Effects of Surface-feature Concreteness on Procedural Learning and Transfer

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

Abstract

One general obstacle to learning is a tendency for learners to focus on surface—rather than deep—features of instructional materials; we have found this tendency among students learning experimental design. The "concreteness" of instructional representations—which is correlated with the number of surface features—has been found to affect initial learning and transfer. We investigated the effect of the concreteness of experiments presented during instruction on sixth- and seventh-graders' learning and transfer of experimental design skills. For this instructional topic, surface features could potentially support *or* undermine understanding. We hypothesized that the effect of concreteness would depend on students' tendency to focus on deep versus surface features. That is, we predicted that students who tend to focus on surface features would benefit most when representations were initially abstract and concrete features gradually added (Abstract-fading condition). However, we predicted that students who tend to focus on deep features would benefit most if initially given a (minimally) concrete representation then either (a) two more minimally concrete examples (Concrete-only condition) or (b) concrete features were gradually "faded" (Concrete-fading condition). There was no effect of condition among surface feature-focus students. However, the Concrete-only condition facilitated transfer among deep feature-focus students by supporting understanding of the rationale for controlling variables.

Keywords: Concreteness of surface features; transfer; experimental design; elementary school students; middle school students.

Effects of Surface-feature Concreteness on Procedural Learning and Transfer

Supporting initial learning and transfer of that knowledge to other domains are primary goals of education. The "concreteness" of instructional materials—or their similarity to real-world entities—is one factor that affects initial learning and transfer (e.g., De Bock et al., 2011; Goldstone & Son, 2005; Koedinger, Alibali, & Nathan, 2008; McNeil, Uttal, Jarvin, & Sternberg, 2009; Kaminski, Sloutsky, & Heckler, 2006, 2008, 2013). However, the effect of concreteness may depend on student characteristics (e.g., Goldstone & Sakamoto, 2003; Author, 2014), including the tendency to focus on surface or deep features of the learning domain.

In the present study, we investigated the effect of concreteness of representations of experiments on sixth- and seventh-grade students' initial learning and subsequent transfer of experimental design skills to new domains. Specifically, we investigated the effect of concreteness on students' learning of the control of variables strategy (CVS), a procedure in which only the variable one is testing is contrasted across conditions. In the following, we begin by elaborating on our definition of concreteness. Then we discuss mechanisms by which concreteness may generally affect initial learning and transfer before discussing how concreteness may affect learning of CVS in particular. Based on these mechanisms, we then discuss our predictions for how concreteness will interact with students' tendency to focus on surface vs. deep features in the present study.

Definition of concreteness

The "concreteness" of a representation is defined here as it is widely used in the literature (e.g., Goldstone & Son, 2005; Koedinger, Alibali, & Nathan, 2008; McNeil, Uttal, Jarvin, & Sternberg, 2009): as the level of realism of the representation or its similarity to an actual physical object. The concreteness of a representation is positively correlated with the amount of information it communicates and inversely related to the number of entities it could feasibly represent (Kaminski et al., 2013).

Because the type of surface of a ramp is one variable used in instruction in the present study, we illustrate how surface may vary in concreteness (see Figure 1). An actual physical surface such as a piece of sandpaper can be considered the most "concrete" representation of "surface" and possesses more associated details and properties than even a realistic representation such as a photograph of the sandpaper. For example, the actual sandpaper has a specific texture and weight, which the photograph does not convey. Visual representations of entities are generally more concrete than even the most concrete textual representations. For example, the photo of sandpaper conveys information about the exact type of sandpaper and its coarseness that the word “sandpaper” does not. Textual descriptions of entities may also vary in concreteness. For example, “sandpaper” is more concrete than “rough surface,” which is more concrete than “surface,” which in turn is more concrete than a generic variable, “Variable X.” Conversely, the most abstract representation (“Variable X”) can potentially represent the greatest number of entities, including surface and countless other variables. The word “sandpaper” can potentially represent more real-world entities than even the most abstractly-represented visual image of sandpaper, because the word does not convey discriminatory information such as shape or color (even if only black/white or grayscale). At the other extreme, a piece of sandpaper is a unique entity.

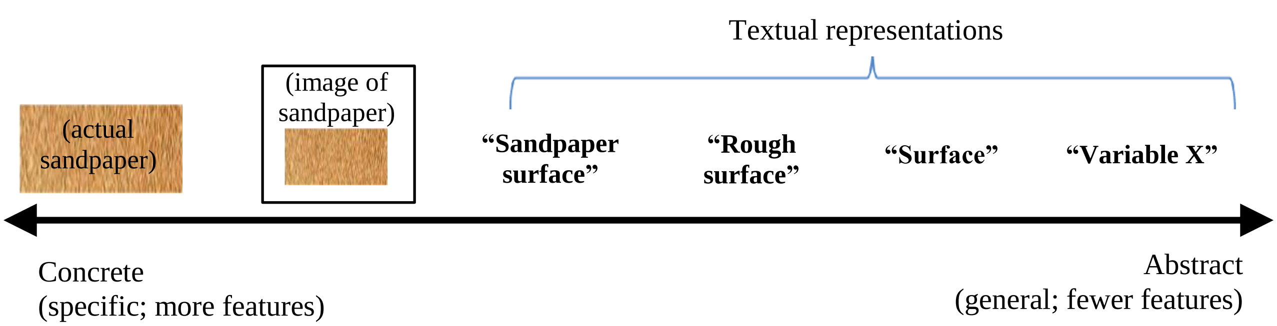

Figure 1. Example of how a representation of “surface” may vary in concreteness.

General effects of concreteness on initial learning and transfer

Depending on whether the additional information conveyed by the concrete features is relevant to the underlying concepts/principles/procedures to be learned, concrete representations may support or impair initial learning/performance. Concrete representations may undermine learning if they include features that distract learners from relevant aspects of the task (e.g., DeLoache, 1991; McNeil et al., 2009) or elicit conceptions that are irrelevant or even contradictory to the task (M1).

In contrast, relevant concrete features may support reasoning in the training domain (M2) (Author, 2014; Cheng & Holyoak, 1985; 1989; Goldstone & Sakamoto, 2003; Goldstone & Son, 2005). In procedural learning, relevant concrete features may elicit schemas that support—or prevent the need for—inferencing. For example, framing a reasoning task in terms of an intuitive, context-specific or concrete, “permission” schema resulted in better performance than when the task was framed using abstract logic (Cheng & Holyoak, 1985; 1989). Relevant concrete features may also support procedural learning by providing information about *why* specific procedures are necessary—and why *not* following the procedures is problematic (cf. Siegler, 2002).

Concrete features may also provide additional information about the deep structure of the task (M3). For example, in Kaminski et al. (2008), concrete training figures were depicted as measuring cups that were 1/3, 2/3, or 3/3 “full.” These proportions were analogous to the deep structure of a modular addition-like task (specifically, the numbers 1, 2, and 3) and the “full” cup could convey the idea that one cannot go beyond a limit. The highlighting of deep structure in the training domain may also allow recognition of analogous underlying structure in other domains and enable transfer of inter-element relationships via structure mapping (cf. Gentner, 1983; 2010).[1]

[1] Concrete features in a training domain may support transfer if the same or similar concrete features are present in the transfer domain (e.g., Fong & Nisbett, 1991; Holyoak & Koh, 1987) by reminding learners of the initial domain and eliciting either cross-domain structure-mapping (cf. Gentner, 1983; 2010) or memory of abstract rules/principles they had previously learned in that domain (Fong & Nisbett, 1991). However, because we are interested in transfer to other domains, with different surface features, this mechanism is not relevant in this paper.

Table 1. Potential mechanisms of concrete features in experiments on initial learning and transfer of CVS.

