PSYCHOLOGY OF SCIENCE IMPLICIT AND EXPLICIT PROCESSES

2ND PURDUE SYMPOSIUM ON PSYCHOLOGICAL SCIENCES

Edited by

Robert W. Proctor

and

E. J. Capaldi

Vosniadou, S., & Brewer, W. F. (1992). Mental models of the earth: A study of conceptual change in childhood. *Cognitive Psychology, 24,* 535–585.

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology, 12,* 129–140.

Wetherick, N. E. (1962). Eliminative and enumerative behaviour in a conceptual task. *Quarterly Journal of Experimental Psychology, 14,* 246–249.

Worth, R. H., Jr., (1995). No choice but war: The United States embargo against Japan and the eruption of war in the pacific. Jefferson, North Carolina: McFarland.

Zimmerman, C. (2000). The development of scientific reasoning skills. *Developmental Review, 20,* 99–149.

Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review, 27,* 172–223.

7 Detecting, Classifying, and Remediating

Children's Explicit and Implicit Misconceptions about Experimental Design

Stephanie A. Siler and David Klahr

It is well established that children do not come to their science classes as "blank slates" but rather, as learners with preconceptions—sometimes deeply entrenched—about the natural world and ways to explore that world in order to learn more about it (Stathopoulou & Vosniadou, 2007). Children's experiences may also lead to biases—both implicit and explicit—that only come to be revealed in the context of instruction. To the extent that preconceptions about both scientific knowledge and scientific processes are *mis*conceptions, they are important to identify and remediate (Vosniadou, Ioannides, Dimitrakopoulou, & Papademetriou, 2001). Once an implicit misconception has been identified by appropriate assessments, it may be possible to craft instruction that not only renders the implicit misconception explicit, but also directly engages, builds upon, and/or remediates that misconception.

Research on early science learning has focused on two relatively distinct classes of misconceptions. One line of work has focused on *misconceptions in specific domains,* such as heat and temperature (e.g., Slotta, Chi, & Joram, 1995; Wiser & Carey, 1983), the solar system (e.g., Vosniadou, 1994; Vosniadou & Brewer, 1992, 1994), mass and density (e.g., Smith, Carey, & Wiser, 1985), or forces and motion (e.g., Clement, 1983; diSessa, 1993; McCloskey, 1983; Minstrell, 1982). A second line of work has examined children's misconceptions in the abstract and *domain-general scientific procedures* that produce and verify that domain knowledge, such as

2003; Morris & Sloutsky, 2002; Pieraut-Le Bonniec, 1980), interpreting anomalous data (e.g., Chinn & Brewer, 1993), understanding the uses of scientific models (e.g., Grosslight, Unger, Jay, & Smith, 1991), and constructing coherent and consistent arguments (Kuhn, 1991; Zohar & Nemet, 2001). In this chapter, we focus on these latter types of difficulties: specifically, the errors and misconceptions that students bring to the task of learning how to design and interpret simple experiments.

Our analysis is motivated by, and based on, a corpus of instructional dialogues produced during a series of extensive tutorial sessions with elementary and middle school students engaged in learning about experimental design concepts and procedures. These tutorial dialogues were generated as part of our ongoing project aimed at building an adaptive computer tutor to teach the basic procedural and conceptual knowledge associated with experimental design (e.g., Klahr, Triona, Strand-Cary, & Siler, 2008; Siler, Klahr, Strand-Cary, Magaro, & Willows, 2009). Because diagnosing students' knowledge and beliefs is an essential prerequisite for instruction in an adaptive tutor, a major goal of the project has been to identify children's conceptual and procedural misconceptions about experimentation. In pursuing this goal, we have drawn on earlier work on this topic by others (e.g., Kuhn, Garcia-Mila, Zohar, & Andersen, 1995; Schauble, Klopfer, & Raghavan, 1991), as well as past work done in our laboratory (e.g., Chen & Klahr, 1999; Klahr & Nigam, 2004; Strand-Cary & Klahr, 2008; Toth, Klahr, & Chen, 2000). We start by outlining an expert model of basic experimental design. Then we describe the phases of our computer development project from which we draw our data, and finally, we describe the types of misconceptions about experimental design expressed by the fifth- through seventh-grade students who participated in the various project phases. We comment on the nature of these misconceptions, including their level of explicitness and their relationship to students' domain-specific beliefs.

Expert Model of Experimental Design

Schauble et al. (1991) were the first to observe the important distinction between two classes of goals that children bring to their initial experimental activities—"science" goals, and "engineering" goals. In order to design an experiment that will support valid causal inferences about a particular variable, one needs to understand the "science" goal underlying experimentation, which is to *find out about* the causal status of that variable. For example, when a child justifies an experimental design by saying: "I wanted *to see if* the different engines would affect the speed of the rockets," she is explicitly expressing the science goal of trying to find out about the causal relation between engine type and rocket speed. A science goal may be considered abstract in two senses. First, it may be applied in any specific context, in or out of school. Secondly, because a science goal involves finding out about one or more variables, it is abstract in the sense that students must be willing to suspend any beliefs about the effect(s) of the variable(s) they are testing in order to design and a particular desirable outcome (e.g., make the fastest rocket) rather than attempt to find out about those effects via experimentation, as with a science goal. As with science goals, engineering goals are abstract in the sense that students can and do apply them across contexts. However, because applying an engineering goal may involve one's beliefs about some domain-specific variables to produce a domain-specific result, an engineering goal may also be concrete.

In addition to holding a science goal, one needs to know, and correctly execute, the abstract procedural steps involved in designing an unconfounded experiment. We call this procedure the Control of Variables Strategy (CVS) and its abstract rules are shown in Table 7.1.

To apply the basic procedure, one must identify the particular variable whose causal status is being investigated, the "target" variable (Rule 1), contrast the levels of the variable being tested across *at least two* conditions to compare their relative effects on the outcome (Rule 2), and control all other variables (Rule 3)[1]. Only after designing such an experiment can one compare the outcomes of the contrasting conditions (i.e., A1 and A2) to infer whether there was an effect of the target variable.

This procedure is abstract in two ways. First, like engineering and science goals, CVS is domain-general and can in principle be applied in any context. Second, the procedure is not dependent on the specific variables and values in the context in which CVS is applied. In addition to this procedural knowledge, we believe it is important for students to also understand the underlying rationale of CVS—that if one runs the experiment and finds different results, then one can logically attribute that result to the one variable that is contrasted. This logical understanding is also not dependent on the specific variables involved.

Overview of Identified Difficulties

Although the core procedure for CVS can be captured in the few simple rules listed in Table 7.1, their simplicity makes them no easier to master and apply in a wide

Table 7.1 Abstract representation of rules for designing, running, and evaluating an unconfounded experiment

In a multivariable situation, if your goal is to determine whether or not a variable plays a causal role in outcome A, then

Rule 1: Identify that variable (X) and its values:

Rule 2: Create a contrast:

a. In Condition 1, Set X to Value 1.

b. In Condition 2, Set X to Value 2.

Rule 3: Set all other variables (Y, Z, W) to the same values in both conditions.

"Run" the experiment: measure A1 and A2.

If A1≠ A2, then X is causal.

range of contexts than does the simplicity of Newton's Second Law make it easy to master, recognize when it is applicable, and apply correctly to physics problems. Kuhn et al.'s (1995) classic study demonstrated that—in a variety of scientific discovery tasks in which participants explored the effects of several variables—even after 20 sessions spread over 10 weeks, fewer than 25% of fourth graders' inferences were valid. For this reason, the procedural—as well as the conceptual or logical—underpinnings of CVS have been the focus of the CVS instruction conducted in our laboratory (e.g., Chen & Klahr, 1999; Strand-Cary & Klahr, 2008).

Similarly, we have found that many children—especially younger children and those with weak science backgrounds—failed to learn from the simple nonadaptive or "straight-line" instruction that we used in our earlier studies with middle-class children (e.g., Klahr & Li, 2005). This instruction focused on the logical underpinnings of the CVS procedural steps.

Over the past several years we have been incrementally developing a computer tutor intended to aid students who had difficulty with the "straight-line" instruction. Our project involves human tutoring of those students who failed to learn from the given instruction at different phases of development. Those instructional moves found to be effective for different knowledge states inform tutor development. Extensive analyses of students' responses during instruction and dialogues from these individualized tutoring sessions revealed that students make characteristic mistakes that interfere with their learning and transfer of CVS.

In addition to these errors, there were three general schema-related misconceptions related to the goal of instruction found in our data:

- Engineering goals, described earlier;
- "Fairness" goals, in which students designed comparisons that were "fair," or could produce the same results. There are several different forms these first two schema-related misconceptions may take, which will be elaborated on in the results section;
- "Domain knowledge" goals, in which students misinterpret the purpose of instruction as a discussion of variable effects within the given domain (e.g., that balls roll faster on a smooth ramp than on a rough ramp), rather than as learning about a domain-general procedure for designing experiments.

Thus, competence in experimental design requires more than mastering the three simple rules in Table 7.1. An inability to apply CVS might be caused by failure to: (a) understand the goal of the task—which may be related to underlying difficulties with abstract concepts and procedures, (b) consider the variable level when testing for causality, (c) understand the need to "compare and contrast" variable values across conditions, and (d) understand that the causal status of only one variable at a time can be investigated in a single experiment.

In the following sections we first describe the sources of our data. Then we pro- or engineering) and students' expressions of beliefs about variable effects (i.e., application of domain knowledge goals). That is, when students applied the relatively more abstract science goal, were they also more likely to suspend their beliefs about variables other than the one(s) they were testing? For each approach, we consider whether or not it involves application of domain-specific beliefs. We also consider differences in the level of explicitness of different underlying goals. Of practical interest to us was how those approaches, which reveal failures of one or more of the knowledge components listed above, could inform the instructional and remedial strategies used by the tutor.

METHODS

Data Sources for Errors and Goal Misconceptions

Data were collected during several phases of an ongoing design-based research project aimed at developing a computer tutor (TED, for "Training in Experimental Design") that provides individualized, adaptive instruction on experimental design to elementary and middle school students. The starting point for the project was the instruction procedure—designed by Klahr and colleagues in studies previously reported (e.g., Chen & Klahr, 1999; Klahr & Nigam, 2004; Strand-Cary & Klahr, 2008; Toth, Klahr, & Chen, 2000)—given by humans using physical materials. This instruction involves asking students to evaluate several experiments, and, for each, further prompting them to consider whether the experiment would allow them to know if the target variable caused an outcome and explain why it would or would not. Following a series of such questions, students are then given an explanation for why the experiment could or could not allow for conclusions about the effect of the target variable.

