teaser Framing: Eliciting Effortful Problem-Solving (H1)

In both Study 1 and Study 2, students from the younger/lower-SES population who were higher-knowledge—or who demonstrated a science goal understanding of the task on the story pretest—showed better transfer performance from the BT than SE framing. Such interactions between instructional framing and learners' initial knowledge levels have been found previously (e.g., McElhaney & Linn, 2010). Lower-SES (sixth-grade) students rated the relative difficulty of solving brain teasers compared to doing science experiments as greater than higher-SES students on the framing survey. There was also a positive relationship between difficulty ratings of solving brain teasers and "thinking hard" while solving them among lower-SES students only. That students may associate task difficulty or "thinking hard" with engaging in knowledge integration or reasoning is suggested by research finding that even 6-7 year-old children associate reasoning with effortful thinking (Amsterlaw, 2006). Consistent with this explanation, there was a trend for time students spent answering initial instructional questions to be greater in the BT than SE framing. Thus, it is plausible that the BT framing primed the (hypothetical-deductive) reasoning that allows one to understand the rationale for controlling variables. These results are consistent with findings that eliciting conceptions can promote associated behaviors (e.g., Dijksterhuis & Van Knippenberg, 1998; Adam & Galinsky, 2012). The framing survey results also suggest the possibility that these effects may be related to lower-ability low-SES students' perceptions of doing experiments as more fun than solving brain teasers compared to higher-ability lower-SES students. Such associations may prime positive affective states in a Science Experiment framing, which may in turn impair critical thinking (cf. Friedman, Forster, & Denzel, 2007) and thus, developing an understanding of controlling variables. This may account for why lower-SES students who demonstrated science goal understanding on the final pretest item in Study 1 or higher-pretest students in Study 2 benefited from the BT framing.

However, these results also suggest that the “comparing and contrasting” aspect of CVS needs to be made even more salient than in Study 2 either prior to or during instruction for this framing to also benefit lower-knowledge students.

Although lower-knowledge students did not benefit from the BT framing, in Study 1, the lower-ability students did benefit from the BT framing. However, in Study 2, higher-ability students benefited. The reason for this is unclear. One possibility is that the majority of lower-SES students in Study 1, who were from an above-average performing charter school—where students were 25% more likely to score at proficient or advanced levels on the Pennsylvania System of School Assessment (PSSA) reading assessment than students in the public school district—were generally higher-ability than students at School B (an average-performing school). Thus, the “lower-ability” Study 1 students may have been more similar to the “higher-ability” Study 2 students. It may be that students must possess a particular level of reasoning skills to benefit from any effort induced by a BT or similar game-like framing.

Another open question is—if the BT framing did induce students to think more logically than the SE framing—then why did the sixth-graders who completed the framing survey not more strongly agree that solving brain teasers requires reasoning than does designing science experiments? One possibility is that—because the students who completed the survey had been in the process of conducting independent science experiments (which they later presented at an in-school science fair)—they had come to realize the amount of thought and reasoning that goes into designing experiments. In contrast, the L5 study participants—who were not required to participate in school science fairs—tended to under-estimate the reasoning involved in designing science experiments. Another possibility is that students around this age (11 years old) may not hold accurate metacognitive beliefs about their mental processes involved in certain activities (such as their levels of concentration or reasoning processes), though they may have more accurate beliefs regarding affective states (e.g., their enjoyment of activities). It is important to note that, although the results of students in the BT framing are consistent with survey results regarding perceived difficulty and task enjoyment, these are correlational findings. Thus, future studies that somehow pinpoint and manipulate only effort or task enjoyment (or other plausible factors) are necessary to allow one to claim these as causal factors.