		Assuming effect, direction of effect	
General Mechanism	Evidence of?	Learning	Transfer
(M1) Elicit irrelevant or contradictory conceptions	Yes (general & CVS)	Negative (variable-effect and engineering goals)	Negative
(M2) Support reasoning/understanding	Yes (general)	Positive (rationale for CVS)	Positive (via more coherent conception)
(M3) Provide information about deep structure	Yes (general) No (CVS)	Positive	Positive (via structure mapping)
(M4) Prompt context-specific reasoning	Weak, if any	(n/a)	Negative

However, concrete features present in the training domain may lead to context-specific forms of reasoning that may be tied only to the initial learning domain and suppress transfer (Goldstone & Son, 2005) (M4). For example, Bassok and Holyoak (1989) found that high school and college students were better able to transfer from algebra to physics than the reverse, which they explained as: “In studying physics, students learn that the physical concepts involved in word problems are critical to the applicability of the relevant equations. Accordingly, they do not expect, and fail to recognize, any direct relation between physics problem-solving procedures and isomorphic problems drawn from non-physics domains” (p. 165). However, a greater tendency to transfer from “content-free” (i.e., algebra) to “content-specific” (i.e., physics) domains than the reverse may not apply to concepts that are not deeply-entrenched within complex subject domains like physics. For example, in Gick and Holyoak (1983) (Exp 2), college undergraduates were equally likely to apply context-specific and abstract versions of a principle in a transfer domain with novel surface features. This finding suggests that the context-specific version did *not* suppress transfer. This lack of an effect is also consistent with Holyoak and Koh’s (1987) summation of activation model, in which the likelihood of source activation is proportional to the number of features—

both surface and structural—in common with the target. When only the underlying structure (but no surface features) is common between domains, source activation is equally probable, as found.

Effects of concreteness on CVS learning and transfer

We now consider how surface feature concreteness may affect initial learning of experimental design, and in particular, learning the control of variables strategy (CVS). In our instruction, the surface level includes details of the experiments presented to students, including the cover story for the experiment, the context-specific variables, and their values. However, we want students to learn the underlying relational structure of CVS, which involves contrasting values of the focal variable and controlling all other relevant variables (see Table 2).

Table 2. Structural (or deep) features of a controlled experimental comparison.

Common variables	Cross-condition	Condition1	Condition2
Focal variable	Different	ValueF1	ValueF2
VariableA	Same	ValueA1	ValueA1
VariableB	Same	ValueB1	ValueB1

We have found that the surface features present in the experimental examples in assessments and instruction appear to distract some students from our intended instructional focus. (M1) In our earlier studies (Authors, 2013), we found that students sometimes misinterpreted the goal of assessment and instructional tasks as either to discuss their beliefs about the effects of the problem-specific variables or apply such beliefs in their designs to “engineer” particular outcomes (Schauble et al., 1991). For example, when asked to design experiments that would allow them to find out about particular parts of ramps, some students simply discussed their beliefs about whether and how particular variable(s) (e.g., the ramp steepness) would affect how far balls roll (e.g., “The steeper ramp would make the ball roll farther”) or indicated a desire to set up ramps in particular ways that satisfied *engineering goals* such as making the balls roll farthest/fastest/different distances.

(M2) However, these same concrete features may also support students' understanding of CVS and the rationale for controlling variables—and why not controlling variables may lead to invalid conclusions of causality. For example, in our instruction, students are asked to evaluate an experiment meant to find out if balls roll different distances on ramps with different surfaces, where ramp steepness is a confound. Students' beliefs about the effect of ramp steepness on rolling distance may support their understanding that—because steeper ramps will cause balls to roll farther than non-steep ramps—the ramp steepness must be controlled. Developing an understanding of the rationale for CVS supports transfer to other domains (Authors, 2010; Authors, 2011; Sao Pedro, Gobert, & Sebuwufum, 2011) especially after long delays (Sao Pedro, Gobert, & Raziuddin, 2010). The improved transfer performance may be caused by the rationale's linking of the procedural rules "contrast the focal variable" and "control all other variables" [so that only the focal variable can affect the outcome], resulting in a more coherent and inter-connected conception of CVS. Research in category learning suggests that explanatory principles in particular increase conceptual coherence (Murphy & Medin, 1985); further, more coherent conceptions are more likely to be applied in novel situations (Patalano, Chin-Parker, & Ross, 2006). Concrete features may indirectly support CVS transfer by increasing the coherence of concepts students learn in the training domain. This suggests a trade-off, where given surface features from concrete representations may either distract students and impair learning (M1) *or* support students' understanding of the rationale for controlling variables, thereby supporting transfer (M2).

(M3) Although for some topics, concrete features may support initial learning of underlying principles by providing information about the underlying structure (Author, 2014), it is unclear how concrete features could support procedural understanding of CVS, per se. For example, there is no reason to expect that the surface (rough/smooth) or steepness would elicit appropriate conceptions related to contrasting or controlling. (M4) In addition, concrete representations may lead to the formation of domain-specific knowledge or reasoning, which may not transfer easily to other domains (Goldstone & Son, 2005). For example, students may only develop an understanding of the rationale for controlling variables in the context of ramps (e.g., if they learn that "The ramps should be the same except for the type of surfaces so that only the surfaces can affect how far the balls roll") and may not realize its relevance in the transfer domain.

Summary of potential mechanisms affecting CVS learning and transfer. In summary, we expected that the first two effects of concreteness outlined in Table 1 (M1 and M2) would be the primary mechanisms underlying initial learning and transfer. That is, we expected there to be a trade-off between concrete representations (a) eliciting students' related beliefs and engineering goals and (b) supporting initial understanding of the rationale for controlling variables and subsequent transfer. Because of this trade-off, the effect of concreteness may depend on whether students tend to focus on the deep (structural) or surface level of instruction. That is, students who tend to focus on surface features may gain a better understanding from an abstract representation whereas students who tend to focus on deep features may gain a better CVS understanding from a concrete representation.

	Ramp 1	Ramp 2
Surface	Rough	Smooth
Steepness	Steep	Not Steep
Ball	Heavy	Heavy
Starting Position	Middle	Top

	Ramp 1	Ramp 2
Surface	Different	
Steepness	Different	
Ball	Different	
Starting Position	Different	

	Setup 1	Setup 2
Variable X	Different	
Variable A	Different	
Variable B	Same	
Variable C	Different	

Figure 2. (a) Concrete (left), (b) Intermediate (right), and (c) Abstract (bottom) representations of experiments given during instruction in the present study.

Present study

In the present study, students evaluated three experiments in one of three conditions: (a) Concrete-only, in which all experiments were concrete and involved ramps (Figure 2a), (b) Concrete-fading, in which the first experiment was concrete, the second intermediate (Figure 2b), and the final one abstract (Figure 2c), and (c) Abstract-fading, in which the first experiment was abstract, the second intermediate, and the final one concrete. An important note is that, across conditions, experiments were represented as text within tables. Even in the most concrete form, only enough detail was given to convey critical information involving potential causality. Thus, the "concrete" condition may be more accurately described as a "*minimally*-concrete" condition. The range of representations given in the current study along the concrete-abstract continuum is shown in Figure 3 (circled). As can be seen, even the "concrete" representation of the present study tends to fall within the middle of the continuum.

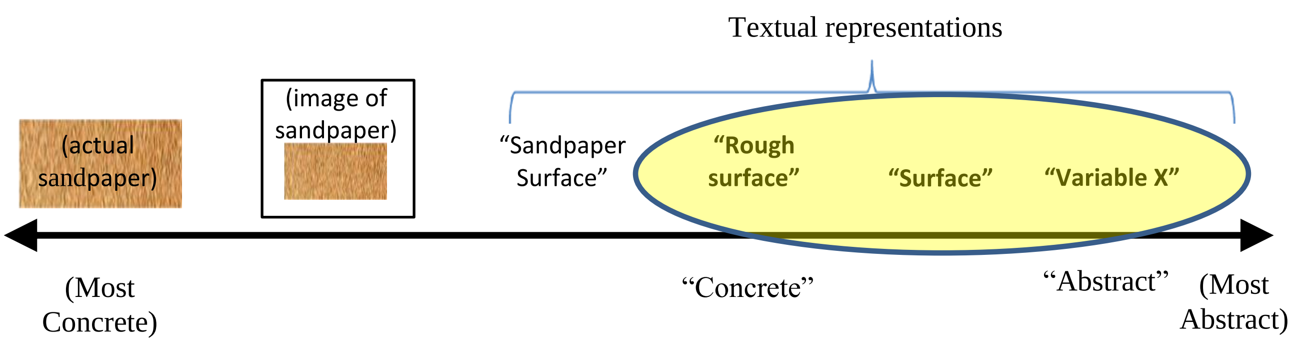

Figure 3. The range of concreteness of representations in the present study (circled).