During the course of TED's development, the instructional steps, student prompts and queries, and CVS assessment procedures have been incrementally replaced with computerized components. In the course of the project, we have worked with students in the classroom, in individual or small-group tutoring sessions, and have developed a more complete taxonomy of the types of misconceptions about the purpose and process of experimental design that students bring to the science classroom. Throughout this process, we collected student data in the form of the experimental designs they created or evaluated, their written responses to probes and queries, and from the explanations they gave during one-on-one and small-group tutorial interactions. These data—collected in four separate development phases of the tutor—provide the basis of the analyses presented in this chapter. Because the nature of instruction varied across development phases, it is important to note that the data reported across phases cannot sensibly be compared (e.g., in terms of frequencies). Rather, the purpose of reporting student responses across phases—and within phases but across schools—was to demonstrate the possible range of student responses and their distributions, given the different instructional instruction in their classrooms using physical ramps apparatuses; in Phase 2, a teacher led instruction using Flash-based virtual ramps, guided by PowerPoint instructional slides; in Phase 3, human tutors provided computer-supported instruction with virtual materials. Finally, in Phase 4, students completed "straight-line" TED computer-delivered instruction, with no instruction-related human interaction. All student interface actions were recorded in log files, which were saved on a server and later analyzed.

Phase 1 included 73 sixth-grade students and three teachers at three different local K–8 Catholic schools. Two of these schools (classrooms L1 & L2) served primarily low-SES students (95% and 59% of students were eligible for free or reduced lunch, respectively), and one (with classroom M1) served primarily middle-SES students (about 20% were eligible for free or reduced lunch) and incorporated an inquiry-based science curriculum. Teachers were trained on CVS instruction similar to the one-to-one instruction given to students in Chen and Klahr (1999), which focused on the rules and rationales for setting up informative (i.e., unconfounded) experiments.

On the first day, all students completed a "Story-evaluation" pretest that required them to evaluate six experiments in three different domains (selling drinks, flying rockets, and baking cookies; an example item from each domain is shown in Figure 7.1), and to correct any experiments they evaluated as "bad." All variables had two possible values (e.g., for drinks, students could choose either lemonade or iced tea). As with assessments used in past studies (e.g., Chen & Klahr, 1999; Klahr & Nigam, 2004), students were not required to give explanations for their responses. Students completed this test in approximately 15 minutes.

The next day, the teacher introduced a ramps apparatus (Figure 7.2) and demonstrated the four variables, or ways it could be changed (slope—steep or not steep; starting position of the ball—at the top or middle; surface—smooth or rough; and ball type—yellow or pink). Students individually designed an experiment to test each of these four variables by entering values for all variables for each ramp into a table on the paper-pencil ramps pretest on their worksheets. Then the teacher led the lesson on CVS by first presenting a confounded experiment using two ramps and asking students to individually evaluate the experiment as a "fair" or "unfair" way to find out about the target variable on their worksheets. Following this evaluation, the teacher led a class-wide discussion on whether or not the experiment under discussion was informative (i.e., whether it enabled one to make inferences about effects of the target variable) and explained why, in fact, the initial (confounded) setup was not informative. As students identified the confounded variables, the teacher controlled them until eventually the experiment was unconfounded (i.e., informative for the target variable). Then students wrote explanations for why the corrected experiment was a fair way to find out about the target variable. This sequence was repeated two more times with other ramps setups and target variables. On the final day of instruction, students completed the ramps posttest in which they again designed experiments to test each of the four

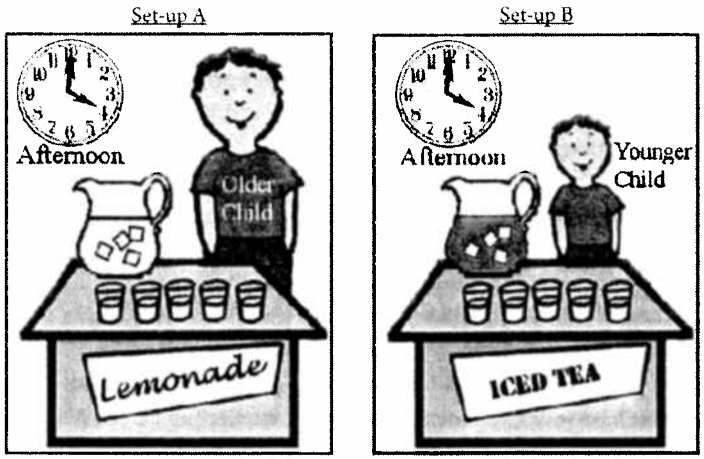
Phases of TED project and data sources

	Phase 1	Phase 2	Phase 3	Phase 4
classrooms	3 (2 low-SES: L1 & L2; 1 middle-SES: M1)	1 (low-SES) (L3)	2 (middle-SES) (M2a1 & M2a2)	4 (2 middle-SES: M2b1 & M2b2; 2 low-SES: L4 & L5)
	6	6	5 & 6	5
students	73 (L1: 23; L2: 17; M1: 23)	21	58 (27 fifth; 31 sixth)	80 (M2b1/2: 50; L4: 16; L5: 14)
truction	Classroom	Classroom	One-on-one or small group	One-on-one (human & computer)
utoring	One-on-one	One-on-one	(n/a)	(n/a)
es	• In-class written responses • Oral and written responses during remedial tutoring	• In-class written responses • Oral and written responses during remedial tutoring • Written/typed explanations on follow-up Story-eval/design test (one year later)[a]	• Oral and written responses during individualized tutoring •	• Oral and written responses during tutoring. • Written/typed explanation responses on Story-eval/design pre and posttests and follow-up[a]

lassifications from this assessment point are shown in Table 7.6.

These two pictures show how they tested whether or not the ***age of the child*** selling the drinks made a difference in how much they sell.
Look carefully at the pictures. Each one shows a time of day (Morning or Afternoon), a child (Older or Younger), and a drink (Iced Tea or Lemonade).
Do you think this is a good way to find out whether the *age of the child* (Older or Younger) makes a difference in how much they sell?

(a) **If you think it is a good way, then circle the word "Good" below. If you think it is a bad way, circle "Bad."**



Good
Bad

(b) ***If you circled "Bad," change the picture(s) above to make it a Good comparison.***
(For example, you might want to change the age of the seller, the type of drink, or the time of day in one or both of the set-ups.)

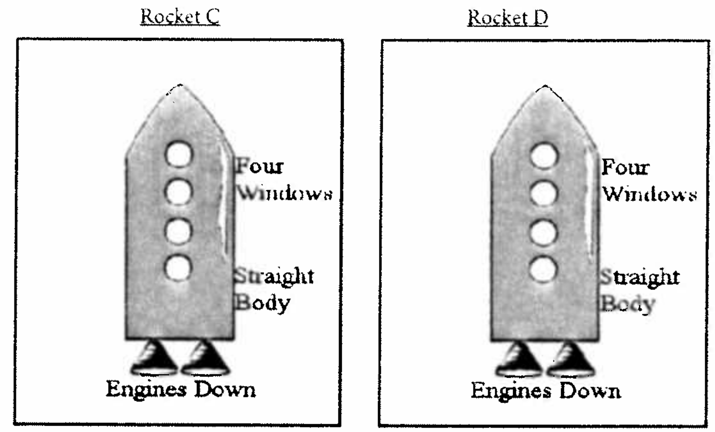
Figure 7.1a Story-evaluation questions used in pre- and post-test during Phases 1 and 2. Drink sales question. (This is a "bad" experiment because age of child is confounded with type of drink.)

Table 7.3 provides a summary of the questions discussed during the tutoring session in this and subsequent phases.

For classroom L1, only 33% of students designed at least three unconfounded experiments on the ramps posttest—much lower than in prior studies with middle-SES students (e.g., Toth et al., 2000). To rule out the possibility that students had simply not paid close attention in class, the instruction given to the full class was repeated in tutoring. However, this repeated instruction was not successful, suggesting that inattentiveness was not the problem. In fact, even after intensive one-to-one tutoring, none of the 10 tutored students showed mastery on the Story-evaluation posttest. Thus, in this and subsequent tutoring sessions in L2 and M1, as well as in all subsequent phases, we sought to identify preconcep-

These two pictures show how they tested whether or not the ***engine direction*** made a difference in how high the rockets fly.
Look carefully at the pictures. Each rocket has a certain body shape (Curved or Straight), number of windows (One or Four), and engine direction (Down or Tilted).
Do you think this is a good way to find out whether the *engine direction* (Down or Tilted) makes a difference in how far the rockets fly?

(a) **If you think it is a good way, then circle the word "Good" below. If you think it is a bad way, circle "Bad."**



Good
Bad

(b) ***If you circled "Bad," change the picture(s) above to make it a Good comparison.***
(For example, you might want to change the body shape, the number of windows, or engine direction for one or both of the rockets.)

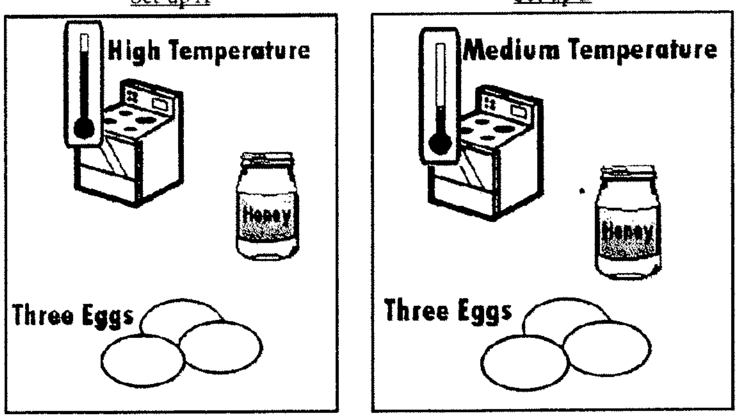
Figure 7.1b Story-evaluation questions used in pre- and post-test during Phases 1 and 2. Rocket design question. (This is a "bad" experiment because the target variable is not varied.)

by breaking it down to focus on one CVS rule (as listed in Table 7.1) at a time. Approximately two weeks after the tutoring phase, all students completed a Story-evaluation posttest, identical to the Story-evaluation pretest. Because of similar mastery rates and response patterns for L1 and L2 (e.g., mastery rates were 33% and 36%, respectively versus 80% for M1), these classes were combined in later analyses.

Phase 2 was conducted with 21 sixth-grade students in classroom L3 at a school where 80% of students received free or reduced lunch. All students first completed the Story-evaluation pretest. One science teacher was trained to administer CVS instruction using a procedure that we had found to be productive with some students in the remedial tutoring of Phase 1. This procedure presented CVS in a more incremental fashion: All CVS rules were discussed sequentially, and prior to dis-

These two pictures show how they tested whether or not the ***oven temperature*** made a difference in which cookies people like.
Look carefully at the pictures. Each one shows the oven temperature (High or Medium), sweetener (Honey or Sugar), and number of eggs (One or Three) used to make the cookies.
Do you think this is a good way to find out whether the *oven temperature* (High or Medium) makes a difference in which cookies people like better?

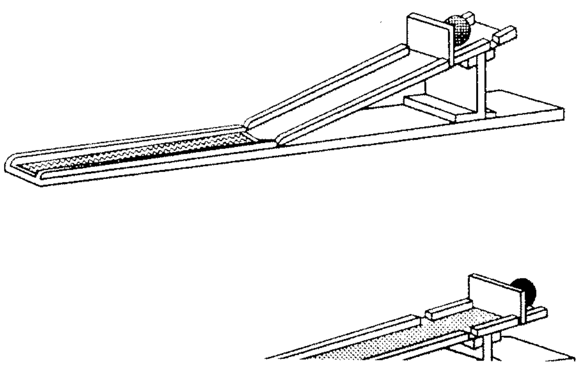
(a) **If you think it is a good way, then circle the word "Good" below. If you think it is a bad way, circle "Bad."**



Good
Bad

(*b*) ***If you circled "Bad," change the picture(s) above to make it a Good comparison.***
(For example, you might want to change the number of eggs, oven temperature, or type of sweetener in one or both of the set-ups.)