Fair Comparison Framing: Eliciting Relevant Knowledge Schemas (H2)

In Study 1, younger and lower-SES students tended to initially misinterpret instruction in the Fair Comparison framing as setting up identical conditions, which likely resulted in poorer CVS learning and transfer performance. However, consistent with other work showing how eliciting knowledge schemas can support reasoning (e.g., Cheng & Holyoak, 1985; Cheng & Holyoak, 1989), when the salience of contrasting the focal variable was increased by engaging lower-SES fifth-grade students in practice setting up experiments and providing explicit remedial instruction on contrasting the focal variable (Study 2), lower-knowledge students in the FC condition demonstrated better transfer performance than lower-knowledge students in the SE condition. That this benefit was among lower-knowledge students (who showed no understanding of CVS on the pretest) suggests that the FC framing supported a complete CVS understanding—or an understanding of both the contrasting and controlling aspects of CVS. Although this may suggest holistic adoption of a fair comparison schema, there was no evidence that students in the FC framing adopted CVS *earlier* in instruction than students in the SE (or BT) framing, as would be expected.

However, unlike the older and higher-SES students, whose initial gains in the FC framing, shown in their initial instructional responses and on the ramps posttest, were not present on the transfer posttest, younger and lower-SES students in the FC framing showed more stable performance between the ramps and transfer posttest than their counter-parts in the SE framing. This contrast suggests that an FC framing may cause more robust learning among younger and lower-SES than older and higher-SES students. It is possible that the ease of initial learning of CVS—as evidenced by its early adoption among M67 students—resulted in weaker CVS schemas that were less likely to transfer to the novel items of the story posttest; this is consistent with the idea of creating desirable difficulties that slow down learning but increase conceptual robustness (cf. Bjork & Linn, 2006). Consistent with this, the framing survey responses indicated a trend for lower-SES students to consider the relative difficulty of "making fair comparisons" verses "doing science experiments" as greater than higher-SES students (though this interaction was not significant). This could indicate that lower-SES students showed more robust CVS knowledge in the FC than SE framing in part because their learning was more effortful. Another possibility is that a "fair comparison" framing (when combined with an emphasis on contrasting) elicits stronger beliefs—and perhaps stronger affect—regarding "fairness" among lower-SES students, which translates into more robust CVS schemas. Further studies are necessary to better understand the source of the differential effect of the Fair Comparison framing across populations.

Implications

Returning to the over-arching theoretical question of whether students' learning and transfer of CVS can be better improved over a baseline "Science Experiment" framing by prompting more *effortful problem-solving* (H1) or activating *knowledge schemas* (H2), as for most questions of this sort, the results of this study suggest that it depends. That is, the better general instructional method depended on student population—which may be related to differences in knowledge and beliefs regarding the instructional terminology (and these beliefs may be influenced by students' personal experiences)—general ability levels, as well as the specific instructional procedures. When instruction was modified (Study 2), among lower-SES fifth graders, the lower-knowledge students tended to benefit more from the Fair Comparison framing (as predicted), and the higher-knowledge students benefited more from the Brain Teaser framing.

In general (see Table 7 for a summary), the alternative initial framings resulted in improved transfer performance among the younger/lower-SES populations but the Science Experiment framing was equal to or better than the alternative framings among the older and higher-SES population. Because Science terminology may be most commonly used by classroom teachers, more advantaged students may typically benefit the most and less advantaged students may generally benefit least from typical instruction. Teachers of younger and/or lower-SES students in particular may want to consider—at least initially—alternative terminologies that encourage effortful and/or problem-solving approaches during learning activities, or elicit students' knowledge schemas that are analogous to the instructional concepts to support students' understanding of the rationale for controlling variables.

However, it is vital that teachers assess their *particular* students' knowledge and beliefs—both cognitive and affective—regarding such terminology prior to instruction. One lesson we learned from this study was the danger of generalizing students' beliefs across ages and schools. This led to our underestimating the potential for misinterpreting the instructional terminology among younger students (who also had less science inquiry experience). Failure to take such knowledge into consideration when designing instruction may lead to confusion and poorer learning outcomes. In the Next Generation Science Standards, controlled experiments are referred to as “fair tests.” As we found in Study 1, using such terminology in instruction may be confusing to particular students—especially younger students and students from lower-SES populations.