Predictions. Concretely-represented experiments may elicit surface-level focus (e.g., variable-effects or engineering goals) and suppress science goal understanding (and thus subsequent CVS development). We predicted that students who (based on their pretest responses) are more likely to focus on surface-level features would benefit most from the Abstract-fading condition because the initial abstract representation would reduce the distracting effect of concrete variables/values and indirectly support understanding of the lesson goal. In addition, the subsequent addition of concrete features in the Abstract-fading condition would support understanding the rationale for controlling variables. Thus, we predicted that the Abstract-fading condition would best support students prone focusing on surface-level features (P1). However, there is a potential trade-off, where the initial absence of concrete features may reduce the chance of distraction from surface features for some students, but the absence of concrete details may also result in poorer initial understanding for others. In addition, because the CVS instruction given in the present study emphasizes the rationale for controlling variables, concrete representations—which may help support this understanding—may be particularly important.

Among deep-focus students, who do *not* misinterpret the assessment or instructional goals, we predicted that instruction using only “concrete” experiments would be most effective because this context would allow students the most opportunities to understand the rationale for controlling variables, supported by concrete features (P2). However, because all three of the experiments students evaluated during instruction involved ramps, it is possible that students’ schemas for experimental design will be tied to the surface features of ramps and thus hinder transfer to other domains (M4) (cf. Goldstone & Son, 2005). Thus, another possibility is that the Concrete-fading condition, in which the final experiment is completely abstract (devoid of surface features), will be most effective by promoting transfer to other domains (alt-P2). However, because evidence of M4 is weak, we consider alt-P2 a possibility rather than a prediction. Predictions are summarized in Table 3.

Table 3. Summary of predictions and respective mechanism(s).

Prediction	Mechanism(s)
P1: Abstract-fading best for likely surface-focusing students	Reduce (initial) distraction (M1); Concrete supports CVS rationale (M2)
P2: Concrete-only best for deep-focusers	Concreteness supports CVS rationale (M2);
P2a: Concrete-only > Abstract-fading	(M2)
P2b: Concrete-only > Concrete-fading	(M2; no M4)
alt-P2: Concrete-fading best for deep-	If concreteness supports initial understanding (M2);

focusing students	but causes domain-specific reasoning (M4)

Method

Participants

A total of 243 students (140 sixth-graders and 103 seventh-graders) at a suburban public middle school serving a primarily middle-SES student population completed the pretest. The 175 students (106 sixth-graders and 69 seventh-graders) who did not show above-chance levels of CVS understanding (i.e., who set up fewer than two unconfounded experiments on the pretest) and completed the study were included in analyses.

Materials and Procedure

This study took place during students' regular 40-minute science class periods. In each class, students worked individually at a computer in the school's computer lab. Each phase of the study—the pretest, instruction, and posttest—was completed during one class period on three consecutive days. Students first completed one of two isomorphic versions of a 6-item computerized "story problems" pretest (the other version was given as the posttest). For each of three domains (e.g., designing rockets), students designed an experiment, then they evaluated (and, if necessary, fixed) a given experiment. Students explained each design and evaluation response. See Figures 4 and 5 for design and evaluate items (respectively). Within each class, students were randomly assigned to one of three instructional conditions: Concrete-only, Concrete-fading, or Abstract-fading.

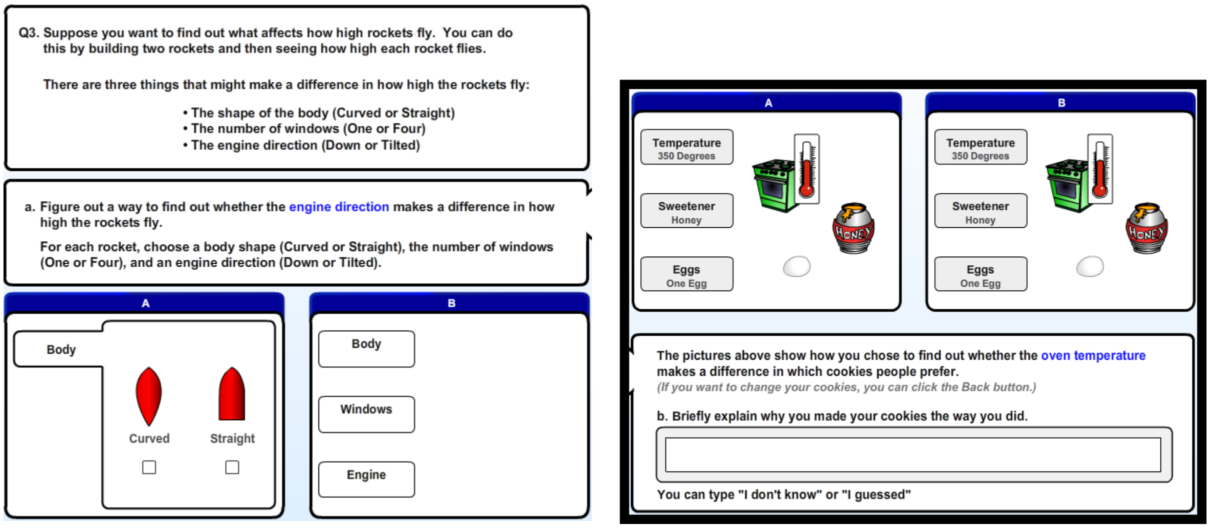

Figure 4. Test design items (a) rockets (set-up in progress), and (b) cookies (set-up completed).

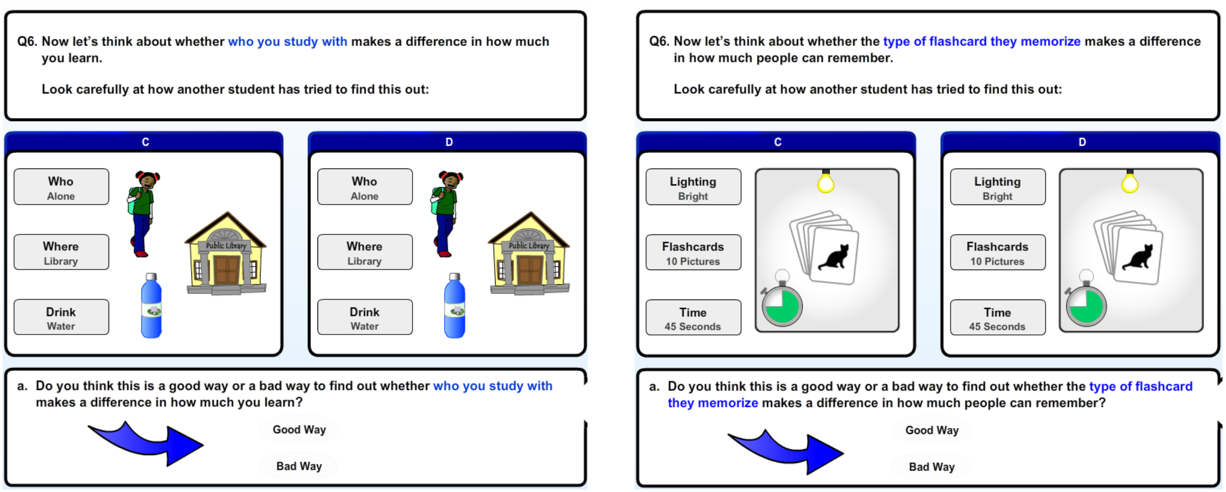

Figure 5. Final question of pretest (version A and version B).

During instruction, students evaluated three confounded experiments as follows: (a) Concrete-only: all experiments were concrete ramps experiments, (b) Concrete-fading: the first experiment was concrete, the second intermediate, and the final one abstract, and (c) Abstract-fading: the first experiment was abstract, the second intermediate, and the final one concrete. (See Figure 2 for each type of representation.) Concrete experiments included all concrete variables and values (e.g., the ramp "surface" could be "smooth" or "rough"). All concrete variables and values involved ramps and were described textually only; no pictorial representations were given. For intermediate experiments, variables were concrete, but variable levels were not referenced. Rather, variables were described as having either “the same” or “different” values across ramps. For abstract experiments, neither variables nor their values corresponded to concrete entities. Variables were labeled as “Variable A[/B/C],” and values were described only as “the same” or “different” across conditions. For each example experiment, students answered a series of multiple-choice (MC) and open-ended (OE) questions. Specifically, students:

- Indicated whether it was a good way to find out whether the [focal variable (“Variable X” or “ramp surface/type of ball/starting position”)] makes a difference (MC_Q1) then received feedback on their response (e.g., “That’s right. It’s not a good way to find out if the [focal variable] makes a difference.”)
- Explained why it was (actually) a bad way (OE_Q1) and received feedback on why it was a bad way (i.e., “Only the [focal variable] should be different.”)
- Indicated if the experiment would allow them to “know for sure” that the [focal variable] caused a difference in outcomes (MC_Q2) and received feedback on their response (e.g., “That’s right. You could not know for sure that the [focal variable] caused the difference in [the outcome].”)
- Responded to: “Why could you [actually] not know for sure?” (OE_Q2)
- Indicated whether each of the three non-focal variables could cause a hypothetical difference in outcomes (MC_Q3). Because these questions directly assess students’ understanding of the link between the experimental set-up and determinacy of the outcome’s cause, they most directly assess the rationale for controlling variables. Students received corrective (i.e., Right/Wrong) feedback on each response.
- Corrected the set-up and explained why the design was now “a good way to find out whether the [focal variable] makes a difference” (OE_Q3) then received feedback (e.g., “Everything is the same except for the [focal variable]….If there is a difference in [outcome], it has to be because of the [focal variable]…”).