Figure 7.1c Story-evaluation questions used in pre- and post-test during Phases 1 and 2. Cookies question. (A "good" experiment because only the target variable is contrasted and all other values are the same in both conditions.)



Contexts of questions serving as data sources by phase

	Phase 2	Phase 3	Phase 4
lass lesson:	*During in-class lesson:*	*Questions tutors could choose among:*	Story-evaluation & design w/explanations pre-, post-, follow-up tests.
mps setups ı worksheets)	Evaluate study experiment (i.e., Does X affect grades?)	Design & evaluate study experiment (i.e., Does X affect grades?) Design of abstract experiment.	Ramps design pre- & posttest w/explanations.
estions	Evaluate ramps experiment.	Design & evaluate running experiment.	
	Immediate Posttest:	Design & evaluate shooting basketballs experiment.	Explicit instruction:
valuate ramps :s.	Evaluate basketball experiment.	Evaluate restaurant experiment.	Evaluate ramps setups (typed or spoken).
ly experi- 'Does X affect	*Tutoring questions included:*	Design of carts experiment (TIMMS . . .).	Standardized item post & follow-up tests (evaluate beetles—NAEP & design carts experiment—TIMMS questions).
	Design & evaluate ramps experiments.	*In the following contexts, tutors could choose among questions addressing R1, R2, and R3 individually, or all three rules (as in the design or evaluate questions):*	
experiment. ning :.	Evaluate basketball experiment.	Hens laying eggs experiments. (for R2:) A farmer wants to know whether older or younger hens lay more eggs. She has some hens that are 1 year old and some hens that are 3 years old. Which ones should she study if she wants to answer her question?	
	Evaluate drink experiment.	a. the 1-year-old hens only	

(*continued*)

Continued)

Phase 2	Phase 3	Phase 4
Delayed posttest:	b. the 3-year-old hens only c. some of the 1-year-old hens and some of the 3-year-old hens	
Story-evaluation & design delayed posttest (drink sales, rocket design, and cookie baking experiments)	Plant growth experiment. (for R1:) James is doing an experiment with plants to find out if the kind of water used with them affects their growth. How should he go about doing that? a. Do something with the soil the plants are living in b. Do something with the type of light the plants receive c. Do something with the type of water given to the plants d. Do something with the location of the plants	
	Dog-calling experiment. (for R3:) James is watching Brittany do her experiment from next door. He can see that for both trials in her experiment, she is standing up, wearing a white shirt, and he can hear her calling for Bandit with the word "Here!" James can't see what she is holding in her hand for each trial, he is too far away. What can James know about the variables he can see Brittany using? a. She is not testing those variables. b. She is testing those variables. c. She's not doing her experiment correctly. d. It doesn't matter what she's holding in her hands.	

included brief discussions of the purpose of science, definitions and examples of key terminology (e.g., "variables," "good experiment"), followed by some CVS rules (e.g., only test one variable at a time). The teacher led the class in designing an experiment (on factors affecting the effectiveness of studying) in a highly scaffolded manner (i.e., by choosing different values for the target variable and the same values for the other variables across conditions).

Students individually designed an experiment for a different target variable on their worksheets and then evaluated an experiment. All worksheet responses were given feedback and returned to students the following day, when students were introduced to virtual ramps apparatus (using the TED computer-based interface) and designed an experiment testing one ramps variable on their worksheets. Students evaluated a confounded experiment and wrote explanations for their evaluations. Then the teacher led a discussion of why the experiment was not informative and how to improve it. Immediately after instruction, students designed four ramps experiments. As in Phase 1, students who designed fewer than three out of four unconfounded experiments were given remedial tutoring. Approximately two weeks after remedial tutoring was completed, all students took the Story-evaluation posttest. One year later, students (now seventh graders) were given a delayed "Story-evaluation/design" posttest that required them to design and then evaluate experiments in three different domains (selling drinks, flying rockets, and baking cookies—see Figure 7.1), and to provide written explanations for each item.

In Phase 3—unlike in previous phases—there was no whole-classroom instruction. First, 27 fifth- and 31 sixth-grade students in classrooms M2a1 & M2a2 of a middle-SES school, where 11% of students received free or reduced lunch, completed the Story-evaluation pretest. Students who corrected fewer than four of the six confounded experiments on the Story-evaluation pretest were tutored—either individually or in small groups—by a member of the research team. The tutor led the student(s) through the computerized instruction that included an introduction to the lesson topic and some introductory definitions (e.g., "variable"), and then required students to evaluate experiments and derive the (CVS) rules for informative experiments. Afterward, tutors selected CVS problems for students from a database of problems, categorized by CVS rule (Rules 1, 2, and 3 from Table 7.1). Tutors selected these problems based on their assessments of students' particular weaknesses. For example, if the tutor believed the student could identify the target variable and vary it across conditions (Rules 1 & 2) but did not understand the need to control nontarget variables (Rule 3), the tutor chose questions targeting Rule 3. All tutoring sessions were recorded and transcribed.

In the final phase described in this chapter, Phase 4, 50 fifth-grade students in classrooms M2b1 and M2b2 from the same middle-SES school as in Phase 3 (but a year later) worked one-to-one with either a member of the research team or used the TED tutor. Students first completed the "Story-evaluation/design" pretest—described in Phase 2—that required them to design and evaluate experiments in their designs. Then they evaluated three different experiments as a "good" or "bad" way to find out about the target variable and explained their responses. Students were given feedback for each evaluation and told why the experiment was or was not informative. After this, students completed the ramps posttest, identical to the pretest. The next day, they completed the Story posttest, identical to the Story pretest. About three weeks later, students completed a follow-up Story posttest. This same procedure was repeated in two low-SES classrooms (L4 and L5, with 16 and 14 students, respectively).

In sum, our analysis of preconceptions about experimental design is based on students' explanations collected during these four tutor development phases. There are three primary sources for these explanations: (a) written explanations given during the teacher-led CVS instruction, (b) written and spoken explanations during one-to-one remedial tutoring sessions, and (c) written and typed explanations on the ramps and Story-evaluation/design pre- and posttests. Table 7.3 provides information about the types of questions given to students across phases.

This explanation-based approach extends and enriches an earlier analysis of CVS misconceptions in which Toth et al. (2000) used a "rule assessment" approach (Siegler, 1981). They classified each student as holding one of four distinct reasoning strategies by examining each student's pattern of evaluations for a 10-item battery of experimental setups including the following four types: unconfounded, singly confounded, noncontrastive, and maximally contrastive (cf., Toth et al., 2000, Table 7.4). However, their analysis rested on the assumption that students approached each of the items in the test battery with a stable—albeit incorrect—strategy. As discussed in Siegler (1996) and also found in our data, the assumption of a stable strategy is tenuous, particularly with students struggling to understand CVS, and whose approaches to it are likely to be changing even within a given assessment. Moreover, even if students had applied the same strategy across problems, it is not clear from this analysis what strategy they had applied, since, as will become apparent, different underlying strategies can lead to the same setup designs or evaluation responses.

Results

In this section, we present an analysis of students' misunderstandings, based on the explanation sources described earlier. The presentation is organized in terms of the goals students held—science, engineering, "fairness," or "domain knowledge"—and the approaches they took within those goals. The analysis is based on five contextual aspects of the student-tutor interaction: (a) student explanations (uttered, typed, or written) in conjunction with the referential setup; (b) our inference about goal type; (c) our inference about error within goal type; (d) the knowledge components addressed in the tutor's response to this information, and (e) the impact of the tutor's responses (i.e., the extent to which the response was effective in reme-

Percentage of responses by goal type, phase, and assessment

	Phase 2: L3	Phase 4: M2b			Phase 4: L4 & L5		
	Story one-year follow-up	Story pretest	Story post	Story follow-up	Story pretest	Story post	Story follow-up
ssification	Q1	Q1	Q1	Q1	Q1	Q1	Q1
l	33.3	12	50	55.3	0	20	26.9
dure[a]	8.3	2	11.4	17.0	0	13.3	15.4
g goal	50	52	22.7	14.9	56.7	33.3	23.1
owledge goal	8.3	0	0	0	3.3	10	0
't know	0	26	9.1	8.5	26.7	16.7	23.1
e	0	2	2.3	0	0	0	3.8
descriptive only)	0	6	4.5	4.3	13.3	6.7	7.7

[a] includes CVS procedure with no explicit science goal.

knowledge component deficits). Later in the chapter, we discuss possible reasons why students hold different goals.

For an overview, Table 7.4 presents the percentages of responses to the first question of the Story pre-, post-, and follow-up tests that indicated a science goal, an engineering goal, an expression of domain beliefs (i.e., "domain knowledge" goal), or another type of response. Initial responses were chosen because students tended to elaborate most on their first response.[2] A second coder categorized a random 10% sample of Story test responses for goal type; intercoder agreement was good (kappa =.90).

CONCEPTUAL AND PROCEDURAL ERRORS WITHIN SCIENCE GOALS

As noted in our task analysis, to understand the point of instruction on experimental design, a student must understand the science goal of the task. That is, the student must realize that the underlying (abstract) goal of designing the comparison is to find out about the causal status of a variable based on the experimental outcome. However, even if a student holds a science goal, errors may occur in the pursuit of that goal due to incomplete or faulty conceptions. Thus, we first discuss the errors students made when designing and evaluating experiments when they held a science goal. As mentioned previously, our inference that a student held a science goal is based on evidence in the explanation indicating that the student was *trying to find out about* the effects of a variable (or variables). For example, we interpreted statements such as "I wanted to see whether [variable X] made a difference" or "I was trying to figure out if downward or straight engines were better" as evidence of science goals. In contrast, statements such as "I wanted to make the balls roll far" or "I wanted to make the balls roll the same," would be explicitly indicative of engineering goals.

Our discussion of these incorrect approaches (the features of which are summarized in Table 7.5) is roughly ordered in terms of decreasing similarity to the expert approach to experimental design. For each approach, we briefly discuss associated underlying knowledge deficits and/or alternative beliefs along with suggestions for remediation. Though the approaches listed do not include every possible feature combination, they include those detected in more than a single instance in our data corpus and/or reported in previous research. Table 7.6 presents a summary of the frequencies of student responses on the first Story test items in Phases 2 and 4 falling into different science goal categories. Table 7.6 also includes responses that indicated CVS application, where students set up unconfounded experiments and gave at least partial CVS responses.

