

COGNITION AND INSTRUCTION

Twenty-Five Years of Progress

Edited by

Sharon M. Carver

David Klahr

Carnegie Mellon University



2001

LAWRENCE ERLBAUM ASSOCIATES, PUBLISHERS

Mahwah, New Jersey

London

3

Cognitive Development and Science Education: Ships that Pass in the Night or Beacons of Mutual Illumination?

David Klahr

Zhe Chen

Eva E. Toth

Carnegie Mellon University

Two beliefs widely shared among the contributors to this volume are that (a) theoretical and empirical advances in cognitive and developmental psychology can provide a solid basis for improved instructional practice; and (b) the challenge of instructional innovation can raise new questions for basic cognitive research. Evidence supporting the first belief—implicit in the large dose of cognitive and developmental psychology contained in most degree programs in education—comes from the type of chapters in this and related volumes (e.g., Bruer, 1993, McGilly, 1994), as well as the articles appearing in two influential interdisciplinary journals—*Journal of the Learning Sciences* and *Cognition and Instruction*—that have appeared since the first Carnegie Symposium on this topic (Klahr, 1976). The second belief is a particularization of the commonly held view that applied work always raises novel questions to be addressed by further basic research. In

this case, the applications involve the development and implementation of effective instructional methods and the basic research is carried out in the psychologist's laboratory. Taken together, these two beliefs support the "mutual illumination" metaphor in the title of this chapter.

However, there is another view—one that supports instead the second metaphor in our title, in which the two enterprises of cognitive research and instructional practice have no more in common than "ships passing in the night." Consider, for example, the somewhat pessimistic assessment by Strauss (1998) that appeared recently in a prestigious handbook on applied topics in developmental psychology (Sigel & Renninger, 1998). Although Strauss acknowledged that one can find, at the margins of both fields, several atypical examples of such mutual influence, he argued that an honest look at the bulk of the work published in each field reveals that "cognitive developmental psychologists rarely involve themselves in topics that are of interest to science educators" (Strauss, 1998, p. 358). In other words, perhaps the education ship and the research ship traverse the same seas and visit the same ports, but they pass in darkness, with neither one illuminating the other.

Strauss offered several explanations for the relatively small proportion of published work that is of interest to both cognitive psychologists and science educators. One problem is lack of common interest in content: Developmentalists often study topics that, while providing useful indices for cognitive development, may have little relevance for science education. Another problem is that developmentalists focus on universal and invariant sequences that may be largely irrelevant to educators who are more interested in what can, rather than what can't, be changed. A third problem is the tendency for researchers in cognitive development to study the child in isolation, whereas educators have to work in complex social and institutional settings in which cognitive processes may account for only a small part of the variance in outcomes. Finally, Strauss argued that there is scant shared knowledge between developmentalists and science educators: The former know a lot about children, but little about topics in the nonpsychological sciences, whereas the latter know a lot about their science, but little about the psychology of thinking, learning, and development.

Unfortunately, there is much that is correct about Strauss's gloomy assessment. Except for a few notable exceptions (e.g., Brown, 1992, 1997; Fennema, Carpenter, Franke, Levi, Jacobs, & Empson, 1996; White & Frederiksen, 1998), most of the research in the intersection between cog-

niton and instruction is carried out by researchers whose predilection is to conduct their work in either the psychology laboratory or the classroom, but not both. Consequently, reports of laboratory-based research having clear instructional implications typically conclude with a suggested instructional innovation, but one rarely finds a subsequent report on an associated specific action resulting in instructional change. Similarly, many instructional interventions are based on theoretical positions that have been shaped by laboratory findings, but the lab procedures have been adapted to the pragmatics of the classroom by a different set of researchers (e.g., Christensen & Cooper, 1991; Das-Smaal, Klapwijk, & van der Leij, 1996). This division of labor between laboratory-based cognitive research and classroom research is understandable, but, in our view, unnecessary and inefficient because much can be lost in the translation from the psychology laboratory to the classroom.