The next day, all students completed the posttest (isomorphic to the pretest), introduced with the following: “Now, we will ask you another six questions. This is not a test and it will not be graded. Please take your time and think about each question before you answer.” Thus, the posttest assessed students’ transfer of what they learned during instruction to novel contexts with no direct cuing. Students received one point for each correctly-designed experiment (out of three), correct evaluation response (out of three), and each given experiment they converted to a non-confounded one (out of three). Thus, the maximum possible posttest score was nine.

Qualitative Analyses

Level of focus. Whether students were likely to focus on surface or deep features during instruction was determined by their responses to the final pretest question (see Figure 5), where the given set-ups were non-contrastive (i.e., the conditions were identical).

Surface-focus: Students who expressed “engineering” goals or were unable to give any response were considered likely to focus on surface features during instruction. Responses were coded as “engineering” if they indicated that the set-up was “good” (or “bad”) because it would (or would not) result in a desired outcome (e.g., the non-contrastive set-up was “good” because “pictures are easier to remember” or “you have quiet and something healthy to drink plus more resources [at the library]”). Engineering responses were the most common type (49%).

In addition, 21% of the students were unable to explain why the set-up was good or bad (e.g., “I guessed” or “I don’t know”). The majority of these students (72%) previously expressed engineering goals in their pretest responses, 17% only responded that they guessed or did not know throughout the entire pretest, and only 11% previously expressed a science-only or contrastive response. There were no pre- or posttest differences between students who expressed “engineering” and students who were unable to explain their evaluation response, again indicating that these groups are similar.

Deep-focus: Students who gave “Science goal” and/or “Contrastive” responses indicating the need to contrast one or more variables were considered likely to focus on deep features during instruction.

"Science goal" responses indicated an understanding that the goal was to *find out about* something. However, only two students gave Science goal responses.

"Contrastive" responses included both general indications that *something* should be contrasted (e.g., "that's a bad way to find out because there are no differences between the [two]") and that specifically the focal variable should be contrasted (e.g., "They are using the same person"). This type of response was given by 29% of all students. (a) In 31% of Contrastive responses, students also explicitly expressed a science goal for contrasting (e.g., "It is bad because it is the same thing and you wouldn't really get to test anything"). Of the remaining students who gave Contrastive responses without indicating a science goal understanding, (b) 43% had previously expressed a science goal understanding of the pretest task, but (c) the remaining 57% did not. However, these three sub-groups of Contrastive responders (a-c) did not differ in pretest (p = .53) or posttest (p = .76), which suggests that these sub-groups are similar.

Likely surface-focus students (who gave engineering or no explanation) scored significantly lower than likely deep-focus students on both the pretest ($p < .001$) and posttest ($p < .001$). Thus, overall, there were no within-focus group differences for either surface- or deep-focus students, but there were between-group (surface-focus vs. deep-focus) differences in pre- and posttest performance. This suggests within-focus group homogeneity and supports this classification of responses within focus category.

Instructional explanations. To assess students' explicit understanding of experimental design during instruction, we coded their open-ended responses to "Why is [the experiment] a bad way to find out about [the focal variable]?" (OE_Q1) and "Why could you [actually] not know for sure [that the focal variable caused a difference in outcomes]?" (OE_Q2) as one of the following:

- *Rationale*: Responses indicating understanding of the rationale for controlling non-tested variables (e.g., "Everything else is different too, not just the surface so anything else could cause it." or "The surface of the ramp, the steepness of the ramp, and the starting position all could have affected the distance the ball rolled because they were all different.")

- *Procedural*: Responses expressing procedural knowledge of controlling variables (e.g., [it's a bad way because] "variable A and C are different" or "because the steepness and surfaces are different"; [you could not know for sure] "because Variable A and C need to be the same instead of different")
- *Other*: Responses not clearly indicating either conceptual or procedural understanding, including:
    - *Engineering/Variable-effect:* Responses indicating an outcome goal or expressing one's beliefs about specific variable effects (e.g., "the balls don't roll too good on rough ramps"; "you could not know because if it's rough then it would be hard to roll, but if it's smooth it rolls a lot easier.").
    - *Don't know:* Indications that students could not explain their response (e.g., "I don't know [why it's a bad way/why you could not know for sure that the variable has an effect]")
    - *Vague:* Unclear responses such as "the variables need to be changed" or [it's a bad way] "because it does not talk about the surface."
    - *All-Different*: Responses indicating that all of the variables should be contrasted across conditions (e.g., [it's a bad way because] "the type of ball are both heavy").
    - *Too Abstract:* Responses indicating the experiment was bad because it lacked detail (e.g., [it's a bad way because] "they do not show the setups" "not enough info" "you wouldn't know how steep the ramp was, what type of ball you used, and the starting position")

Results

Subject variables. Among students included in analyses, there was no effect of condition for pretest score (p = .57) or focus expressed on the final pretest item (p = .79). Among students who did not demonstrate above-chance level CVS understanding on the pretest (and were included in analyses), sixth- and seventh-graders did not differ in pretest score (p = .22). However, consistent with prior findings that younger students are more likely to express engineering goals than older students (Authors, 2013), sixth-graders were significantly more likely to be classified as surface-focus students than seventh-graders (78% vs. 57%, respectively), $\chi^2(1, N = 175) = 9.39$, p = .002.

Main effect of condition. Although we did not have an overall prediction, among all students (sixth- and seventh-graders), there was a marginally significant effect of condition on total posttest score, $F(2, 171) = 3.03$, $p = .05$, $\eta_P^2 = .05$, shown in Figure 6. Students in the Concrete-only condition scored significantly higher than those in the Abstract-fading ($p = .03$) and Concrete-fading ($p = .045$) conditions. However, there was no difference between the Abstract-fading and Concrete-fading conditions ($p = .89$).

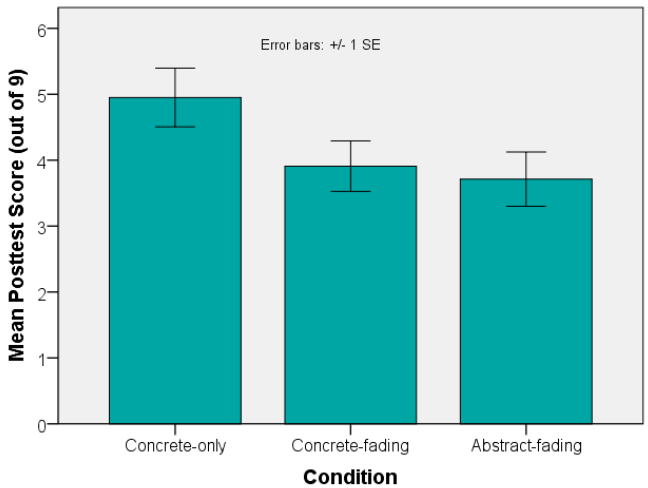

Figure 6. Main effect of condition (all students).

Interactions with condition. We predicted that surface-focus students would show better learning and transfer outcomes from the Abstract-fading condition (P1) and that deep-focus students would show better outcomes from the Concrete-only (P2) or Concrete-fading (alt-P2) conditions. When grade and focus were included with pretest and condition in ANCOVA, condition interacted with focus, $F(2, 162) = 3.64$, $p = .03$, $\eta_P^2 = .04$, pretest, $F(2, 162) = 3.23$, $p = .04$, $\eta_P^2 = .04$, and marginally with grade, $F(2, 162) = 2.34$, $p < .10$, $\eta_P^2 = .028$. We discuss each of these interactions next.