Due to unique response patterns, data from different classes is presented separately in Table 7.6. The first column provides the frequencies of the different approaches for L3 students in Phase 2 on the first design question of the delayed follow-up Story

ummary of science goal experimental approaches and feature(s) erroneous or absent in that approach

	Aspect of CVS missing or misconceived						
	Focus on variable level	Test one variable	Identify correct target variable[a] (R1)	Compare across conditions/outcomes	Vary what testing (R2)	Control variables not testing (R3)	Understand setup/outcome causal relationship
wrong variable			X				
ry-"noncausal"						X (Noncausal belief)	
tiple variables		X	X			X (No CVS logic)	X
onfound(s)						X (No CVS logic)	X
ondition comparison	X	X	X			X (No CVS logic)	X
trastive target variable					X	X	X
ondition experiment				X	X	X	X

tify the correct target variable may occur in each of the approaches, but is only a necessary criterion in three.

ercentage of approach within science goal and CVS responses (and overall percentages) by test and phase

	Phase 2: L3	*[a]Phase 4: M2b*			*Phase 4: L4 & L5*	
	Story one-year follow-up Q1	*Story pretest[b] Q1*	*Story post Q1*	*Story follow-up Q1*	*Story post Q1*	*Story follow-up Q1*
cience goal & CVS responses (%)	5 (41.6%)	7 (14%)	27 (61.4%)	34 (72%)	10 (33%)	12 (42%)
logic	0	14.3 (2.0)	11.1 (6.8)	17.7 (12.7)	0	16.7 (7.0)
cience goal	0	14.3 (2.0)	55.6 (34.1)	50 (36.0)	20 (6.6)	16.7 (7.0)
10 goal stated)	20 (8.3)	0	14.8 (9.1)	11.8 (8.5)	40 (13.2)	25 (10.5)
science goal but incomplete)	0	0	0	2.9 (2.1)	0	8.33 (3.5)
	20 (8.3)	**28.6 (4.0)**	**81.5 (50.0)**	**82.4 (59.3)**	**60 (19.8)**	**66.7 (28.0)**
: test wrong variable	20 (8.3)	0	3.7 (2.3)	0	10 (3.3)	0
ry- "noncausal"	0	0	0	0	0	0
ltiple variables	40 (16.6)	0	11.1 (6.8)	5.9 (4.2)	0	8.3 (3.5)
onfound(s)	0	14.3 (2.0)	0	2.9 (2.1)	0	8.3 (3.5)
ondition comparison	20 (8.3)	57.1 (8.0)	0	5.9 (4.2)	20 (6.6)	0
trastive	0	0	0	0	0	0
ondition experiment	0	0	0	0	0	8.3 (3.5)
ce goal	0	0	3.7 (2.3)	2.9 (2.1)	10 (3.3)	8.3 (3.5)

ı was not included in this table because unlike in the other phases, there were no assessment events that were common to all students.

posttest taken one year after instruction. Next, frequencies are given for the Phase 4 fifth-grade M2b students' responses on the first Story pretest design question and on the first design question of the immediate and follow-up Story posttests—administered three weeks after instruction. The same data is provided for L4 and L5 (because they did not express any CVS or science goal responses on the first Story design pretest item, this column is not included in the table). Data from this table is referenced throughout the discussion of various science goal approaches.

As shown in Table 7.6, among students expressing science goals on the pretest, CVS responses were relatively rare. However, given that students held a science goal, Phase 4 students in both M2b and L4/5 were most likely to express a CVS understanding on the first Story posttest question (82% and 60%, respectively). However, the L3 Phase 2 students were only 20% likely to give a CVS response on the one-year follow-up. Furthermore, M2b students were relatively more likely than L4 and L5 students to explicitly indicate a science goal when they expressed a CVS explanation (e.g., "I only made the age of the child different because I want to find out if the age makes a difference") on the posttest. CVS explanations given by the M2b students generally included explicit indications of science goals whereas those given by L4 and L5 students did not. It is possible that L4 and L5 students were more likely to learn the CVS procedure without understanding its purpose in serving science goals.

Types of Error-Prone Approaches within a Science Goal

Here we describe the set of experimental design errors exhibited by children expressing a science goal at that moment (summarized in Table 7.6), along with corresponding remediation for each approach. The first of these (a) is an unconfounded experiment but for the "wrong" factor. The next four (b–e) are different types of errors that led to confounded designs. In these, students held science goals of designing experiments to find out about a variable (or variables) and understood the need to "compare and contrast" the tested variable(s) across conditions, but failed to control the other variables. The final two (f & g) are other types of errors. A second coder categorized a random sample of science goal responses; intercoder reliability was good (kappa =.86).

(a) CVS for wrong variable (R1 failure): This approach occurred when students were attempting to find out about a single variable and set up an experiment that was unconfounded, but with respect to the wrong variable (i.e., a variable other than the one given in the question).[3] For example, when asked to design an experiment to test time of day in the drink sales story problem (Figure 2a), one student designed an unconfounded experiment to test age instead. His justification was:

I made one variable so I will find out if age sells more.

Students may design unconfounded experiments to test something other than the target variable because they believe they already know the causal role of the (teacher)-intended target variable, and substitute it for one that they would prefer to find out about. This is an action they would likely take when engaged in open-ended science inquiry, and not an unreasonable one. The "error" here is bringing to bear beliefs about likely causal factors in a domain rather than following the "academic" exercise of suspending prior beliefs in the service of a kind of Platonic approach to experimental design. Another possible cause of this error is simply that children did not read the problem statement carefully.

Remediation: In general, when students who made this mistake were prompted to reread the problem statement, they realized that the variable to be investigated was given. They subsequently design unconfounded experiments to test that variable with minimal help from the tutor, as shown in the following exchange between a tutor (T) and student (S):

T: Ok, so is this [CVS but wrong target variable] a good experiment?
S: Ah, yes?
T: Ok... all right, so let's think about that a little more. The first thing you want to do is find out which of the variables they want to find out about. OK? So which variable do they want to find out about?
S: Where they study?
T: Exactly. So which variable would that be?
S: Location?
T: Exactly, so they want to find out about location. OK? So now that you know that, what do you think? Is that a good experiment?
S: No.
T: OK, why not?
S: Because they're both the same?
T: OK, and how do you want to change that?
S: Shayna should study on the couch.
T: OK, all right, so now we made this different. OK? So is this a good experiment now?
S: No.
T: OK, what do you wanna change about it?
S: The time of day?
T: OK, so what do you want to do?
S: Make both of them... late?
T: OK, so it doesn't matter, they could be both late or both early. But the important thing is to make them both...
S: The same. (unconfounded experiment for the target variable now)
T: Exactly. And anything else you want to change?
S: [No]

likely to make this error on the Story tests. In all of these cases, there was no indication from students' written explanations that this error was more than a slip—as opposed to students' desire to test a different variable due to their beliefs about specific variable effects. Thus, converging evidence suggests that this error was *not* generally caused by students applying their beliefs about variables.

(b) CVS but vary "noncausal": Kuhn et al. (1995) described a type of behavior in a task where participants (fourth graders and adults) had to choose evidence from which to infer the effect of a particular variable where participants simply ignored variables they believed were noncausal when making their choices of which evidence to compare. This behavior led to confounded comparisons.

However, we failed to find evidence of students explicitly indicating they did not control a variable because they did not think it mattered. For example, as shown in Table 7.6, identified instances of this approach were nonexistent in the pre- and posttest responses to the first Story question, as well as in responses to the other items. This may indicate that, when students are applying science goals, they tend to suspend their beliefs about variable effects. As discussed later, varying what one believes to be noncausal was more common within engineering goal applications.

Though infrequent, instances of what may be this approach were identified in other student responses, including those given during tutoring sessions (refer to the excerpt below). In this approach, students understand the causal relationship between the experimental setup and outcome (i.e., to determine whether a variable affects an outcome, only that variable can differ; otherwise, the cause cannot be determined) when making inferences about variable effects. However, they fail to realize their beliefs may not be consistent with reality, and that therefore they must control even nontarget variables they believe are not causal.

Remediation: During the tutoring sessions, we used two different methods to remediate this approach when we suspected students of using it. In the first, we told students: "even if you think a variable doesn't matter, you must still control that variable, because you can never be completely certain that it does not affect the outcome." This explanation led to the adoption of the expert approach. When it did not, we used a second method in which we either suggested or asked students for a plausible explanation for why the (confounding) variable *might* make a difference. This method is shown in the example below, where a ramps experiment testing for starting position is confounded by ball type (red or yellow):

T: Now can we tell whether one ball rolls farther than the other because they have different starting positions?

S1: Yes.

T: How about those balls? What if one's a ping-pong ball and the other's a marble? Could we tell whether it's just the starting position of the balls that's affecting [illegible]

T: If the red ball's a ping pong ball and the yellow ball's a marble, how would you...

S2: Make them the same ball.

Between these two remedial methods, this particular error was almost always corrected.[4] It is important to note, however, that the same behavior shown in the tutoring excerpt may be due to students ignoring confounding variables (discussed later).

Failures in Understanding CVS Logic

In the next two approaches, students demonstrate lacking an understanding of the indeterminate nature of confounded experiments by varying other variables. Two alternative explanations for this behavior exist. Toth et al. (2000) assumed that students who designated as "good" any experiment that varied the target variable, regardless of the settings for the other variables, were ignoring the confounded variables. However, without the additional evidence provided by participants' explanations, students may have had other reasons for their "good" and "bad" designations of different experimental designs. For example, Toth et al.'s participants may have been trying to find out about more than one variable at a time, as Kuhn and Dean (2005) speculated:

> Students who have developed an understanding of the need to access an available database as a source of information may nonetheless still initially pose ineffective questions, *in particular because they aim to discover the effects of all variables at once. It may be this ineffective intention that leads them to simultaneously manipulate multiple variables (in effect, overattending to them, rather than underattending by failing to control them, as is typically assumed)* [italics added]. (p. 867)

Our explanation database suggests that students both overattend to the uncontrolled variables (by trying to test multiple variables), and underattend to the uncontrolled variables (by simply ignoring them). As shown in Table 7.6, these two response categories were detected at similar rates, though were relatively uncommon overall (e.g., in Phase 4, these responses represented between 0 and 16% of responses—an average of 10% of science goal responses). Additionally, with instruction, the rate of overattending responses increased while the rate of underattending responses decreased, suggesting that variables become more salient over the course of instruction. The corresponding knowledge components comprising these two categories are described in Table 7.5 (c–d), and these categories are further elaborated on below.

(c) Test multiple variables: One reason that students design confounded experiments is that they attempt to find out about *each* of the contrasted variables in

a single experimental contrast. Students almost always indicated wanting to test all of the variables; thus, they typically did not apply any beliefs about variable effects (such as controlling variables they believed were noncausal), but rather, held a "pure" abstract science goal. This is shown in the following student's explanation for why she evaluated a maximally confounded experiment as good:

> I think this is a good way because testing the differences can probably turn out to be a good experiment and you could see what effect the time, drink, and person [age] have on them.

Kuhn et al. (1995) reported similar conceptual errors in students who attempted to find out about multiple variables simultaneously. This approach reveals two related misconceptions: first, a belief that the effect of the target variable can be determined from a confounded experiment, and second, a belief that the effect of the nontarget variables can be determined from the same experiment. That is, students believe that they can "do it all" in a single experiment. These misconceptions stem from failure to understand the logic of CVS: that the cause of any differences in outcomes cannot be uniquely determined in a confounded experiment.

