

Alternative Beliefs about Experimental Design

Stephanie Siler (siler@cmu.edu), David Klahr, Kevin Willows, and Cressida Magaro

Carnegie Mellon University

It has long been known that students' preconceptions about the physical world influence their ability to learn. To the extent that a preconception is a misconception – i.e., at variance with the knowledge to be acquired – it is important to identify and remediate. There have been two relatively distinct foci in the research on early science learners' preconceptions: one of which is a focus on domain knowledge and the second on the domain-general scientific procedures that produce that knowledge. In this paper, we focus on a domain-general procedure, specifically, children's naïve or alternative understandings that emerge in the context of experimental design. Building upon previous research (e.g., Kuhn et al., 1995; Schauble et al., 1991), we provide a taxonomy of the common types of preconceptions we identified in our corpus of instructional materials that students completed and in tutoring transcripts. Afterwards, we discuss the results of an instructional intervention designed to reduce the elicitation of alternative goals and thus increase student learning of experimental design.

Task analysis of experimental design.

To design an informative experiment that will allow one to make valid inferences about the causal nature of a variable (i.e., an experiment following the Control of Variables Strategy, “CVS”), one first needs to understand that the goal of this type of experimentation is to *find out about* a particular variable or hold a “science” goal.

One must then identify the variable to find out about (i.e., the “target” variable), make the levels of that variable different across at least two conditions to compare their relative effects, and control all other variables. Having run the experiment, one can use the results of the two experimental conditions to infer whether there is any effect of the target variable (and the direction of that effect). Thus, when one holds a science goal, inferences are made about the effect of the target variable only *after* the experiment is run.

Though this task analysis may suggest otherwise, learning the scientific approach to experimental design is difficult for many students.

Procedure: Phase 1

In Phase 1 of our project, we worked with 6th-grade students and teachers at three different local K-8 private Catholic schools. Two of these schools (classrooms L1 & L2) served low-SES urban and predominantly African-American students, and one (with classroom M1) served suburban middle-SES and primarily Caucasian students. Classrooms L1 and L2 differed somewhat in their science instruction; in L1, science instruction tended to be more “traditional”, with primarily teacher-led question/answer interactions, but, due to limited resources, no in-class student experimentation or inquiry. In L2, the science curriculum tended to be more diverse, with both teacher-led discussions and student experimentation and inquiry. Students in L2 were also required to complete a yearly science project, which they presented at an end-of-year in-school fair.

All teachers first received researcher-provided training on the CVS instruction (which was very similar to the one-to-one instruction given to students in Chen and Klahr, 1999, and described shortly). It focused on the rules and rationales for setting up informative (i.e., unconfounded) experiments. First, all students completed a “Story-evaluation” pretest in which they evaluated six experiments in three different domains (selling drinks, rockets, and cookies), and corrected any experiments they evaluated as “Bad” (**Figure 2, below**). All variables had two types or levels (e.g., for drinks, students could choose either lemonade or iced tea).

The next day, teachers presented the ramps apparatuses and students designed an experiment to test each of the four ramps variables (slope, run length, surface, and ball type). Then, teachers administered that CVS instruction in their classroom. Specifically, the teacher presented a confounded experiment using two ramps and asked students to individually evaluate the experiment as a fair or unfair way to find out about the target variable on their worksheets. Afterwards, the teachers led a class-wide discussion about whether or not the experiment under discussion was informative (i.e., enabled one to make inferences about effects of the target variable) and why students believed that to be the case. As students identified the confounded

variables, the teacher controlled them on the ramps apparatus until eventually the experiment was unconfounded (i.e., informative for the target variable). Then students wrote explanations for why the corrected experiment was now a fair way to find out about the [target variable]. This sequence was repeated two more times. On the final day of instruction, students completed the ramps post-test in which they again designed an experiment to test each of the four ramps variables. Students who did not show CVS mastery by designing at least three out of four unconfounded (or CVS) ramps experiments were tutored by a member of our research team. Finally, about three weeks later, students completed the Story post-test (identical to the Story pre-test).