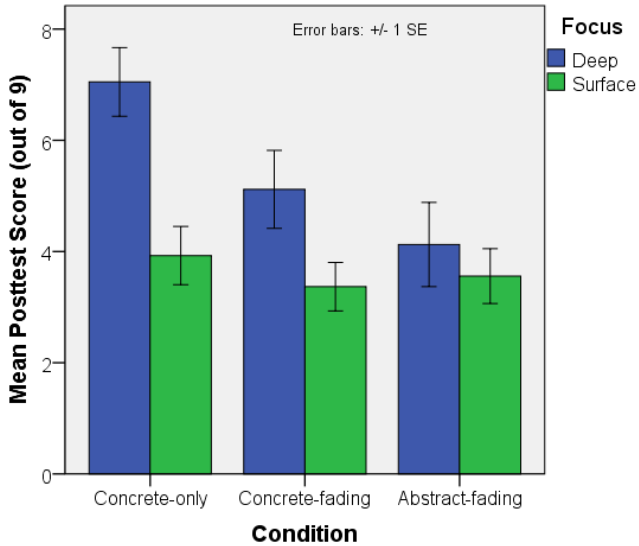

Figure 7. Condition by focus interaction.

Condition by focus interaction. As shown in Figure 7, the effect of condition was significant only among deep-focus students, $F(2, 49) = 4.92$, $p = .01$, $\eta_P^2 = .17$. Among surface-focus students, there was no effect of condition ($p = .69$). Among deep-focus students, those in the Concrete-only condition scored higher on the posttest than students in either the Concrete-fading ($p = .03$) or Abstract-fading ($p = .004$) conditions. The Concrete-fading and Abstract-fading conditions did not differ ($p = .45$). Thus, the overall advantage of the Concrete-only condition was due to its effect on deep-focus students in particular.

Thus, our prediction (P1) that students who were more likely to focus on surface features (based on their final pretest responses) would benefit most from the Abstract-fading condition was *not* supported. Consistent with P2, among deep-focus students, those in the Concrete-only condition showed better transfer than in the other conditions. However, counter to alt-P2, deep-focus students in the Concrete-fading condition did not.

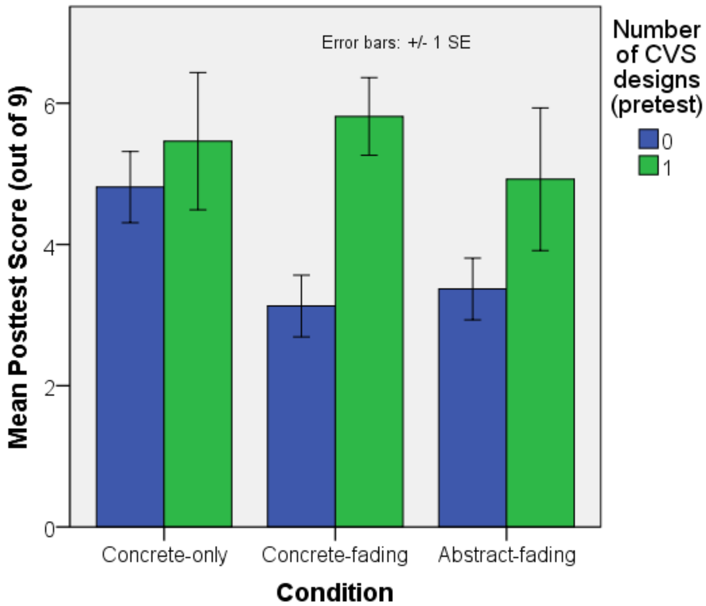

Figure 8. Condition by pretest interaction.

Condition by pretest interaction. As shown in Figure 8, the effect of condition was significant only among students who failed to design *any* unconfounded experiments on the pretest, $F(2, 129) = 3.87$, $p = .02$, $\eta_p^2 = .06$ (there was no effect among students who designed an unconfounded experiment on the pretest, $p = .60$). Posttest scores were higher in the Concrete-only than in the Abstract-fading ($p = .03$) and Concrete-fading ($p = .01$) conditions. The Concrete-fading and Abstract-fading conditions did not differ ($p = .74$). The Concrete-only condition benefited deep-focus students who understood the need to contrast variables but failed to set up any unconfounded experiments on the pretest. Thus, the Concrete-only condition appeared to support learning to control variables in particular. A summary of findings vis-à-vis each prediction is shown in Table 4.

Table 4. Summary of outcomes per prediction.

Prediction	Consistent with outcomes?
P1: Abstract-fading best for surface-focus	No
P2: Concrete-only best for deep-focus:	**Yes**
P2a: Concrete-only > Abstract-fading	Yes
P2b: Concrete-only > Concrete-fading	Yes
Alt-P2: Concrete-fading best for deep-focus students	No (Concrete-fading < Concrete-only)

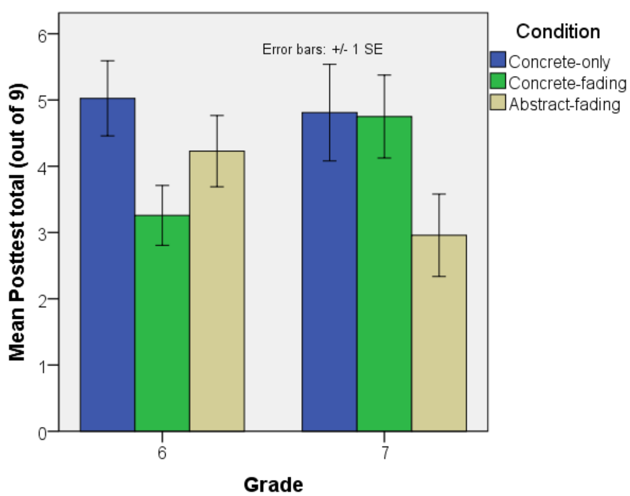

Figure 9. Condition by grade interaction trend.

Condition by grade interaction trend. Finally, there was a marginally significant condition by grade interaction trend, $F(2, 162) = 2.34$, $p < .10$, (see Figure 9). Among sixth-graders, Concrete-only students out-performed Concrete-fading students ($p = .03$), but there were no other pair-wise differences. However, seventh-graders in Concrete-only and Concrete-fading conditions performed similarly on the posttest ($p = .95$) and better than seventh-graders in the Abstract-fading condition ($p = .05$ for each). These results suggest differential effects of the fading conditions: the younger students performed worse in the Concrete-fading condition, but older students performed worse in the Abstract-fading condition.

Learning Mechanisms

We now investigate the learning mechanisms underlying the overall advantage of the Concrete-only condition among deep-focus students in particular. First, we compare the Concrete-only and Abstract-fading conditions then compare the Concrete-only and Concrete-fading conditions. In the first comparison, we also investigate why there was no benefit of the Abstract-fading condition among surface-focus students, as we had predicted (P1). For each comparison, we also examine whether any differences in final understanding may explain advantages of the Concrete-only condition. In the comparison between the Concrete-only and Concrete-fading conditions, we also look at the effect of concreteness on transferring one's final understanding to the posttest.

Concrete-only vs. Abstract-fading. We predicted that the Abstract-fading condition would help prevent surface-focus students (P1) from focusing on surface features (M2). However, surface-focus students in the Concrete-only and Abstract-fading conditions performed similarly on the posttest. We first explore the expected trade-off between distraction-reduction effects of abstract representations (M1) and supportive effects of minimally-concrete representations (M2) in students' responses to Q2_OE of the initial experiment (i.e. "Why would you not know for sure [that the focal variable caused the difference]?").

Initial responses: M1 vs. M2. As expected, overall, students in the Abstract-fading condition were less likely to give engineering or variable-effects responses than students in the Concrete-only and Concrete-fading conditions (combined), $\chi^2(1, N = 119) = 6.84$, $p = .009$. Only one student in the Abstract-fading condition gave this type of response, compared to 9 students in the Concrete-only condition (see Table 5). This overall effect was due to surface-focus students, $\chi^2(1, N = 84) = 6.48$, $p = .01$. (Only one deep-focus student expressed an engineering/variable effect belief, and the effect of condition was not significant, $p = .35$.)

Table 5. Frequency of response for explanations given for the initial experiment by condition.

	Correct responses		Incorrect responses					
	Rationale	Procedural	Engineer/ VE	Don't know	Vague	All-Different	Too Abstract	Other
Abs-F	5 (8%)	4	1	6	18	1	14	10
Conc	10 (16%)	12	9	3	11	6	1	9
Con-F	10 (18%)	9	10	3	6	3	3	11
Total	25	25	20	12	35	10	18	30

However, students in the Concrete-only condition *still* gave more correct (rationale or procedural) responses than students in the Abstract-fading condition, $\chi^2(1, N = 119) = 7.08$, $p = .008$. This relationship was significant for both surface- and deep-focus students. Thus, the supportive effects of concrete representations (M2) outweighed the distraction-reduction effects of abstract representations (M1), even among surface-focus students.