In this chapter, we propose a two-part remedy to this situation. The first part provides a conceptual framework for classifying research on children's scientific thinking along lines that are relevant to science education. We hope that by providing a kind of "reader's guide" to some of the basic research on the development of scientific reasoning, we may clarify its relevance to science education while at the same time providing some insight into why such work is not always immediately embraced by those facing the challenge of improving instruction. The second part of our chapter provides a counterexample to Strauss's claim: We offer a concrete instance of a productive two-way flow between the psychology lab and the science classroom. The example is based on a project in which, over the past several years, we have been developing, implementing, and assessing a set of instructional materials for teaching children in grades two to four the concepts and skills associated with the design of unfounded experiments and the derivation of valid inferences.

A TAXONOMY OF APPROACHES TO STUDYING AND TEACHING SCIENTIFIC THINKING

Scientific reasoning—both as it is studied by developmental psychologists and as it is taught by elementary school science teachers—can be classified along two dimensions: one representing the degree of domain specificity or domain generality, and the other representing the type of

discovery processes involved, such as generating hypotheses, designing experiments, and evaluating evidence (see Table 3.1). During the course of normal scientific discovery, the various cells in Table 3.1 are traversed repeatedly. However, it is very difficult to study thinking processes that involve all of them simultaneously. Consequently, much of the research on scientific thinking has been intentionally designed to focus on only one or two of the cells in Table 3.1, although some studies have used complex contexts involving several cells. The entries in Table 3.2 illustrate some of the ways that psychologists have attempted to study different aspects of scientific thinking in isolation. (For a more complete description, see Klahr & Carver, 1995, and Klahr, 2000.)

Integrative Investigations of Scientific Reasoning

The types of investigations spanning the bottom row in Table 3.1 and summarized as the final entry in Table 3.2 reveal the large grain of truth in Strauss's complaint about the perceived irrelevance of psychological research to science instruction. On the one hand, the Bruner concept formation task and the Wason "2-4-6" task are among the most widely cited and replicated studies in the cognitive psychology literature. On the other hand, few science teachers would deem it worthwhile to teach children these kinds of skills or to use these puzzlelike materials in their classrooms, even though, from a psychologist's perspective, they elegantly illustrate some of the fundamental cognitive processes involved in scientific thinking.

The classroom teacher does not have the laboratory psychologist's luxury of isolating different components of scientific thinking in order to better understand them. Instead, the teacher must attempt to orchestrate all of these aspects in various combinations. For example, consider the complexity faced by a teacher attempting to teach her students the classic

TABLE 3.1
Types of Foci in Investigations of Scientific Reasoning Processes

	<i>Generating Hypotheses</i>	<i>Designing & Executing Experiments</i>	<i>Evaluating Evidence</i>
Domain-specific knowledge	A	B	C
Domain-general knowledge	D	E	F

TABLE 3.2
Examples of Investigations Located in Various Cells of Table 3.1

<i>Cell(s) from Table 3.1</i>	<i>Focus of Study</i>	<i>Reference</i>
A	Domain-specific hypothesis generation. Participants are asked to make predictions or give explanations in a specific domain in order to reveal their intuitive theories of mechanical or biological phenomena. They are not allowed to run experiments, and they are not presented with any evidence to evaluate.	Carey, 1985; McCloskey, 1983.
B	Domain-specific experimental design. Participants are asked to decide which of a set of prespecified experiments will provide the most informative test of a prespecified hypothesis. There is no search of the hypothesis space and the experiment space search is limited to choosing from among the given experiments, rather than generating them.	Tschirgi, 1980.
E	Domain-general experimental design. People are asked to design factorial experiments in relatively sparse contexts. The use of domain-specific knowledge is minimized as is search in the hypothesis space and the evidence evaluation process.	Case, 1974; Kuhn & Angelev, 1976; Siegler & Liebert, 1975.
C & F	Domain-specific and domain-general evidence evaluation. Studies in this category focus on people's ability to decide which of several hypotheses is supported by evidence. Typically, participants are presented tables of covariation data, and asked to decide which of several hypotheses is supported or refuted by the data. In some cases, the factors are abstract and arbitrary—in which case we would classify the studies in Cell F—and in others, they refer to real world factors, such as studies that present data on plant growth in the context of different amounts of sunlight and water.	Amsel & Brock, 1996; Bullock, Ziegler, & Martin, 1993; Ruffman, Perner, Olson, & Doherty, 1993; Shaklee & Paszek, 1985.
A & C	Domain-specific hypothesis generation and evidence evaluation. Children are asked to integrate a variety of forms of existing evidence in order to produce a theory that is consistent with that evidence. They do not have the opportunity to generate new evidence via experimentation, and the context of their search in the hypothesis space is highly domain specific.	Vosnradou & Brewer, 1992.
A, C, & F	Domain-specific hypothesis generation and domain-specific and domain-general evidence evaluation. In these studies, participants are presented with a complex mixture of covariation data, possible causal mechanisms, analogous effects, sampling procedures, and alternative hypotheses from which they are asked to make a decision about a potentially causal factor. People are given the opportunity to go beyond just the covariation data—that is, to use both their domain-specific knowledge as well as other domain-general features, such as sample size, in making their decisions.	Koslowski, 1996; Koslowski & Okagaki, 1986; Koslowski, Okagaki, Lorenz, & Umbach, 1989.