Kuhn et al. (1995) noted that as their students progressed through a series of experimental design problems, they became less likely to express "an intent to assess effects of multiple features by examining a single instance or pair of instances" (p. 65). Thus, at least some students who hold science goals and are trying to find out about the effects of more than one variable at a time can eventually learn to focus on only one variable when no feedback is provided. However, there was a point in the Kuhn et al. intervention in which the number of features students intended to investigate stopped decreasing, suggesting the need for more instructional support.

Remediation: Because they only lack an understanding of the rationale for controlling the nontarget variables, students applying this approach are ideal candidates for explicit CVS instruction (e.g., as given in Chen & Klahr, 1999), which explicitly focuses on this rationale. In Phase 1, all of the five students (in L1 or L2) who held science goals and understood the need to vary what was being tested but lacked the procedural knowledge of controlling for nontarget variables (as shown by their ramps pretest setups) designed at least three out of four informative experiments on the ramps posttest following explicit CVS instruction. However, for some students, simply designing experiments during the Story or ramps pretest—where no feedback was provided—allowed for their development and application of CVS.

(d) Ignore confound(s): Students may set up maximally confounded experiments or evaluate them as good, but only refer to the target variable, ignoring the other variables. Thus, again, students do not apply their beliefs about the nontarget variables. For example, the following statement was made by a student evaluating a maximally type of drink. Similarly, a student who had designed a maximally contrastive rocket experiment simply explained: "I wanted to see if the *engines* would affect the speed and direction of rockets." Again, the student ignored the nontarget variables (body shape and number of windows).

Such explanations suggest three possibilities: (1) students did not attend to the uncontrolled variables, indicating that they did not recognize the importance of controlling variables either; (2) students did attend to the uncontrolled variables, but did not understand that the (potentially causal) confounded variables prevent valid inferences about the effect of the target variable; (3) students attended to the uncontrolled variables and understood the rationale for controlling variables, but believed they were noncausal and not necessary to control (i.e., they actually held a "CVS but vary noncausal" approach).

Remediation: In the first two possibilities, students are prime candidates for explicit CVS instruction (e.g., as given in Chen & Klahr, 1999, and Strand-Cary & Klahr, 2008), which explicitly focuses on the nontarget variables and their impact on inferences that can be made. For the third possibility, the remediation described for CVS vary noncausal is appropriate.

(e) Whole condition comparison: Another approach within a science goal is comparing two conditions rather than an individual variable or variables. This approach—typically associated with maximally contrastive designs—once again does not involve application of beliefs about variable effects but is an "abstract" science goal application. This approach is exemplified in the following explanation: "I set it up the way I did so everything would be different to see which one [cookie] people like better."

A second example is from a tutoring excerpt in which the tutor and student discuss an experiment about plant growth (in Phase 3):

S: Umm... no [changes experiment from CVS to maximally contrastive]
T: So why did you set it up like that?
S: To be different to see which one would grow better.

The frequency of this approach (shown in Table 7.6) ranged from 0 to 57% of Story test responses within science goals, and was greatest on the pretest responses of Phase 4 M2b students. In addition to not considering the variable level, these students probably also lack an understanding of controlling variables and the rationale for doing so.

Remediation: Perhaps because it is natural for students to think at the concrete level of variable values, simply prompting them to consider the variable level by identifying the given target variable and then prompting them to think through the logic of the relationship between the target variable and outcome resulted in subsequent consistent CVS application. This is shown in the following dialogue (continued from the dialogue above):

T: . . . makes a plant grow better. So if Plant #1 grew better than Plant #2, would he know for sure, the way he has this set up, that it's because of the fertilizer?

S: Yeah.

T: How would he know that?

S: No.

T: How could he set up his experiment so he would know for sure that it's the soil that the plant is in that is making any difference in how the plants grow.

S: These would all have to be the same.

T: So why don't you go back and make those changes.

(f) Noncontrastive target variable(s) (R2 failure): Students sometimes held a science goal but did not contrast the levels of the target variable across conditions. Thus, these students did not seem to understand the need to compare different values of a variable (Rule 2). However, these cases were relatively rare in our data. This may be because, as Kuhn et al. (1995) noted, the idea of comparing things to see if there is a difference is intuitive for students and understood from a young age. In the overwhelming majority of cases in which students designed setups where the target variable was not contrasted, they indicated engineering rather than science goals. Of the six analyzed interventional points shown in Table 7.6, there was *no* evidence of students failing to contrast what they indicated they were testing. In all of the Story evaluation/design pre-, post-, and delayed posttest responses to design questions (from Phases 2 and 4), there were only three instances out of 622 responses (less than .5% of responses) where (two different) students expressed a science goal but did not contrast the variable they indicated they were testing. For example, one student designed a noncontrastive comparison for the cookies item: "so I can see what they like the best." It is possible that such responses were slips, where students mistakenly failed to contrast variable(s) in their setups. Regardless, it is safe to say that cases of failing to contrast the variable(s) of interest when the student held a science goal were rare in our data. Thus, science goals and the idea of contrasting variables appear to be strongly associated.

Remediation: Our remediation for failure to contrast the target variable focused on helping students understand the rationale for comparing outcomes across conditions. Students generally understood that comparing the same value of a variable would not provide information about whether that variable had an effect (because it would always produce the same results, so it would not be possible to find out if the different values have different effects on the outcome). When asked "If the [values of the target variable] were the same, would you be able to tell if [the target variable] made a difference in the result?," students generally came to realize the need to contrast values of the target variable across conditions, as in the following:

T: OK, so now we have two ramps that are identical. If you did this experiment, can

T: How could you change this so that you could do an experiment to figure out how starting position affects [the result].

S2: You could put one in the middle and leave the other one there [at the top].

(g) Single-condition experiment: The least sophisticated science goal approach found in our corpus is one in which students did not compare results or contrasting variables across conditions, but rather viewed each condition as a separate experiment. As shown in Table 7.6, instances of this approach were rare, identified in only one analysis point (where it was 8% of responses of the low-SES students in Phase 4 on the Story follow-up). In this approach, students consider an outcome from a single condition to allow for inferences to be made about the effects of individual variable values, exemplified by the following justification:

> I set up this experiment like the way I did because I wanted to see if tilted [engine] had an effect on how fast the rocket went and I picked straight [engines] for the other one to see if down had an effect on to [*sic*] see how fast the rocket went also.

Here, the student's use of "also" implies that she isn't comparing the outcomes of conditions to each other, but rather viewing the two setups and their outcomes independently. This type of error was noted by Kuhn et al. (1995), who found that when trying to discover the effects of variables given instances with variable settings and an associated outcome, fourth-grade students did not compare instances for about half of the inferences they made.[5]

There are likely two deficits underlying this approach. First, rather than thinking of outcomes as relative (e.g., that one is "better" than another), students view outcomes as absolute (e.g., "good" or "bad"), without realizing that an absolute outcome is subjective (i.e., how does one determine what a "good" outcome is?). Second, students may ignore the nontarget variables. That is, they may attribute an outcome solely to one variable, ignoring any potential impact of the other variables. However, if the student is attending to the other variables, he or she may either (a) fail to understand the causal ambiguity of the experimental contrast they have created, or that it is impossible to know the effects of the individual variables on the outcome, or (b) understand this, but apply their beliefs about the effects of the other variables to estimate the target variable's contribution to the outcome. For example, if the student believes a rocket's windows and engine direction have no effect on how high it flies, he or she may attribute a "good" outcome to the rocket's body shape.

Remediation: Because of the small number of identified cases of this single-condition experiment approach, evidence for the success rates of remedial methods

[5] One possible reason that this type of noncontrastive science approach was more common in Kuhn

are not available. However, at minimum, this remediation would involve addressing all CVS rules.

Summary of Erroneous Approaches within Science Goals

As demonstrated in the prior discussion, students tended to explicitly state their science goal intentions when explaining their designs. Additionally, when students applied science goals, they rarely expressed or applied their beliefs about the variables. The most common error within a science goal orientation was failing to control nontarget variables (refer to Table 7.5). Rather than ignoring variables considered noncausal, this error was due to students not realizing that confounds render experimental results uninterpretable. Why might this be the case? According to Kuhn et al. (1995), in the real world, people do not see many unconfounded experiments. Instead, they typically observe correlations among lots of data—even over time—and make inferences based on those. These correlations are embedded in the noise of confounds that cannot be controlled. From a child's point of view, this turns out to be a relatively effective, thus practical, way to discover causal patterns. However, it can lead to conceptual errors. For example, people commonly infer that heavier objects fall faster than lighter ones based on their correlational observations, but fail to account for confounds such as surface area or object shape. But given its apparent utility, it is not surprising that children draw inferences from confounded experiments, without thinking through the logic of the causal relationship between the setup and possible outcomes.

When students begin with a science goal and understand the idea of comparing outcomes to see whether a variable has an effect, prompting them to think through the logic of a confounded experiment helps them to realize the "power" of CVS. We believe that this is one reason why the CVS instruction developed by Klahr and colleagues, which does just that, promotes more rapid and immediate CVS gains than discovery learning (Chen & Klahr, 1999; Klahr, 2009; Klahr & Nigam, 2004; Kuhn & Dean, 2005; Strand-Cary & Klahr, 2008).

Engineering Goals

As stated previously, students may hold or adopt goals other than the science goal of trying to find out about some variable(s) in tasks that—from the instructor's perspective—involve experimenting. Others (e.g., Kuhn, Amsel, & O'Loughlin, 1988; Schauble et al., 1991; Tschirgi, 1980) have noted this as well. Unlike erroneous approaches within a science goal, which can be compared to the expert model of experimental design in terms of missing and faulty knowledge, approaches involving alternative goals cannot be sensibly mapped onto the expert model due to their fundamentally different knowledge structures.

As discussed earlier, Schauble et al. (1991) termed one alternative class of goals

they are domain specific in the sense that students intend to produce an effect that is specific to the particular context. Whether students apply their beliefs about variable effects within various types of engineering goals will be explored throughout this section.

In what follows, we discuss the different kinds of engineering goals expressed by students in various phases of our development project. Unlike with science goals, students rarely stated their engineering goal intentions explicitly (as in, e.g., "I am trying to make the balls roll far"); their goals had to be inferred from the rationales they gave for their design selections. This suggests that, while science goals are generally linked to explicit knowledge, engineering goals are not. Frequencies for the different types of engineering goal approaches were assessed at the same intervention points as in the analyses of science goal approaches and are given in Table 7.7. Intercoder reliability for categorization of engineering goal responses was good (kappa = .90). After discussing the alternative goals, we discuss potential methods of eliciting science goals.

Types of Engineering Goal Approaches

(1) Maximize outcome: According to Scauble et al. (1991), "the main objective of engineering practice is to optimize a desired outcome" (p. 860). One type of optimal outcome is the *maximum* outcome, such as the fastest car, the highest flying rockets, or the best-tasting cookies. As shown in Table 7.7, maximize outcome goals were common, particularly on the pretest, where they accounted for 88% and 96% of all engineering responses for Phase 4 low- and middle-SES students, respectively. However, they still accounted for a fairly large percentage of responses on the posttest, between 71% and 90% of engineering responses for Phase 4 students, and all engineering responses of the L3 students in the Phase 2 follow-up. Students applied maximize outcome goals in three ways. In the first two, students designed setups to maximize effects in *both* conditions, assuming independent or dependent variable effects. In the third, students maximized the outcome in just one condition. Maximize outcome responses were roughly evenly divided among their three types.