In conclusion, framing instruction on CVS in non-traditional ways improved primarily transfer performance, particularly among the lower-SES populations. The finding that a Brain Teaser framing supported learning of the most difficult aspect of CVS, controlling variables, suggests the possibility that using game-like framings may improve learning outcomes, provided that they meet the needs of the particular students *and* task. Furthermore, the framing may need to highlight and infuse challenge into aspects of the task that induce learning, rather than peripheral aspects of the task (e.g., Habgood & Ainsworth, 2011). Using a brain teaser framing may have done just this, focusing students’ attention on the design of the experiment as it relates to the rationale for controlling variables. However, further research is necessary to determine the robustness of outcomes of the present study and more fully understand how task framing interacts with student-specific characteristics to maximize learning for all students.

References

(2000). FOSS full option science system. 2000 ed. Nashua, NH: Delta Education Inc.

Adam, H. & Galinsky, A. (2012). Enclothed cognition. Journal of Experimental Social Psychology, 48(4), 918-925.

Amsterlaw, J. (2006). Children’s beliefs about everyday reasoning. Child Development, 77, 443-464.

Anderson, C.W., & Roth, K.J. (1989). Teaching for meaningful and self-regulated learning of science. In Brophy, J., (Ed.), Advances in Research on Teaching: Vol. 1 (pp. 265–310). Greenwich, CT: JAI Press.

Anderson, R. C., & Pearson, P. D. (1984). A schema-theoretic view of basic processes in reading comprehension. Handbook of reading research, 1, 255-291.

Anderson, R. C., Spiro, R. J., & Anderson, M. C. (1978). Schemata as scaffolding for the representation of information in connected discourse. American Educational Research Journal, 15(3), 433-440.

Author (2013, 2012, 2011, 2010a, 2010b, 2006, 2005, 2004, 2000, 1999).

Bianco, A. T., Higgins, E. T., & Klem, A. (2003). How “fun/importance” fit affects performance: Relating implicit theories to instructions. Personality and Social Psychology Bulletin, 29(9), 1091-1103.

Bjork, R. A., & Linn, M. C. (2006). The science of learning and the learning of science. *APS Observer*, 19(3).

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. Journal of Verbal Learning and Verbal Behavior, 11(6), 717-726.

Bullock, M. & Ziegler, A. (1996). Thinking scientifically: conceptual or procedural problem? Second International Baltic Psychology Conference, Tallinn, Estonia.

Case, R. (1974). Structures and strictures: Some functional limitations on the course of cognitive growth. Cognitive Psychology, 6, 544–574.

Castelli, I., Massaro, D., Sanfey, A. G., & Marchetti, A. (2013). “What is fair for you?” Judgments and decisions about fairness and Theory of Mind. European Journal of Developmental Psychology, (ahead-of-print), 1-14.

Cheng, P. W., & Holyoak, K. J. (1985). Pragmatic reasoning schemas. Cognitive Psychology, 17, 391-416.

Cheng, P. W., & Holyoak, K. J. (1989). On the natural selection of reasoning theories. Cognition, 33, 285-313.

Cheng P. W., Holyoak, K. J., Nisbett, R. E., & Oliver, L. (1986). Pragmatic versus syntactic approaches to training deductive reasoning. Cognitive Psychology, 18, 293-328.

Clark, R. E. (1982). Antagonism between achievement and enjoyment in ATI studies. Educational Psychologist, 17(2), 92-101.

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

Dijksterhuis, A. & Van Knippenber, A. (1998). The relation between perception and behavior, or how to win a game of trivial pursuit. Journal of Personality and Social Psychology, 74(4), 865-877.

Evans, J. (2006). The heuristic-analytic theory of reasoning: Extension and evaluation. Psychonomic Bulletin & Review, 13(3), 378-395.

Fiedler, K. (1988). Emotional mood, cognitive style, and behavior regulation. In K. Fiedler & J. Forgas (Eds.), Affect, cognition, and social behavior (pp. 100–119). Toronto: Hogrefe International.