(Continued)

TABLE 3.2
(continued)

Cell(s) from Table 3.1	Focus of Study	Reference
D, E, & F	<p>Domain-general hypothesis generation, experimental design, and evidence evaluation.</p> <p>Participants are asked to discover an arbitrary rule or concept based on formal properties of the stimulus. No domain-specific knowledge of any kind is required, but participants have to use domain-general reasoning processes such as hypothesis formation, instance selection, and rule induction.</p>	<p>Bruner, Goodnow, & Austin, 1956; Wason, 1960.</p>

problem in mechanics of discovering the period of a pendulum. As illustrated in Table 3.3, her students would traverse all of the cells as they worked with this problem. Even though the instruction would tend to focus on the domain-specific aspects of force and acceleration that underlie the phenomenon being investigated, the teacher would also attempt to convey some important domain-general processes and knowledge about scientific methodology. Thus, if they are to be of relevance to educators, psychological studies must somehow be more representative of the complexity faced by the teacher.

First, they must cross the row boundaries in Table 3.1 in order to study the interaction between domain-specific and domain-general knowledge. Second, they must integrate the processes of hypothesis search, experimentation, and evidence evaluation in order to examine their mutual influence. In recent years, several investigators have begun to address these questions by integrating the six different aspects of the scientific discovery process represented in Table 3.1 while still posing the research questions at a sufficiently fine grain so as not to lose relevant detail about the discovery process (cf. Dunbar, 1993; Klahr, 2000; Klahr & Dunbar, 1988; Klahr, Fay, & Dunbar, 1993; Kuhn, 1989; Kuhn, Amsel, & O'Loughlin, 1988; Kuhn, Schauble, & Garcia-Mila, 1992; Kuhn, Garcia-Mila, Zohar, & Andersen, 1995; Schauble, 1990; Schauble, Glaser, Raghavan, & Reiner, 1991).

The study that we describe later in this chapter also focused primarily on a domain-general skill—or what is usually called a “process skill” (in contrast to “content knowledge”). The particular skill had to do with how to design unconfounded experiments, and our study can therefore be classified as belonging primarily in Cell E. However, as will become evident, our experiments also involved evaluation of real experiments with real devices in the physical world, and thus were experiments “about” something. So we would also implicate Cells B, C, and F.

TABLE 3.3
Cells Traversed During Typical Elementary School Science Lab Sessions
on Finding the Period of a Pendulum

	Generating Hypotheses	Designing & Executing Experiments	Evaluating Evidence
Domain-specific knowledge	<ul style="list-style-type: none"> • Length? • Initial Force? • Mass? 	<ul style="list-style-type: none"> • Selecting and isolating some aspect (length, mass, force, etc.) • Counting cycles of pendulum • Establishing a timing basis 	<ul style="list-style-type: none"> • Averaging and comparing several trials of same setup • Cross-setup comparisons to look for differences in period length • Eliminate noncasual factors (mass, initial height, initial force, etc.)
Domain-general knowledge	<ul style="list-style-type: none"> • Asking “good” questions • Proposing plausible causal mechanisms • Inducing “rules” from regularities 	<ul style="list-style-type: none"> • Varying one thing at a time • Choosing tractable values for variables • Minimizing error • Observing relevant outcomes 	<ul style="list-style-type: none"> • Recording data • Making tables • Distinguishing determinate from indeterminate data patterns • Finding most representative measures

To summarize, we have briefly described a taxonomy in which scientific reasoning is classified along two dimensions—domain specificity and type of processes—and we have attempted to illustrate how this classification can be useful for understanding and characterizing both basic lab investigations and classroom teaching. One problem that this taxonomic exercise has revealed is that although science education aims to impart domain-general scientific reasoning skills, it is almost always couched in—perhaps even overwhelmed by—specific context, whereas lab research, although unambiguously identifying general reasoning components, often fails to indicate its relevance to classroom practice.