Results: Phase 1

In M1, the middle-SES classroom, the rate of CVS ramps mastery¹ (80%) was similar to what was found in prior studies of one-to-one instruction of middle-to-high-SES students (e.g., Strand-Cary & Klahr, in press). However, in the other two classrooms, the rate of CVS learning was much lower (only about one-third of the students in both classes gained CVS mastery through the classroom instruction). Similarly, the mean percentage of CVS designs in M1 was 73%, but in L1 and L2 were 37% and 48%, respectively.

Students who did not gain CVS mastery from classroom instruction were tutored one-to-one. Our initial approach was to repeat the classroom instruction to rule out the possibility that students had simply not paid attention to it. As expected, this was not successful; thus, we sought to identify preconceptions that might have been affecting students' ability to learn.

Engineering Goals: We found that some of our Phase 1 students held what Schauble et al. (1991) termed “engineering” goals. They seemed to be using their knowledge of the effects of relevant variables to “engineer” a particular desirable result or outcome rather than attempting to *find out about* those effects. According to Schauble et al. (1991), engineering goals arise because “People are accustomed to making simple changes in the environment to generate desired outcomes; this goal seldom requires analytic or extended reasoning” (p. 862). This behavior is reinforced in science classes, where the focus is on the

¹ In these analyses, students who designed more than 2 out of 4 CVS ramps experiments were excluded.

production of an effect rather than on understanding its cause(s). The following is a synopsis of our findings of different types of alternative goals we found in Phase 1 and in later phases of our project, from students written and oral explanations.

1) One type of desirable outcome, identified by Schauble et al. is “maximizing the effect”. Our students who held this goal tried, for example, to design a ramp that would make the ball roll farthest. We found this goal to take two forms:

A. In one version of this goal, students designed *both* condition to produce maximum effects. We saw two versions of this. (a) One way students accomplished this was by designing each condition identically, using the optimal level of each variable in each condition, resulting in a non-contrastive comparison. In this form of maximizing goal, students are not *comparing* the two conditions, but rather considering the conditions separately². In the excerpt below, the tutor tried to prompt the student to vary the target variable (slope), where the set-up is non-contrastive and both slopes are steep (the optimal setting). The bolded student statements are evidence of his engineering goal, which is consistent with his subsequent expressed desire to keep both slopes steep.

T: So we want to know if a high slope or a low slope would cause the ball to roll farther. So in order to do that, what do we need to do with the slope?

J: **if we wanted to make a *better* slope,**

T: what would we have to do with the slope? How would we set up the slopes if you want to see if a high slope or a low slope...

J: well, I would make a slope steep as it is,

T: uh-huh.

J: **I would make it a smooth surface** [change from a rough surface]

T: OK, what do we do with the other slope?

J: the other slope?

² It was not uncommon for kids to think that one of the conditions in an experiment *was* the experiment.

T: mm-hmm.

J: (pause) isn't that already steep?

(b) Sometimes students had the goal of maximizing outcomes in both conditions, but, because their domain theories were in terms of interactions between variables, they designed maximally-contrastive set-ups (i.e., ones where all of the variable settings differ across conditions). For example, one student designed a maximally-contrastive “experiment”³ (Condition A: noon, older child, and iced tea; Condition B: 3pm, younger, lemonade), giving the following explanation: “Most younger kids are outside around 3pm and older kids like to get up in the noon. Also more younger kids like lemonade than iced tea.”

B. In a second version of this maximizing outcome goal, students simply ignored one of the conditions in their experiment. For example, one student evaluated a maximally-contrastive experiment as “good”, explaining: “Because the younger children can sell more. The older children and adults will give them more money.” This student, who held a maximizing engineering goal, was evaluating only the condition in which a younger child was selling drinks⁴. Or, students would set up an experiment with two conditions, but only refer to one when explaining their rationale for the set-up. For example, one student set up a maximally contrastive drinks “experiment” (Stand A: Noon, Younger child, Lemonade), and explained “I thought that noon would be a better time for a younger child to be out and younger children drink more lemonade than an older child would.” This student, typical of those with this approach, did not mention Stand B at all. Therefore, it is unclear why the student set it up opposite of Stand A.