Table 6. Number of correct instructional responses (out of five) by condition and example experiment.

Experiment	Concrete-only	Abstract-fading	Concrete-fading
#1	3.63 (1.26) ^{a1}	2.90 (1.54) $^{a1, a2}$	3.67 (1.04) ^{a2}
#2	4.45 (0.81) ^{b}	4.09 (0.98) ^{b}	4.25 (0.91)
#3	4.65 (0.66) $^{b\ c}$	3.91 (1.13) ^{c}	4.24 (1.15) ^{b}
Overall mean	4.24 (0.67) ^{c}	3.63 (0.82) ^{c}	4.05 (0.77)

$^{a1, a2}$ Means differ at $p < .005$. ^{b} Means differ at $p < .05$.

^{c} Means differ at $p < .001$.

Students' responses to instructional multiple-choice questions also showed initial advantages of concrete representations (see Table 6). Students in the Concrete-only condition averaged more correct responses for the first example experiment than students in the Abstract-fading condition, $F(1, 116) = 8.10$, $p = .005$, $\eta_P^2 = .065$. This initial advantage of the Concrete-only condition was similar for deep-focus and surface-focus students.

However, students in the Concrete conditions were significantly more accurate only on the MC_Q3 (but not MC_Q1 or MC_Q2) questions, which asked students to indicate whether each non-focal variable could cause a hypothetical difference in outcome, $F(1, 116) = 17.47$, $p < .001$, $\eta_P^2 = .13$. This effect was significant for surface- and deep-focus students. MC_Q3 questions directly assessed students' understanding of the link between variable set-up and causal determinacy and may most directly assess students' understanding of the rationale for controlling variables. In sum, both the open-ended and multiple-choice responses indicate that concrete features supported initial CVS understanding—for both surface- and deep-focus students.

***Final understanding*.** We expected concrete representations to support students' understanding of the rationale for CVS (M2). Thus, students in the Concrete-only condition—who had spent the most time learning in the context of concrete representations—should be more likely to express this understanding by the end of instruction. We coded students' responses to Q1_OE for the final example experiment. Consistent with M2, students in the Concrete-only condition were more likely to express an explicit understanding of the rationale for controlling variables than students in the Abstract-fading condition (33% vs. 11%, respectively), $\chi^2(1, N = 116) = 8.52$, $p = .004$. The relationship between condition and rationale expression was significant among deep-focus students, $\chi^2(1, N = 34) = 9.63$, $p = .002$, but *not* surface-focus students, $\chi^2(1, N = 82) = 1.38$, $p = .24$.

For responses to the multiple-choice questions of the final example experiment, there was a significant condition by focus interaction, $F(1, 113) = 4.867$, $p = .03$, $\eta_P^2 = .041$. Among deep-focus students, there was a highly significant effect of condition, $F(1, 31) = 31.25$, $p < .001$, $\eta_P^2 = .50$, where Concrete-only students gave a higher proportion of correct responses (97%) than Abstract-fading students (71%). When the proportion of correct responses to the final example experiment was covaried, the effect of condition on transfer performance was no longer significant ($p = .49$). Surface-focus students also scored significantly higher in the Concrete-only than Abstract-fading condition (91% vs. 81%), $F(1, 81) = 5.11$, $p = .03$, $\eta_p^2 = .059$. (But this did not translate into better posttest performance.)

Summary Concrete-only vs. Abstract-fading. These results suggest that concrete representations supported students' initial understanding of the relationship between variable set-up and outcome determinacy. Furthermore, continuous exposure to concrete representations led to a better final understanding of the rationale for controlling variables—but only among deep-focus students. This better final understanding is a plausible account for the advantage of the Concrete-only condition in transfer among deep-focus students.

Concrete-only vs. Concrete-fading. We reasoned that whether the Concrete-fading or Concrete-only condition better supported transfer would depend on the trade-off between continuous exposure to concrete representations supporting understanding the rationale for controlling variables (M2) and concrete representations inhibiting transfer (M4). The Concrete-only condition better supported transfer—in particular, among students who expressed deep-focus on the final pretest item.[5] Here, we investigate whether and the extent to which each mechanism played a role in this outcome.

Final understanding. Constant exposure to concrete representations led to a better final CVS understanding (M2): Students in the Concrete-only condition were more likely to express an understanding of the rationale for controlling variables than students in the Concrete-fading condition (33% and 15%, respectively), $\chi^2(1, N = 114) = 5.26$, $p = .02$. This relationship was significant among both deep- and surface-focus students. Further, among the non-rationale responders, Concrete-only students were more likely to express a procedural understanding of CVS than Concrete-fading students (50% vs. 33%, respectively). Students in the Concrete-only condition also gave more correct responses to the

[5] As expected (because the same concrete representations were used in both conditions), there was no difference in the number of correct responses to multiple-choice questions for the initial instructional experiment ($p = .86$) or likelihood of expressing at least a procedural understanding of controlling for the initial experiment ($p = .81$).

multiple-choice questions for the final example experiment, $F(1, 113) = 6.82$, $p = .01$, $\eta_P^2 = .06$ (response accuracy did not differ on the first or second experiments; see Table 6). This relationship was significant among both deep- and surface-focus students. As with the initial responses of Concrete-only vs. Abstract-fading students, Concrete-only students were only more accurate in their responses to MC_Q3 questions.

Because there was a significant condition by multiple-choice response interaction ($p = .048$), multiple-choice response accuracy could not be co-varied to determine whether it could account for the transfer advantage of the Concrete-only condition. Among students who correctly answered *fewer than five* questions (below the mean 4.42), there was no effect of condition on posttest score. However, among "high-understanding" students who correctly answered all five multiple-choice questions, there was a further condition by focus interaction ($p = .04$). The effect of condition for high-understanding students was still significant among deep-focus students ($p = .01$).[6] Thus, there was still an advantage of the Concrete-only condition among deep-focus students who also answered all multiple-choice questions correctly. This was not due to a difference in the likelihood of expressing an understanding of the rationale for controlling variables ($p = .38$) or total number of correct multiple-choice responses given throughout instruction ($p = .92$).

Likewise, among sixth-graders, who showed better transfer in the Concrete-only than Concrete-fading condition (Figure 9), there was a significant condition by multiple-choice response interaction, $F(1, 67) = 4.41$, $p = .04$, $\eta_P^2 = .062$. Among sixth-graders who answered fewer than five questions correctly, there was no effect of condition, $F(1, 20) = 0.14$, $p = .71$, $\eta_P^2 = .007$. However, among "high-understanding" sixth-graders who correctly answered all five multiple-choice questions, Concrete-only students scored significantly higher than Concrete-fading students, $F(1, 46) = 4.85$, $p = .03$, $\eta_P^2 = .095$.

[6] Although surface-focus students in the Concrete-only condition demonstrated better final CVS understanding, this did not translate into better transfer performance because, unlike deep-focus students, surface-focus students were unable to apply their final understanding to the posttest.

Therefore, not only did the concrete representations support understanding of the need to control variables, the concrete representation seems to have supported an understanding that *better transferred to novel domains* than the abstract representation—especially among deep-focus students and the younger students. This result is counter to M4, that concrete representations of experiments used during instruction prompted a type of context-specific reasoning that did not transfer well. A summary of results of the analyses of learning mechanisms is shown in Table 7.

Table 7. Summary of outcomes regarding effects of learning mechanisms during instruction.

	Initial Understanding (Exp #1)	Final Understanding (Exp #3)
Concrete-only vs. Abstract-fading	Fewer ENG/VE responses: Abs-fading (surface-focus students) Better understanding: Concrete-only • Open-ended (deep- & surface-focus) • MC (deep & surface) ○ MC_Q3 (deep- & surface-focus)	Better understanding: Concrete-only • Open-ended rationale (deep-focus) • MC (deep & surface) ○ Accounts for better transfer in Concrete-only condition among deep-focus students
Concrete-only vs. Concrete-fading	(n/a; no differences, as expected)	Better understanding: Concrete-only • Open-ended rationale (deep & surface) • (Also open-ended procedural) • MC (deep & surface) ○ Does *not* account for deep-focus transfer. ○ Does not explain sixth-graders' better transfer performance.