In the next section, we offer an example of how lab research can generate a solid basis for classroom research, which in turn can generate new theoretical issues for further study in the lab. We describe the process whereby we translated a theoretically motivated, carefully crafted, and laboratory-based instructional procedure of proven effectiveness into a

classroom intervention, making minimal modifications to the instructional components while adapting to the constraints of a real classroom. The research-to-practice interface described here further supports both of the widely held beliefs cited in the opening paragraphs of this chapter. First, instruction based on prior laboratory research was educationally effective. Children learned and transferred what they had been taught. Second, once the instruction was situated in a real classroom, a new set of basic research issues arose, and they are currently under investigation. Because we view the move from the laboratory-based research environment to the classroom as fraught with potential pitfalls, we took a very small step—a “baby step” perhaps—in making this move. Nevertheless, or perhaps consequently, we were able to devise an effective curriculum unit that maintained what we believe to be the fundamental features of the laboratory instruction, while still being consistent with everyday classroom practice.

The rest of this chapter is organized as follows. First, we describe the topic of the instruction—the design of controlled experiments—and its place in the elementary school science curriculum. Then, we briefly introduce a contentious issue in instructional methodology, the use of direct instruction versus discovery learning. Next, we summarize the laboratory training study that led us to use direct instruction as the basis for our classroom intervention. With this as background, we describe the design and implementation of the classroom study that aimed to verify the laboratory findings in classroom situations, followed by the basic findings of this study. Finally, we revisit the issue of the mutual influence and relevance of the fields of cognition and instruction.

Before we embark, some terminological clarification might be helpful. Throughout this chapter we use “lab study” when referring to the type of one-on-one study that is typical of the research psychologist—exemplified by the first study described in this chapter. By “classroom study” we mean the kind of study described in the second part of the chapter, where a teacher introduces an experimental curriculum unit and we do several assessments of its effectiveness. The terminology can get confusing because our lab study, although carried out one-on-one with an experimenter and a child, was actually conducted (in a quiet room) in the school and our classroom study took place in the normal science lab in the school. The one additional complexity is that immediately before and after the classroom study, we assessed some children in a one-on-one lab fashion in order to compare their performance to the earlier (true) lab study and to calibrate the lab assessments with the classroom assessments.

DESIGNING UNCONFOUNDED EXPERIMENTS: THE CONTROL OF VARIABLES STRATEGY

There is widespread agreement among science educators that “Even at the earliest grade levels, students should learn what constitutes evidence and judge the merits or strength of the data and information that will be used to make explanations” (NSES, 1995). But evidence does not spring forth unbidden. Instead, it must be actively sought or generated. Thus, the ability to create informative experiments and to derive valid inferences from the evidence they yield is one of the fundamental design skills underlying scientific thinking (Klahr, 2000).

A central component of this skill is the control of variables strategy (CVS). Procedurally, CVS is a method for creating experiments in which a single contrast is made between experimental conditions. The full strategy involves not only creating such contrasts, but also being able to distinguish between confounded and unconfounded experiments. Conceptually, CVS involves making appropriate inferences from the outcomes of unconfounded experiments as well as understanding the inherent indeterminacy of confounded experiments.

Both the educational and the psychological literature suggest that elementary school children find these concepts and procedures extremely difficult to acquire. Ross’s (1988) meta-analysis of over 60 CVS training studies from the 1970s and 1980s indicated that a variety of training methods can generate improvement in CVS performance, but only a handful of the studies in his sample included young elementary school children (i.e., below grade five). The results of those few studies, as well as more recent ones in that age range, present a decidedly mixed picture of the extent to which young elementary school children can understand and execute CVS (Bullock & Ziegler, 1999; Case, 1974; Kuhn, et al., 1995; Kuhn & Angelev, 1976; Schauble, 1996). Moreover, for those studies showing statistically significant differences between trained and untrained groups,¹ the absolute levels of posttest performance are well below educationally desirable levels. Indeed, to get ahead of our story a bit, our first study (Chen & Klahr, 1999) showed that even in schools with strong elementary science programs in which components of CVS were taught repeatedly during the early science curriculum, fourth graders could correctly construct unconfounded experiments on fewer than 50% of their attempts.