2) Another type of maximizing goal (reported by Kuhn & Dean, 2005) involves maximizing the *difference* in outcomes *between* the two conditions. This might lead to, for example, a comparison between a ramp that is high, smooth, and long and a ramp that is low, rough, and short. We consider this to be the more sophisticated form because it is “closer” to the prescriptive strategy that requires contrasting the outcomes

³ The intended goal of the experiment was to design an experiment that would allow one to tell if “time of day” has an effect on the number of drinks that are sold.

⁴ Note that without having access to this student’s rationale, he may have been thought to hold a maximally-contrastive (comparative) goal (discussed next).

of the two conditions. One underlying reason for this type of goal is to produce a result that is consistent with the student's expectation about the effect of the variable (Tschirgi, 1980), to "prove" that the variable has an effect. For example, if the student is asked to set up an experiment to show whether the slope of a ramp affects how far a ball will roll down it (which the student believes to be true), then the student may set up a comparison between a steep, smooth, long ramp and a low, rough, short ramp. Not only is the target variable contrasted, but the other "best variable settings" are paired with the "best target variable setting" to ensure the result. In this particular underlying reason, the effort of using one's beliefs about variable effects to "prove" the effect of the target variable makes this an engineering goal. However, it differs from other engineering goals in which variable effect beliefs are applied to produce some practical benefit (e.g., a maximum outcome). Though students frequently designed maximally-contrastive experiments, we did not find any direct evidence that any of our students held this particular maximizing differences goal in their spoken or written explanations.

We did, however, see indications that students held the goal of getting different (though not necessarily *maximally* different) outcomes. One example of a statement expressing this goal is from a student who set up a maximally contrastive experiment: "because they will both fly a different way". It is possible that such students were actually trying to produce maximally different outcomes to "prove" that a particular variable made a difference, but just did not (or could not) express this.

3) Another type of engineering goal, which has not been previously identified in the literature was to produce the *same* outcomes in both conditions. We believe that this particular goal emerged from use of the word "fair" in the classroom instruction (e.g., "Is this a *fair* test of [target variable]?"). Students applied their "fair/same outcome" goal in several ways. (a) The simplest way was by designing the two conditions identically, for example, by comparing a high ramp with a smooth surface to a high ramp with a smooth surface. For this goal, beliefs about the effects of the variables are not even necessary. The following student gives an explanation for why his non-contrastive (NC) experiment is a fair test of slope: "...I think it's fair because the surface is both the same way, the height is the same, they're both long."

(b) We also found that students designed set-ups in which the variables they believed “mattered” were held constant, whereas variables they believed “didn’t matter” were varied across conditions. For example, they made the heights and surfaces the same, but the color of the ball (which students generally say does not matter) different.

(c) Alternatively, students designed experiments with the goal of producing the same outcome by “balancing” the effects of variables across conditions, for example, by making one ramp high (better outcome) but with a rough surface (worse outcome) and the second ramp low (worse outcome) but with a smooth surface (better outcome). Because the two variables could “balance each other out” across the ramps, the outcome (e.g., how far a ball will roll down the ramp) can conceivably be the same. To illustrate, after being asked why he varied the length of the ramp (the target variable), the student said: “Because the rough one [inaudible] probably mess the ball up so I made it longer so it could move faster.”

This type of “hybrid” approach may arise when students learn to vary the target variable, but still hold the engineering goal of producing the same outcome. This is akin to Vosniadou and Brewer’s (1992) students’ hybrid models of the earth (round, but with a flat interior, representing the earth’s surface), which incorporate both students’ preconceptions (that the earth is flat) and new information they’ve acquired through instruction (that the earth is round).

“Variable effect” responses: We found that students sometimes did not seem to have *any* particular goal when designing or evaluating experiments (one type of what Kuhn et al., 1995, refer to as “theory-motivated processing”). Rather, they simply stated or applied their beliefs about the general effects of the variables. In the following example, both students (but S1 to a greater extent) misinterpreted the tutor’s questions as asking about the effects of the variable (surface) on an outcome rather than as asking about the relationship between the set-up and possible outcome causes. Again, the student does not consider the logic of the experimental set-up or whether it allows for making a valid inference about the target variable:

T: OK, is there something else that’s different about how we have these ramps
set up that might cause the ball to roll different distances?