(a) Maximize outcomes (independent effects): Students applied the goal of producing a maximum effect in each condition by choosing the "best" value for one or more variables for both conditions. This often resulted in students setting up identical conditions, or designing noncontrastive "experiments." In the following, a student expresses this type of maximizing goal for her design, where both rockets had straight bodies, downward engines, and one window:

> I don't know, it just makes sense to have a straight body and down[ward] engines so it goes straight up.

able 7.7 Percentages of approaches within engineering goals (and overall percentages) on Story test Q1 by test and phase

pproach	*Phase 2: L3*	*Phase 4: M2b*			*Phase 4: L4&L5*		
	Story follow-up	*Story pretest*	*Story post*	*Follow-up*	*Story pretest*	*Story post*	*Story follow-up*
lumber (percent) engineering response	6 (50)	26 (52)	10 (23)	7 (15)	17 (56.7)	10 (33)	6 (23)
a. Maximize both (independent effects)	33 (16.5)	26.9 (14.0)	30 (6.9)	57.1 (8.6)	35.3 (20.0)	30 (9.9)	50 (11.5)
b. Maximize both via interactions	33 (16.5)	34.6 (18.0)	40 (9.2)	0	11.8 (6.7)	10 (3.3)	0
c. Maximize one	33 (16.5)	34.6 (18.0)	20 (4.6)	14.3 (2.1)	41.2 (23.4)	40 (13.2)	16.7 (3.8)
laximize outcome (unclear for one/both)	0	0	0	0	0	0	16.7 (3.8)
otal max outcome	**100 (50)**	**96.2 (50.0)**	**90 (20.7)**	**71.4 (10.7)**	**88.3 (50.1)**	**80 (26.4)**	**83.4 (19.2)**
. Different outcomes	0	3.8 (2.0)	10 (2.3)	28.6 (4.3)	5.9 (3.3)	10 (3.3)	0
. Same outcomes	0	0	0	0	0	0	0
)ther engineering[a]	0	0	0	0	5.9 (3.3)	10 (3.3)	16.7 (3.8)

For example, "normative" goals, such as "That's how I would do it."

opposite the behavior noted by Schauble et al. (1991), where students *contrasted* those variables they believed mattered. In this form of maximizing goal, students are not *comparing* the two conditions, but rather considering the conditions independently. (It was not uncommon for students with engineering goals to think of one of the conditions *as* an "experiment.") This behavior is demonstrated in the following explanation for what appears to be an unconfounded experiment testing for number of eggs (Cookie A: 350 degrees, sugar, three eggs; Cookie B: 350 degrees, sugar, and one egg)[6]:

I did 350 because it's sort of in the middle. Not a lot of people like honey in cookies. I guessed about the eggs.

(*b*) *Maximize outcomes via interaction*: There were also cases where students intended to maximize the outcomes of both conditions, but, because their domain theories involved interactions—where different combinations of variable values produced good outcomes—their resulting setups were contrastive, and often maximally contrastive. For example, one student designed a maximally contrastive drink stand "experiment" (Setup A: noon, older child, and iced tea; Setup B: 3:00 p.m., younger child, lemonade), and explained:

Most younger kids are outside around 3 pm and older kids like to get up in the noon [*sic*]. Also more younger kids like lemonade than iced tea.

In this example, the student's belief that the better values for time and drink depended on the age of the child resulted in two two-way interactions.

(*c*) *Maximize outcome in one condition*: The final form of the maximize goal involved maximizing the outcome (either referencing dependencies among variables or not) in only *one* condition. Students who expressed this form of maximizing goal typically did not provide a rationale for setting up the other condition, which they tended to set up as maximally contrastive from the maximized-outcome condition (e.g., 12 of the 14, or 86%, of identified maximize outcome one condition explanations in Phase 4 middle-SES classroom Story pretest responses were associated with a maximally contrastive setup). Less commonly, students set up only one condition, completely ignoring the other.

In the following, a student who set up a maximally contrastive drinks "experiment" (Stand A: noon, younger child, lemonade; Stand B: 3:00 p.m., older child, iced tea) with dependent variable effects, expresses this type of engineering goal:

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 using this approach, did not mention Stand B at all. Perhaps in these instances, students did not address the other condition because they did not consider it essential to achieving their maximize outcome goal. However, it is unclear why students tended to set up the second stand opposite to the first. Perhaps they were setting variables to their opposite values for aesthetic reasons (e.g., for "variety"). Alternatively, perhaps students were contrasting those variables they believed to be causal, the behavior noted by Schauble et al. (1991), implicitly attempting to produce different outcomes (discussed next).

(2) *Different outcomes:* Student responses sometimes indicated intents of producing different outcomes across conditions. Unlike the maximizing goals just discussed, where students considered the conditions independently, this form of engineering goal is comparative. This type of engineering goal often requires students to apply their beliefs about variables' effects. As shown in Table 7.7, these responses were less common than maximize outcome responses at every analysis point. As with maximizing goals, it was quite rare for students tô explicitly state different outcome goals, as in: "because they will both fly a different way." More often, students' different outcomes goals had to be inferred from their explanations, as with the following:

> I set it up the way I did because I think it is more likely for stand A to sell more because she's young and cute, because more people are around at noon and because lemonade sometimes sells more. Then I set up stand B to be the exact opposite.

The behavior of making variables "that matter" different, noted by Schauble et al. (1991), too, may result from implicit goals of producing different outcomes. In our data, students occasionally expressed a desire to [only] vary those variables they believed to be causal. For example, for the drink sales question in which students were asked to design an experiment to test time of day, a student designed an experiment in which only the type of drink differed and explained: "I know that it does not matter whether how old or what time all that matters is the type of drink!" These responses—though again not explicit—may be indicative of different outcome goals.

What is the underlying motivation for different outcome goals? One possibility is that children try to produce results that are consistent with their expectations about variable effects. 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—and if the student already believes slope to be causal—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. For this particular goal, the effort of using one's beliefs about variable effects to "prove" the effect of the target variable makes this a type of engineering goal. However, it differs from other engineering goals in which variable effect beliefs are applied to produce some optimal

A more likely explanation for a different outcome goal—supported by our data—is a (perhaps implicit) belief that a "good" experiment is one that produces different outcomes. In other words, some students may not conceive of an "experiment" in terms of science goals, but rather within an engineering interpretation. For example, students often evaluated the unconfounded experiment testing the number of windows on a rocket as "bad" because they believed that the number of windows did not have an effect, as expressed in the following:

> Even though they have different amount of windows doesn't mean that if one has 4 windows it might fly more faster and farther than another.

This student changed the setup to maximally contrastive, which (presumably) would be more likely to produce a different outcome. Implicit beliefs that a good experiment is one that produces different outcomes might have been triggered during instruction when students were asked: "Is this a good way to find out whether the balls go different distances just because of the ball?" Consistent with this possibility, the proportion of different outcome goals increased from the Story pretest to immediate Story posttest (Table 7.7).

(3) *Same outcomes*: Some student responses indicated that their specific engineering goal in designing an experiment was to produce the same outcomes in both conditions. For example, students set up two ramps with the goal of making the ball roll the same distance on each ramp. Or they evaluated an experiment as good because they believed that both setups would (or *should*) produce the same outcome (e.g., equal sales). To our knowledge, this specific type of engineering goal has not been previously reported, likely because it is relatively rare. We found no evidence of this in Story pre- or posttest responses (Table 7.7). However, students in Phase 1 expressed this misconception in about 15% of their explanations given during the classroom instruction. Students in Phase 1 may have expressed this type of approach more than in most other phases due to the instructional wording—which asked students if the experiments were "fair tests" of the target variable. Some students may have interpreted "fair" to mean "the same," thus eliciting this particular type of engineering goal. In subsequent versions of instruction, the word "fair" was replaced with "good." No evidence of this type of misconception was uncovered again until the evaluation phase of computerized instruction in Phase 4. When L4 and L5 students worked with the TED ("Training in Experimental Design") tutor, 7.4% (2 out of 27) of responses to "Tell me why this [unconfounded experiment] is a good way to find out whether the balls go different distances just because of the ball" indicated a same outcome goal. Immediately before answering this question, students heard the following explanation:

> To fix this experiment [to test for ball type], we would need to make everything

It is possible that students—especially those students who did not hold science goals and thus did not understand why controlling nontarget variables is necessary—interpreted these explanations as "everything must be the same," eliciting (or elicited by) "same outcome" goals.

(a) Same outcomes via noncontrastive setups: Students can apply "same outcome" goals in several ways. The simplest is by using identical conditions. For example, students might compare a high ramp with a smooth surface to a high ramp with a smooth surface. Note that here it is *not* necessary for students to apply their beliefs about the effects of the variables. This goal is expressed by a student in her written explanation for why an unconfounded experiment testing ramp height "is a fair comparison for height" in the following:

I think we should have made them both high because I think it would have been really close or the same. I think that's what we should have done.

Interestingly, though, this student mentioned that the ramps should be "high" rather than "the same," indicating that she may have also been applying an implicit maximizing goal.

(b) Same outcomes vary "noncausal": Another manifestation of the "same outcome" engineering goal *is* influenced by students' beliefs about variable effects. In this case, students design setups they believe will produce the same outcome by making the variables they believe "matter" the same across conditions, but, as in previous approaches described, vary the variables they believe "don't matter." Students may express this approach by creating a setup in which the slope and surface of the ramps (which they typically believe to be causal) are the same, but the color of the ball (which students generally say does not matter) is varied. The following student explanation for why an experiment in which only ramp height differs is "a fair way to find out about the height" exemplifies this approach:

[It's] fair b/c high/low probably doesn't matter because it would probably travel the same length.

It appears that this student's goal of producing the same outcomes actually *overrode* his belief that a high ramp would make the ball roll farther, which he had expressed earlier.

(c) Same outcomes via "balanced" setups: Alternatively, students tried to engineer same outcomes by setting variables to "balance" each other across conditions. This approach requires applying one's beliefs about the effects of context variables. For example, students made 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 (surface and height) "cancel each other out," the student explained, the ramps should roll the same distance. In a related case, a student explained why he varied the length of the ramp (the target variable), in relation to the surface:

why he varied the length of the ramp (the target variable), in relation to the surface:

Because the rough one [inaudible] probably mess the ball up so I made it longer so it could move faster.

This approach may arise when students who hold the engineering goal of producing the same outcome learn to vary the target variable and assimilate this new (correct) information onto their prior beliefs. Such assimilation was identified in Vosniadou and Brewer (1992), where some students developed hybrid models of the earth (e.g., as round, but with a flat interior, representing the earth's surface), by assimilating new information (e.g., that the earth is round) with their intuitive preconceptions (i.e., that the earth is flat).