Effect of concreteness on transfer (within-condition analysis) (M4)

As just shown, deep-focus students were actually *better* able to transfer their understanding of CVS to the posttest when the final experiment was in the most concrete form than when it was represented abstractly. To supplement this between-condition analysis, we also performed within-condition analyses of students' final responses, to see whether context-specific reasoning led to poorer transfer. Among Concrete-only students who expressed at least a procedural understanding of controlling variables at the end of instruction, those whose explanations included domain-specific entities (e.g., [it is not a good experiment] "because the surface and steepness are different") had similar posttest scores to those whose explanations were generalized (e.g., [it is not a good experiment] "because the other characteristics are different"), (*M* = 6.62, *SD* = 3.12; *M* = 6.38, *SD* = 2.99, respectively), $F(1, 31) = 0.06$, $p = .82$. Similarly, the generality of students' CVS explanations on the final experiment was unrelated to posttest performance in the Abstract-fading ($p = .55$) or Concrete-fading ($p = .37$) conditions. Thus, neither the between- nor within-condition analysis suggests that concreteness led to a type of reasoning that impaired transfer (M4).

Discussion

Previous work (e.g., Authors, 2012; Schauble et al., 1991) has shown that engaging in tasks involving experimental design elicits alternative goals among some students. These goals may be elicited by "concrete" surface features of the tasks. In this paper, we investigated the effect of varying the concreteness of experiments presented during instruction on sixth- and seventh-grade students' learning and transfer of the control of variables strategy (CVS).

Examining the effect of concreteness on learning CVS is unique because the same concrete features (i.e., variables and their values) could support or impair learning. That is, the same concrete feature could elicit learner conceptions that are either relevant or irrelevant—or even contradictory—to the underlying task. Concrete features may impair learning by eliciting student conceptions that are irrelevant to the task (M1), including engineering goals and beliefs about the domain-specific variables and values. Conversely, these same concrete features may support learning by helping students understand why controlling variables is necessary to determine the causal status of focal variables (M2). Further, understanding the rationale for controlling variables has been found to predict CVS transfer (Authors, 2011). In this study, we chose concrete representations of experiments given during instruction that were just "concrete enough" to convey potentially relevant information about causality. For example, we used "rough" and "smooth" surfaces rather than the more concrete and detailed values "sandpaper" and "glass" surfaces because the more concrete values may elicit other conceptions that are irrelevant to learning to design controlled experiments and may distract students from the intended task goal.

We hypothesized that students' attentional focus—whether they were likely to focus on surface or deep features of the example experiments—would interact with condition to affect learning and transfer outcomes. We predicted that, because the initial abstract representations would support students' understanding of the goal of the lesson, students who were more prone to focusing on surface features would show the best outcomes in the Abstract-fading condition (P1). We expected that the additional concrete details provided in the final experiment of the Abstract-fading condition would then help students understand the rationale for controlling variables (M2). We also predicted that students who were *not* prone to misinterpreting the goal of instruction—or deep-focus students—would do best in the Concrete-only condition (P2) because the initial and continued presentation of concrete representations would support students' understanding of the rationale for controlling variables (M2). However, although not expected, we acknowledged the possibility that concrete representations may lead to domain-specific reasoning that does not transfer to other domains, and consequently, deep-focus students may actually benefit most from a Concrete-fading condition (alt-P2).

P1. Although we predicted that surface-focus students would benefit most from an Abstract-fading approach (P1), there was no effect of condition among surface-focus students. Although surface-focus students in the Abstract-fading condition expressed fewer surface feature-related misinterpretations initially in the instruction, they were still less likely to express an initial procedural or conceptual understanding of CVS than students given concrete representations. However, this early advantage of concrete representations disappeared by the end of instruction, when surface-focus students in the Abstract-fading and Concrete-only conditions were equally likely to express an understanding of the rationale for controlling. This lack of difference in final understanding may explain the similar transfer performance of students in the Concrete-only and Abstract-fading conditions.

P2. Consistent with P2a, deep-focus students showed significantly better transfer in the Concrete-only than Abstract-fading condition. As with the surface-focus students, the minimally-concrete representation supported students' initial CVS understanding and understanding that the contrasted variables are problematic because they may affect the experimental outcome. This initial difference in understanding persisted to the end of the instruction, where students in the Concrete-only condition gave more correct multiple-choice responses and were more likely to express an understanding of the rationale for controlling variables in their open-ended responses than students in the Abstract-fading condition. Further, differences in understanding expressed at the end of instruction accounted for the better transfer performance among deep-focus students in the Concrete-only condition.

Given the initial advantages of the Concrete-only condition for both surface- and deep-focus students, why did only deep-focus students show better transfer? We previously found that students tend to develop an understanding of CVS incrementally rather than holistically (Author, 2013), where students generally are able to articulate an understanding of the rationale for controlling only after having developed an explicit understanding of the procedure. That is, students' explicit understanding of the rationale for controlling appears to be the final step of CVS development. It is possible that—even with the initial "head-start" provided by concrete representations—surface-focus students simply did not have enough time to develop an understanding of CVS—including the rationale for controlling—during the 40-minute instructional period. (Surface-focus students in the Concrete-only and Abstract-fading conditions did not differ in rate of rationale expression on the final instructional question; however, deep-focus students in the Concrete-only condition were significantly more likely to do so.) Had instructional time been extended, there may have been benefits of concrete representations among surface-focus students as well. However, as discussed next, the nature of the concrete features may have prevented surface-focus students from developing a complete CVS understanding.

Findings of advantages of the concrete representations for transfer are generally consistent with Authors (2014). In Authors (2014), sixth- through eighth-grade students who learned rules analogous to those of modular arithmetic (modulo-3) with minimally-concrete figures that conveyed additional information relevant to the modular arithmetic topic performed better on the transfer task (modulo-3 arithmetic problems) than students given more abstract versions of the training figures. However, the benefit of concrete representations was greater among students with poorer deductive reasoning skills (or lower-ability students). In the current study, only deep-focus students, who are likely higher-ability than surface-focus students—benefited from concrete representations. Differences in the relevance of the concrete features to the transfer tasks may explain why different ability students benefited from the concrete features across studies. That is, in Authors (2014), there was no evidence that the concrete features elicited alternate goals in the modulo-arithmetic task, whereas surface features of experimental representations in the present study elicited engineering goals and variable-effects misinterpretations, as shown in responses to the initial open-ended questions.

Consistent with P2b, deep-focus students in the Concrete-only condition demonstrated better transfer performance than those in the Concrete-fading condition, as did sixth-graders. Analyses of students' final instructional responses indicated that both deep- and surface-focus students in the Concrete-only condition developed better understanding of the rationale for controlling variables by the end of the instruction than students in the Concrete-fading condition. Not only did the deep-focus students in the Concrete-only condition generally demonstrate a better final understanding than Concrete-fading students, they were better able to apply this understanding on the posttest. This same pattern of better final understanding and better transfer of that understanding to the posttest was found among the sixth-graders.

The results of both between-condition (Concrete-only vs. Concrete-fading) and within-condition analyses (in which the concreteness of students' responses to the final instructional experiment was unrelated to transfer performance) are inconsistent with concreteness leading to a type of reasoning that impaired transfer performance. If anything, the concreteness of the final experiment appears to have *supported* transfer among the deep-focus and younger (sixth-grade) students.

Thus, the current results suggest more supportive effects of concrete features than Gick and Holyoak's (1983) finding that college undergraduates were equally likely to apply context-specific and abstract versions of a principle in a transfer domain with different surface features. We believe that a likely explanation for this difference may have to do with the relative coherence of learners' conceptions derived from the concrete and abstract representations in the current vis-à-vis Gick and Holyoak (1983) studies. That is, in the current study, concrete representations helped students understand the reason for applying the given procedure. However, in the Gick and Holyoak (1983) study, information regarding the rationale for the procedure (i.e., “attacking” a target in a distributed manner to prevent negative effects of a single direction of attack) was available in both representations.