S1: **the surface**⁵? (but both surfaces are smooth)

T: OK, why do you say the surface?

S1: because if you have a rough surface, it won't go so smoothly down. It won't go as far.

T: okay, but in this case do we have a rough surface?

S2: no.

T: no, ok, so we have both smooth surfaces. OK, so should, would the surfaces cause any differences in how far the balls rolled?

S1: **yes**⁶.

[S2 nods]

T: OK, just in this comparison, since we have both smooth surfaces, would that cause any differences?

S1: [no]

Another student evaluated a maximally-contrastive experiment as “Bad” and explained: “... it really should not matter how old the child is. The children are both selling drinks and it should not matter if the older person was 14 and the younger child was 10.” This student was not evaluating the validity of the experimental set-up, but interpreted the question as: “Does age make a difference?”

Another example of a student stating a variable-effects belief to predict an outcome occurred when the tutor was actually asking her about the logic of the link between the experimental design and outcome (i.e., whether a non-contrastive target variable would allow her to know whether or not the target variable made a difference) follows:

T: Now they're [the balls] both at the top. If we ran this, could you tell whether there's a difference caused by the starting position?

⁵ This student seems to be answering “What might cause the ball to roll different distances?”

⁶ This student persists in answering “Does surface make a difference?”

S: It doesn't gain that much speed.

Table 1 below shows the frequencies of the various engineering goal types and variable effects statements for all Phase 1 classrooms, and each of the classrooms individually. In L1, 45% of the students (9 out of 20) produced an engineering or variable-effect explanation on the first written experimental evaluation question during classroom instruction, compared to 8% (only 1 out of 12) in L2 and 17% of students (3 out of 18) in M1. This difference was significant, $\chi^2(2, 50) = 6.51, p = .04$. Comparing individual schools, L1 was significantly higher than L2, $p(\text{Fisher}) = .03$, and L1 was marginally higher than M1, $p(\text{Fisher}) = .06$. M1 and L2 did not differ. Thus, a greater percentage of students who were less accustomed to experimenting in their science classes (i.e., in L1) interpreted the CVS instruction as either an engineering task or as about the effects of variables. Thus, these results are consistent with the hypothesis that students' experiences in the classroom influence how they interpret instruction on experimental design, though other factors are likely at play as well (e.g., reading ability, discussed next).

Table 1. Percentage of approach within engineering goal and variable effects (and overall percentage for all explanations given) by phase and classroom.

Approach	Phase 1				Phase 4
	All	M1	L1	L2	M2
1a. Maximize outcomes (independent effects)	7.7 ^a (1.6)	0	11.1 ^b (4.5)	0	39.5 (20.4)
1b. Maximize outcomes via Interactions		0		0	19.7 (10.2)
1c. Maximize outcome one condition	15.4 (3.2)	0	22.2 (9.1)	0	22.4 (11.6)
Maximize outcome (unknown) ^b	0	0	0	0	2.6 (1.4)
Total max outcome	23.1 (4.8)	0	33.3 (13.6)	0	84.2 (43.5)
2a. Different outcomes	15.4 (3.2)	33.3 (5.6)	0	100 (8.3)	7.9 (4.1)
2b. Maximize outcome differences	0	0	0	0	0
Total 2	15.4 (3.2)	33.3 (5.6)	0	100	7.9 (4.1)
3a. Same outcomes via NC set-ups	30.8 (6.3)	33.3 (5.6)	33.3 (13.6)	0	0

3b. Same outcomes vary non-causal	15.4 (3.2)	0	22.2 (9.1)	0	0
3c. Same outcomes via “balanced” set-ups	0	0	0	0	0
Total Same outcomes	46.2	33.3 (5.6)	55.5 (22.7)	0	0
Variable effects	15.4 (3.2)	33.3 (5.6)	11.1 (4.5)	0	7.9 (4.1)

^a Students referred to maximum outcomes in both conditions but not individual variables.

^b Ambiguous but maximize outcome responses.