Goal Hybrids and Goal Consistency

In a few rare instances, students applied *both* engineering and science goals within a given setup. For example, on the first question of the ramps pretest, only 3 out of 70 (4%) responses indicated both types of goal. This "hybrid" goal is shown in the following explanation, where the student applied engineering goals for her choices of slope and expressed a goal of finding out about the starting position:

I thought that you should use steep because if it was flat it wouldn't be able to roll that well. I don't know why I used [different surfaces], I just thought it would work. I picked the middle because it's a shorter distance and it might roll farther. And I picked the top to see if it might roll farther than the middle.

It was less rare for students (especially those in the Phase 4 human-tutor condition) to express both science and engineering goals on different questions during the ramps pretest, where 16 out of 70 students (23%) expressed both types of responses. Figure 7.3 shows the number of L1&L2 and M2b students expressing only science goals throughout the ramps pretest, only engineering goals, at least one science goal and at least one engineering goal, and neither of these goals (e.g., they gave descriptions of their setups but not the underlying rationale). However, most students (48 out of 70, or 69%) expressed only one goal type throughout the ramps pretest. Thus, in a feedback-free task in a single domain, goals were stable for the majority of students.

Overextended "Fairness" Schema

Closely related to "same outcomes" engineering goals, students occasionally expressed the goal of making the setups "fair" or giving the conditions—rather than

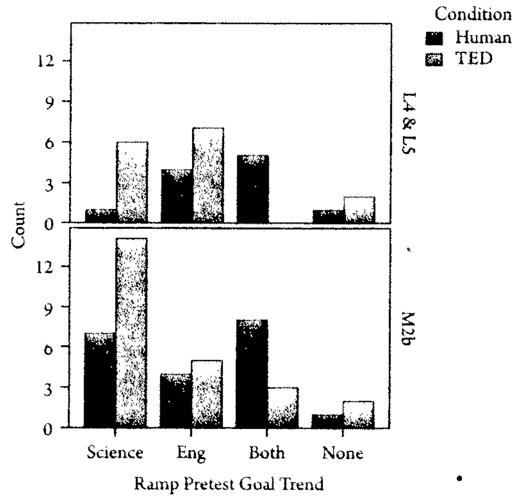


Figure 7.3 Frequency of goal response trends during ramps pretest, by school and condition.

that the setup at least allows for a reasonable possibility of the same outcomes. Additionally, this is not considered a science goal because rather than attempting to discover causal relations in variables, the student is attempting to make the comparison "fair." For this reason, this goal differs from the noncontrastive target variable(s), or Rule 2 failure, science goal approach. This type of response was not detected in students' responses to the Story test questions. However, it was detected in explicit instruction in 10.5% (4 out of 38) of Phase 1 low-SES students' responses to "Why is this a fair comparison for height?" and in 11% (3 out of 27) of responses to "Tell me why this is a good way to find out whether the balls go different distances just because of the ball" during the instruction in Phase 4. This type of response was not found at the same instructional points for the middle-SES students.

This approach is expressed in a student's explanation for why his noncontrastive (NC) experiment is a "fair test of [the target variable]," even though the balls might roll different distances:

Because the surface... one might go farther, one might go [less far] and if the one ball goes less and one ball goes more I think it's fair because the surface is both the same way, the height [slope] is the same, they're both long.

Another example comes from a student responding to why an unconfounded experiment "is a fair comparison for height." This student's overextended fairness schema seems to have overridden his perception of the experimental setup, in which the heights were actually different:

Like the same outcome goal, this misconception may arise when students interpret instructional explanations about controlling variables within the context of a "fairness" schema. Students' explanations such as these indicate an explicit understanding of "fairness," which they apply to their evaluations of ramps setups when they justify the "fairness" of the comparison in terms of whether the variable settings are the same.

"Domain Knowledge" Misinterpretations and Relationship to Engineering Goals

Perhaps the "deepest"—and most likely implicit—misinterpretations that occurred during instruction were when students interpreted questions about experimental design as either asking them to apply their beliefs about the effects of variables to *predict* the experimental outcome or as asking about the effects of specific variables on an outcome (without necessarily predicting the experimental outcome). This misunderstanding is analogous to the behavior of judging the validity of a logical syllogism based on whether the conclusion corresponds to one's beliefs rather than by applying abstract deductive reasoning. As shown in Table 7.4, these responses were relatively rare on the Story tests. This type of response occurred more often during instruction when students were asked to evaluate experiments for whether they would provide evidence of the causal status of a target variable (e.g., "If we ran this experiment, could you tell whether the [target variable] made a difference in how far the balls roll?"). The following is an example of a student misinterpreting the tutor's question as about predicting the outcome:

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?
S: It doesn't gain that much speed.

Students sometimes responded to questions about the determinism of experiments by simply stating their beliefs about the effect of a variable on an outcome, without necessarily using them to predict the specific experimental results. For example, in the following instance, a student simply stated her belief about the effect of the age of the seller on the experimental outcome when asked whether a maximally confounded experimental setup was a good way to find out whether age makes a difference:

"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 setup, as asked.

whether the experiment would produce results that could lead to valid claims about the target variable as asking about the effect of that variable on the result:

T: What if I put it this way: can we tell if it's just the slope of the ramp that makes a difference in how far the balls roll? Using this (maximally contrastive) experiment that you have set up here, could you answer that question?
S: Yes. It would affect how far the ball rolls.

Though it is possible that students who expressed these "domain knowledge" misinterpretations actually held science goals and simply "skipped" to the prediction stage of experimentation, this is not supported by our data. Students who gave "domain knowledge" responses on the Story-evaluation/design pre-, post-, and follow-up tests almost always gave engineering explanations but no science explanations in their other responses. Thus, when students gave responses about the effects of variables on outcomes, they most likely held engineering goals.

More evidence of the relationship between domain knowledge and engineering goals comes from Phase 1, where students who expressed engineering or domain knowledge goals in their written evaluations of experiments were as likely to give the same type of response in a subsequent question (i.e., students giving engineering responses to both questions) as they were to give the other type of response (i.e., students giving an engineering response to one question and a domain knowledge response to the other). However, these students generally did not shift from an engineering or domain knowledge goal to a science goal. For example, in Phase 1, of the nine students in L1 who gave engineering or domain knowledge explanations on the first in-class written evaluation, seven (78%) continued to give engineering and domain knowledge responses throughout instruction. Furthermore, three more students in this class expressed engineering or domain knowledge interpretations later in the instruction. Similarly, the one student in L2 who gave an engineering response on the first evaluation question continued giving these explanations throughout the instruction, and one student "regressed" from a science to an engineering goal during the course of the instruction. However, only one of the three students in inquiry-based M1 continued to give engineering/domain knowledge responses during instruction, and no students in this class regressed to engineering goals. This correlation between engineering and domain knowledge goals is not surprising, given that in the vast majority of engineering goal approaches and, by definition, in all "domain knowledge" interpretations, students state or apply their beliefs about the effects of the variables.

These findings also show that, for students not in the inquiry-based science classroom (M1), misconceptions about the nature of the task that surface during CVS instruction persisted throughout instruction and interfered with learning the skills of experimental design.

plausibly misinterpret within an alternate goal were clarified. In addition, the finding of a strong association between science goals and varying the variable(s) one is testing, discussed earlier, informed remediation in the TED tutor. In a recent phase of development, sixth graders from a science and technology magnet school with a diverse student population completed an instructional unit on the TED tutor. After the ramps pretest but prior to the "explicit" instruction portion of the tutor, students identified as not applying science goals or knowledge of the need to contrast the variable(s) one is testing were provided with remedial tutoring, which required them to identify the target variable and select its values across conditions. Students were given immediate feedback on their responses. All students were then asked why it is necessary to vary the target variable then given the explanatory feedback described in the remediation section for R2 failure. This instruction simultaneously enforced the science goal of the activity and its associated contrasting the target variable element. These students showed significantly better transfer performance on an immediate Story posttest than eighth graders from the same school given similar instruction but without the remedial tutoring and wording clarifications.

DISCUSSION

This chapter reports the first in-depth compilation of the errors and misconceptions that arise when students are instructed in experimental design. Our findings of the different errors middle school students make and their misconceptions about experimental design add to those of previous researchers (e.g., Kuhn et al., 1988; Kuhn et al., 1995; Schauble et al., 1991; Tschirgi, 1980) to encompass a wide range of preconception types. On a practical note, the analysis of misconceptions presented here is being used to further inform the development of our adaptive computer-based tutor for simple experimental design (Siler et al., 2009), the TED (Training in Experimental Design) tutor.

Students' motivations underlying their responses can be broadly categorized as employing science goals, engineering goals, or overextended "fairness" goals. Engineering goals—where students attempt to produce desired outcomes—were the most common goal-related misconceptions identified. Most often, students applied the practical engineering goal of maximizing the outcomes of one or both conditions by selecting the variable values they believed would produce the best effect. Somewhat less commonly, students designed setups aimed at producing different outcomes by varying one or more variables they believed to be causal. This goal may be related to (perhaps an implicit) belief that a "good" experiment is one that produces different outcomes, or, relatedly, one that tests a causal variable(s).

Finally, students designed setups intended to produce the same outcomes across conditions. They did this by setting up the conditions exactly the same and by setting only the variables they believed mattered to the same values. They also did this by "counterbalancing" variable settings by using one "good" and one "bad" variable when instructional cues triggered intuitive "fairness" schemas. Such schemas have been identified in students as young as second graders (Wollman, 1977).

Science goals also appear to be tied to the notion of contrasting values of the variable(s) they were testing. However, students often designed experiments that did not allow for valid causal claims, either by intentionally testing multiple variables or by ignoring experimental confounds. Thus, when students adopt science goals, they seem to have most difficulty with controlling variables (and understanding why that is necessary).

Whereas intentions within science goals appear to be largely explicit, as evidenced by students' explicit goal statements such as "I am trying to find out about the [target variable]," engineering goals appear to be largely implicit. That is, when expressing engineering goals, students rarely made analogous explicit statements of intention (e.g., "I am trying to make the balls roll far"). It is plausible that—though students likely have copious experiences applying engineering goals (Schauble et al., 1991), if not accompanied by verbalizations or higher-level thinking, such goals may remain implicit. In contrast, people almost certainly have fewer experiences designing or evaluating experiments (Schauble et al., 1991); however, such experiences may be more likely to be associated with higher-level cognitive processes that result in explicit knowledge. Moreover, students likely hear more explicit statements of science than engineering goals. For example, in their science classes, teachers may say "Today, we are going to do an experiment to find out about X" though may be less likely to say "Today, we are going to try to make the fastest X." The nature of students' "fairness" beliefs are less clear, due to the infrequency with which they were expressed—particularly when explaining their own designs. As with applying engineering goals, students likely have abundant experiences enforcing "fairness" goals, however, these may be more likely to occur in social situations involving verbal negotiations which could contribute to their becoming (or staying) explicit knowledge.

As with the relationship between engineering goals and domain beliefs, engineering goals and domain knowledge "goals," or interpretations of instructional questions, appear to be inter-related. Students who expressed one of these goals during the course of instruction tended to express the other much more often than they expressed science goals. Moreover, that they were more common in the noninquiry science classrooms suggests that domain knowledge and engineering goals may be more likely to be adopted when classroom experiences consist primarily of "learning facts" or "producing results." Because general domain knowledge goals are not likely to be explicitly expressed by the teacher (e.g., by stating "We are going to learn facts," though a teacher might say, "Today, we are going to learn about the layers of the earth"), they may be learned and adopted implicitly.