These results also appear to contradict research findings that abstract training domains may better facilitate transfer (e.g., Bassok & Holyoak, 1989; Goldstone & Son, 2005). For example, in Goldstone and Son (2005), an “idealized” representation led to better transfer performance than the “concrete” representation. However, when looking beyond the “concrete” labels, similarities emerge. The idealized version of the training simulation—although visually abstract—was labeled “concretely” (i.e., the components were still referred to as “ants” and “food”). Their relatively more abstract representation is more analogous to the concrete representation of the current study, which did not include any pictorial representations but components were described more concretely (e.g., as “smooth/rough”). In addition, like the (minimally) concrete representations of the present study to deep-focus students, the idealized representations of the Goldstone and Son (2005) study likely provided primarily relevant information that supported transfer. Thus, in both the Goldstone and Son and present studies, representations that included additional—and only—relevant information better supported transfer. However, as the Bassok and Holyoak (1989) results suggest, relevant concrete features may impede transfer if they are integrated within more complex knowledge networks.

Age-related differences. Another possible factor involved in the effect of concreteness on transfer is learner age. We found an unexpected treatment by age interaction, where the transfer performance of sixth-graders in the Concrete-fading condition was worse than in the Concrete-only condition, but transfer performance of seventh-graders was similar. As previously discussed, this was at least partially due to the inability of sixth-graders in the Concrete-fading condition to transfer their final understanding to the posttest. However, several prior studies have found advantages of concrete-fading approaches over all concrete representations on transfer performance (e.g., McNeil & Fyfe, 2012; Goldstone & Son, 2005). Unlike the current study, which was run with younger participants (sixth- and seventh-graders), these studies were run with adult participants. This raises the possibility that the ability to transfer from an abstract representation to a specific domain—or the benefit of a Concrete-fading condition—may be age-related. However, further research is needed to investigate this relationship.

Another possible explanation for why the transfer performance of sixth-graders was better in the Concrete-only than Concrete-fading condition whereas the transfer performance of seventh-graders was similar is that the sixth-graders required the additional support provided by concrete representations throughout instruction to develop an understanding robust enough to transfer to the posttest. This latter possibility is consistent with Author (2014), where both lower-reasoning ability and younger (sixth-grade vs. eighth-grade) students required the additional support of concrete-relevant features to show gains in transfer performance.

Study limitations. Benefits of concrete representations found in the current study may be related to the nature of the topic (CVS). That is, prior research indicates that humans have a tendency to believe that variables are causal rather than non-causal (e.g., Schauble et al., 1991; Gilovich, Vallone, & Tversky, 1985; Xu & Harvey, 2014). This tendency may actually support cross-domain transfer. For example, believing that a particular non-focal variable has an effect on the outcome may remind students of the need to control it. Thus, learning experimental design may be a topic in which learner biases and surface-level relevant concrete features act in concert to support transfer. Furthermore, the specific instructional emphasis on the rationale for controlling variables may have been particularly suitable for concrete representations. That is, the additional causal information provided in the concrete representations may have been particularly helpful in supporting understanding of the rationale for controlling. Outcomes of the present study may have differed if instruction had instead focused on only procedural aspects of CVS.

In addition, as discussed previously, in order to convey minimal additional information beyond that of potential causality, the “concrete” experiments shown during instruction tended to be “minimally” concrete. For example, they did not include iconic representations of the ramps experiments and physical properties were represented at a relatively abstract level (e.g., “rough”/“smooth” were used as variable values for type of surface rather than “sandpaper”/“glass” surface). Thus, the representations used in the current study did not fall within the “concrete” end of the concreteness continuum (see Figure 3). It is possible—and we think, likely—that using representations that are more concrete and perceptually richer would result in poorer learning outcomes. However, this is a question for future research. More generally, further research is needed to continue identifying the boundary conditions under which using more or less abstract representations is better for promoting learning and transfer.

References

Authors (2010, 2011, 2012, 2013, 2014).

Bassok, M., & Holyoak, K. J. (1989). Interdomain transfer between isomorphic topics in algebra and physics. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*(1), 153-166. doi: 10.1037/0278-7393.15.1.153.

Cheng, P. W., & Holyoak, K. J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology*, *17*, 391-416. http://dx.doi.org/10.1016/0010-0285(85)90014-3.

Cheng, P. W., & Holyoak, K. J. (1989). On the natural selection of reasoning theories. *Cognition, 33*, 285-313. doi: 10.1016/0010-0277(89)90031-0.

De Bock, D., Deprez, J., Van Dooren, W., Roelens, M., & Verschaffel, L. (2011). Abstract or concrete examples in learning mathematics? A replication and elaboration of Kaminski, Sloutsky, and Heckler's study. *Journal for Research in Mathematics Education*, *42*(2), 109-126. http://www.jstor.org/stable/10.5951/jresematheduc.42.2.0109.

DeLoache, (1991). Symbolic functioning in very young children: Understanding of pictures and models. *Child Development, 62*, 736–752. doi: 10.1111/j.1467-8624.1991.tb01566.x.

Fong, G. T. & Nisbett, R. E. (1991). Immediate and delayed transfer of training effects in statistical reasoning. *Journal of Experimental Psychology: General, 120*(1), 34-45.

Gentner, D. (2010). Bootstrapping the mind: Analogical processes and symbol systems. *Cognitive Science*, *34*(5), 752-775. doi: 10.1111/j.1551-6709.2010.01114.x.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science, 7*, 155–170. doi: 10.1207/s15516709cog0702_3.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive psychology, 15*(1), 1-38.

Gilovich, T., Vallone, R., & Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive psychology, 17*(3), 295-314.

Goldstone, R. L., & Sakamoto, Y. (2003). The transfer of abstract principles governing complex adaptive systems. *Cognitive Psychology, 46*(4), 414-466. http://dx.doi.org/10.1016/S0010-0285(02)00519-4.

Goldstone, R. & Son, J. (2005). The transfer of scientific principles using concrete and idealized simulations. *The Journal of the Learning Sciences, 14*(1), 69-110. doi: 10.1207/s15327809jls1401_4.

Holyoak, K. J., & Koh, K. (1987). Surface and structural similarity in analogical transfer. *Memory & Cognition, 15*(4), 332-340.

Kaminski, J. A., Sloutsky, V. M. & Heckler, A. F. (2006). Do children need concrete instantiations to learn an abstract concept? *Proceedings of the XXVIII Annual Conference of the Cognitive Science Society*, 1167-1172. Mahwah, NJ: Erlbaum.

Kaminski, J. A., Sloutsky, V. M. & Heckler, A. F. (2008). The advantage of abstract examples in learning math. *Science, 25*, 454-455.

Kaminski, J. A., Sloutsky, V. M., & Heckler, A. F. (2013). The cost of concreteness: The effect of nonessential information on analogical transfer. Journal of Experimental Psychology: Applied, 19(1), 14. doi: 10.1037/a0031931.

Koedinger, K. R., Alibali, M. W., & Nathan, M. M. (2008). Trade-offs between grounded and abstract representations: Evidence from algebra problem solving. *Cognitive Science, 32(2),* 366-397. doi: 10.1080/03640210701863933.

McNeil, N.M., & Fyfe, E.R. (2012). “Concreteness fading” promotes transfer of mathematical knowledge. *Learning and Instruction*, *22*, 440-448. http://dx.doi.org/10.1016/j.learninstruc.2012.05.001

McNeil, N., Uttal, D., Jarvin, L. & Sternberg, R. (2009). Should you show me the money? Concrete objects both hurt and help performance on mathematics problems. *Learning & Instruction, 19*, 171-184. http://dx.doi.org/10.1016/j.learninstruc.2012.05.001

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92*(3), 289.

Patalano, A. L., Chin-Parker, S., & Ross, B. H. (2006). The importance of being coherent: Category coherence, cross-classification, and reasoning. *Journal of Memory and Language, 54*(3), 407-424.

Sao Pedro, M., Gobert, J., & Raziuddin, J. (2010). Long-term Benefits of Direct Instruction with Reification for Learning the Control of Variables Strategy. In V. Aleven, J. Kay, & J. Mostow (Ed.), *Proceedings of the Tenth International Conference on Intelligent Tutoring Systems, ITS 2010*. LNCS 6095, pp. 257-259. Pittsburgh, PA: Springer Berlin/Heidelberg.

Sao Pedro, M., Gobert, J., & Sebuwufu, P. (2011). The effects of quality self-explanations on robust understanding of the Control of Variables Strategy. Paper presented at The Annual Meeting of the American Educational Research Association. New Orleans, LA.

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*, 859–882.

Siegler, R. S. (2002). Microgenetic studies of self-explanations. In N. Granott & J. Parziale (Eds.), *Microdevelopment: Transition processes in development and learning*. 31-58.

Xu, J., & Harvey, N. (2014). Carry on winning: The gamblers' fallacy creates hot hand effects in online gambling. *Cognition, 131*(2), 173-180.