Ford, M. (2005). The game, the pieces, and the players: Generative resources from two instructional portrayals of experimentation. The Journal of the Learning Sciences, 14(4), 449-487.

Foresman, S. (2000). Scott Foresman Science. Addison-Wesley Educational Publishers.

Friedman, R. S., Förster, J., & Denzler, M. (2007). Interactive effects of mood and task framing on creative generation. Creativity Research Journal, 19(2-3), 141-162.

Habgood, M. P. & Ainsworth, S. E. (2011). Motivating children to learn effectively: Exploring the value of intrinsic integration in educational games, *Journal of the Learning Sciences, 20*(2), 169-206.

Kahneman, D. & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), Heuristics and biases (pp. 49–81). New York: Cambridge University Press.

Kuhn, D., Garcia-Mila, M., Zohar, A., & Andersen, C. (1995). Strategies of knowledge acquisition. Monographs of the Society for Research in Child Development, 60(4), Serial no. 245.

Lorch, R.F., Jr., Lorch, E.P., Calderhead, W.J., Dunlap, E.E., Hodell, E. C. & Freer, B.D. (2010). Learning the control of variables strategy in higher and lower achieving classrooms: Contributions of explicit instruction and experimentation. Journal of Educational Psychology, 104(3), 553.

Marshall, S. P. (1995). Schemas in problem solving. Cambridge University Press.

McElhaney, K. W., & Linn, M. C. (2010, June). Helping students make controlled experiments more informative. In Proceedings of the 9th International Conference of the Learning Sciences-Volume 1 (pp. 786-793). International Society of the Learning Sciences.

Nathan, M. J. & Koedinger, K. R. (2000). An investigation of teachers' beliefs of students' algebra development. Cognition and Instruction, 18(2), 209-237.

National Academies (1996). National Science Education Standards. Washington, DC: The National Academies Press. Retrieved from: http://www.nap.edu/openbook.php?record_id=4962

National Center for Education Statistics (2009). The Nation's Report Card: Science 2009. Retrieved from: http://nces.ed.gov/nationsreportcard/pdf/main2009/2011451.pdf

Next Generation Science Standards (2013). Washington, DC: Achieve, Inc. Retrieved from http://www.nextgenscience.org/

Nickerson, R. S. (1999). How we know—and sometimes misjudge—what others know: Imputing one's own knowledge to others. Psychological Bulletin, 125(6), 737-759.

Olson, K. R., & Spelke, E. S. (2008). Foundations of cooperation in young children. Cognition, 108, 222–231.

Pennsylvania Department of Education (2012). Academic Standards for Science & Technology. http://static.pdesas.org/content/documents/PreK-2_Science_and_Technology_Standards.pdf

Rai, D., & Beck, J. E. (2012). Math learning environment with game-like elements: an incremental approach for enhancing student engagement and learning effectiveness. In Intelligent Tutoring Systems (pp. 90-100). Springer Berlin Heidelberg.

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

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.

Sloane, S., Baillargeon, R., & Premack, D. (2012). Do infants have a sense of fairness? Psychological Science, 23, 196-204.

Sodian, B., Zaitchik, D., & Carey, S. (1991). Young children's differentiation of hypothetical beliefs from evidence. Child Development, 62, 753–766.

Tang, T. L., & Baumeister, R. F. (1984). Effects of personal values, perceived surveillance, and task labels on task preference: The ideology of turning play into work. Journal of Applied Psychology, 69(1), 99-105.

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

Tse, D., Langston, R. F., Kakeyama, M., Bethus, I., Spooner, P. A., Wood, E. R., Witter, M. & Morris, R. G. (2007). Schemas and memory consolidation. Science, 316(5821), 76-82.

Vosniadou, S., Vamvakoussi, X. & Skopeliti, I. (2008). The framework theory approach to the problem of conceptual change. In S. Vosniadou (Ed.), *International handbook of research on conceptual change*. New York, NY: Routledge.

Webster, J., & Martocchio, J. J. (1993). Turning work into play: Implications for microcomputer software training. Journal of Management, 19(1), 127-146.