Furthermore, students generally explicitly expressed their beliefs about variable effects when they applied engineering goals, but rarely did when applying science goals. This suggests that (at least when first learning about experimental design) stu-

the hypothesized "better" value of a variable when they were supposed to "prove" that that particular variable value was responsible for either a good or bad outcome. That is, when given a good outcome, participants were more likely to choose a second instance with the same value than with a different value of the target variable. However, when given an analogous bad outcome, they were more likely to choose a second setup with the opposite value.

Identifying a range of specific errors and misconceptions—especially those that are resistant to typical classroom activities or instruction on experimental design and thus prevent learning—is an important first step in improving instruction in experimental design because it allows for easier identification during instruction. Knowledge of the ways that students can misinterpret instruction can also alert those involved in instructional design to the kinds of wordings that may elicit engineering goals, fairness schemas, or "domain knowledge" misinterpretations of instruction. Designers of instruction may decide to intentionally elicit such misinterpretations in order to address them or to avoid eliciting them. Our knowledge of how students misinterpreted instruction led to revisions in the on-screen text and audio voice-overs in our intelligent tutor. These "low-tech" revisions were not only time- and cost-effective, but were correlated with significant performance improvements.

Furthermore, identifying the types of misconceptions that arise in this context is relevant in devising means of instruction to *remediate* them. As discussed earlier, remedial tutoring in the TED-tutor that required students to identify and vary the target variable prior to the "explicit" instruction on the rationale for controlling the other variables was associated with gains in transfer performance. We believe this remediation elicited science goals; however, it may not have alerted students of their alternative goal interpretations. Other methods, which do elicit this awareness, may lead to even stronger gains. For example, alternative goals may be directly contradicted when detected in order to both make students' goal assumptions explicit and induce cognitive conflict. And induced cognitive conflict that elicits such cognitive processes as knowledge building may promote conceptual change (Chan, Burtis, & Bereiter, 1997) such as goal shifting. The use of refutational texts designed to directly contradict students' misconceptions has been shown to promote conceptual change in such domains as physics (e.g., Diakidoy, Kendeou, & Ioannides, 2002; Hynd, McWhorter, Phares, & Suttles, 1994), biology (e.g., Mikkilä-Erdmann, 2001), and ecology (e.g., Ozkan, Tekkaya, & Geban, 2004). However, to our knowledge, this method has not targeted goal-related misconceptions.

To apply this remedial method in the TED tutor, when students are asked to design experiments to test a particular variable, they will first be asked to select their goal in setting up the experiment, thereby making explicit any implicit goals. If students select a response indicating an engineering goal, such as "to make the balls roll far," they will receive immediate feedback, such as "Actually, the question is asking you to design an experiment that will let you figure out if the surface affects how far balls roll, not to make the balls roll far." If students persist in selecting responses indicating

Finally, identification of these goal misconceptions provides information about possible alternative instructional approaches. For example, students who fail to learn CVS via the "explicit" instruction may fare better with instruction that builds upon their intuitive ideas, such as their understandings of "fairness" of comparisons in combination with their understandings of "comparing and contrasting" a target variable. The relative effectiveness of such instructional methods will be investigated in future studies.

However, we believe that even if these methods are effective in helping students of all backgrounds to develop robust understandings of experimental design, engaging students in scientific investigations on a regular basis is still of vital importance. Such "mindful" experiences may not only reinforce understanding the procedures and logic of CVS, but also the development of other inquiry skills, better enabling students to explore and learn about their world.

ACKNOWLEDGMENTS

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 the project. We also thank the participating schools and teachers who allowed us to conduct research in their classrooms.

REFERENCES

Bindra, D., Clarke, K., & Schultz, T. (1980). Understanding predictive relations of necessity and sufficiency in formally equivalent "causal" and "logical" problems. *Journal of Experimental Psychology: General, 109,* 422–443.

Braine, M. D., & Rumain, B. (1983). Logical reasoning. In J. H. Flavell & E. M. Markman (Eds.), *Cognitive development: Handbook of child psychology* (Vol. 3, pp. 261–340). New York: John Wiley.

Byrnes, J., & Overton, W. (1986). Reasoning about certainty and uncertainty in concrete, causal, and propositional contexts. *Developmental Psychology, 22,* 793–799.

Chan, C., Burtis, P., & Bereiter, C. (1997). Knowledge building as a mediator of conflict in conceptual change, *Cognition and Instruction, 15,* 1–40.

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

Chinn, C. A., & Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of Educational Research, 63,* 1–49.

Clement, J. (1983). A conceptual model discussed by Galileo and used intuitively by physics students. In D. Gentner & A. Stevens (Eds.), *Mental models* (pp. 345–362). Hillsdale, NJ: Lawrence Erlbaum.

Diakidoy, I. Kendeou, P. & Ioannides, C. (2002). Reading about energy: The effects of

diSessa, A. (1993). Towards an epistemology of physics. *Cognition and Instruction, 10*(2–3), 105–225.

Fay, A. L., & Klahr, D. (1996). Knowing about guessing and guessing about knowing: Preschoolers' understanding of indeterminacy. *Child Development, 67,* 689–716.

Grosslight, L., Unger, C., Jay, E., & Smith, C. (1991). Understanding models and their use in science: Conceptions of middle and high school students and experts. *Journal of Research in Science Teaching, 28,* 799–822.

Hynd, C., McWhorter, Y., Phares, V., & Suttles, W. (1994). The role of instructional variables in conceptual change in high school physics. *Journal of Research in Science Teaching, 32,* 933–946.

Klahr, D. (2009). "To every thing there is a season, and a time to every purpose under the heavens": What about direct instruction? In S. Tobias & T. M. Duffy (Eds.), *Constructivist theory applied to instruction: Success or failure?* (pp. 291–310). London: Taylor & Francis.

Klahr, D., Fay, A. L., & Dunbar, K. (1993). Developmental differences in experimental heuristics. *Cognitive Psychology, 25,* 111–146.

Klahr, D., & Li, J. (2005). Cognitive research and elementary science instruction: From laboratory, to the classroom, and back. *Journal of Science Education and Technology, 14,* 217–238.

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

Klahr, D., Triona, L., Strand-Cary, M., & Siler, S. (2008). Virtual vs. physical materials in early science instruction: Transitioning to an autonomous tutor for experimental design. In J. Zumbach, N. Schwartz, T. Seufert, & L. Kester (Eds.), *Beyond knowledge: The legacy of competence meaningful computer-based learning environments.* (pp. 163–172). New York: Springer.

Kuhn, D. (1991). *The skills of argument.* New York: Cambridge University Press.

Kuhn, D., Amsel, E., & O'Loughlin, M. (1988). *The development of scientific thinking skills.* New York: Academic Press.

Kuhn, D., & Dean, D. (2005). Is developing scientific thinking all about learning to control variables? *Psychological Science, 16,* 866–870.

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), 1–157.

Masnick, A. M., & Klahr, D. (2003). Error matters: An initial exploration of elementary school children's understanding of experimental error. *Journal of Cognition & Development, 4,* 67–98.

McCloskey, M. (1983). Naive theories of motion. In D. Gentner & A. L. Stevens (Eds.), *Mental models* (pp. 299–324). Hillsdale, NJ: Lawrence Erlbaum.

Mikkilä-Erdmann, M. (2001). Improving conceptual change concerning photosynthesis through text design. *Learning and Instruction, 11,* 241–257.

Minstrell, J. (1982). Explaining the "at rest" condition of an object. *Physics Teacher, 20,* 10–20.

Morris, B. J., & Sloutsky, V. M. (2002). Children's solutions of logical versus empirical problems: What's missing and what develops? *Cognitive Development, 16,* 907–928.

Ozkan, O., Tekkaya, C., & Geban, O. (2004). Facilitating conceptual change in students'

Pieraut-Le Bonniec, G. (1980). *The development of modal reasoning: The genesis of necessity and possibility notions*. New York: Academic Press.

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. (1981). Developmental sequences within and between concepts. *Society for Research in Child Development Monographs, 46* (Whole No. 189).

Siegler, R. S. (1996). *Emerging minds: The process of change in children's thinking*. New York: Oxford University Press.

Siler, S. A., Klahr, D., Strand-Cary, M. S., Magaro, C., & Willows, K., (2009). *Adapting an effective lesson plan for a computer-based tutor*. Paper presented at the 13th Biennial Conference for the European Association for Research on Learning and Instruction (EARLI), Amsterdam, Netherlands.

Slotta, J. D., Chi, M. T. H., & Joram, E. (1995). Assessing students' misclassifications of physics concepts: An ontological basis for conceptual change. *Cognition and Instruction, 13*, 373–400.

Smith, C., Carey, S., & Wiser, M. (1985). On differentiation: A case study of the development of size, weight, and density. *Cognition, 21*, 177–237.

Stathopoulou, C., & Vosniadou, S. (2007). Conceptual change in physics and physics related epistemological beliefs: A relationship under scrutiny. In S. Vosniadou, A. Baltas, & X. Vamvakoussi (Eds.), *Re-framing the conceptual change approach in learning and instruction*, Advances in Learning and Instruction Series (pp. 145–165). Oxford: Elsevier.

Strand-Cary, M., & Klahr, D. (2008). Developing elementary science skills: Instructional effectiveness and path-independence. *Cognitive Development, 23*, 488–511.

Toth, E. E., Klahr, D., & Chen, Z. (2000). Bridging research and practice: A cognitively-based classroom intervention for teaching experimentation skills to elementary school children. *Cognition & Instruction, 18*, 423–459.

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

Vosniadou, S. (1994). Universal and culture-specific properties of children's mental models of the earth. In L. Hirschfeld & S. Gelman (Eds.), *Mapping the mind* (pp. 412–430). New York: Cambridge University Press.

Vosniadou, S., & Brewer, W. F. (1992). Mental models of the earth: A study of conceptual change in childhood. *Cognitive Psychology, 24*, 535–585.

Vosniadou, S., & Brewer, W. F. (1994). Mental models of the day/night cycle. *Cognitive Science, 18*, 123–184.

Vosniadou, S., Ioannides, C., Dimitrakopoulou, A., & Papademetriou, E. (2001). Designing learning environments to promote conceptual change in science. *Learning and Instruction, 11*, 381–419.

Wiser, M., & Carey, S. (1983). When heat and temperature were one. In D. Genter & A. Stevens (Eds.), *Mental models* (pp. 267–297). Hillsdale, NJ: Erlbaum.

Wollman, W. (1977). Controlling variables: Assessing levels of understanding. *Science Education, 61*, 371–383.

Zohar, A., & Nemet, F. (2001). Fostering students' knowledge and argumentation skills through dilemmas in human genetics. *Journal of Research in Science Teaching, 39*,

PART THREE IMPLICIT AND EXPLICIT PROCESSES IN THE COGNITIVE PSYCHOLOGY OF SCIENCE