¹Ross found a mean effect size of .73 across all of the studies in his sample.

THEORIES OF INSTRUCTION, LEARNING, AND TRANSFER

Given that CVS is a fundamental scientific reasoning skill and given that few elementary school children master it even after several years of good science instruction, it is important to know whether there are effective ways to teach it and whether age and instructional method interact with respect to learning and transfer. One controversial issue in instruction is whether or not discovery learning is more effective than the traditional didactic method (called here simply “direct instruction”). Part of the controversy derives from a lack of definitional consensus, so we need to clarify our use of the terms. Although the details will become apparent when we describe our studies, it is important to note at the outset that we do not associate one with “active” and the other with “passive” learning. In all of the learning situations described in this chapter, students were actively engaged in the design and manipulation of experimental apparatus. The main distinction between the situations is that in direct instruction, the instructor told the students how and why CVS worked, whereas in other situations there was no such direct telling.

Even with these distinctions, the relative efficacy of discovery learning versus direct instruction depends on many factors, one of which is the content of the learning tasks. Discovery learning has been considered an effective approach for the acquisition of domain-specific knowledge. Its advocates argue that children who are actively engaged in acquiring new knowledge are more likely to be successful in retaining and applying it than children who passively receive direct instruction (e.g., Jacoby, 1978; McDaniel & Schlager, 1990). Although discovery learning might be effective when problem outcomes provide informative feedback (e.g., Siegler, 1976), direct instruction may be appropriate in those cases where it is unlikely that a multistep strategy would be discovered spontaneously. For example, Klahr and Carver (1988) found that a brief period of direct instruction in how to debug computer programs was more effective than hundreds of hours of discovery learning. Here, too, both groups of children were active, that is, they were writing and running computer programs. But one group was told how to debug, and the other was not. With respect to CVS, unguided experimental designs typically do not provide informative feedback concerning their quality. This lack of feedback might render the discovery of procedures such as CVS particularly difficult for early elementary school children.

BACKGROUND: A LABORATORY TRAINING STUDY

It is clear that the issue of the relative effectiveness of direct instruction versus discovery learning is extremely complex (and, unfortunately, somewhat politicized). Rather than examine the issue in the “messy” context of an ongoing classroom, we decided to begin by studying it in the relatively controlled confines of a laboratory study. Thus, we compared the effectiveness of different instructional methods for teaching CVS in a situation where children had extensive and repeated opportunities to use CVS while designing, running, and evaluating their own experiments.

Materials

We used three different domains in which children had to design unconfounded experiments: (a) springs, in which the goal was to determine the factors that affected spring elongation; (b) sinking, in which children had to assess the factors that determined how fast various objects sank in water; and (c) ramps, in which children needed to figure out which factors determined how far a ball rolled down the slope. The underlying CVS logic in all three domains was identical. In each, there were four variables that could assume either of two values. In each domain, children were asked to focus on a single outcome that could be affected by any or all of the four variables.

For example, in the springs domain,² children had to make comparisons to determine the effects of different variables on how far springs stretch. Materials consisted of eight springs varying in length (long and short), coil width (wide and narrow), and wire diameter (thick and thin). The springs were arranged on a tray such that no pair of adjacent springs made an unconfounded comparison. A pair of heavy and a pair of light weights were also used. Heavy and light weights differed in shape as well as in weight, so that they could be easily distinguished. To set up a comparison, children selected two springs to compare and hung them on hooks on a frame and then selected a weight to hang on each spring. To execute a comparison, participants hung the weights on the springs and observed as the springs stretched. The outcome measured was how far the springs stretched down toward the base of the frame. Figure 3.1 depicts the materials and an experiment from the spring domain.

²In this chapter, we describe only the springs in detail. See Chen and Klahr (1999) for a detailed description of all three domains.

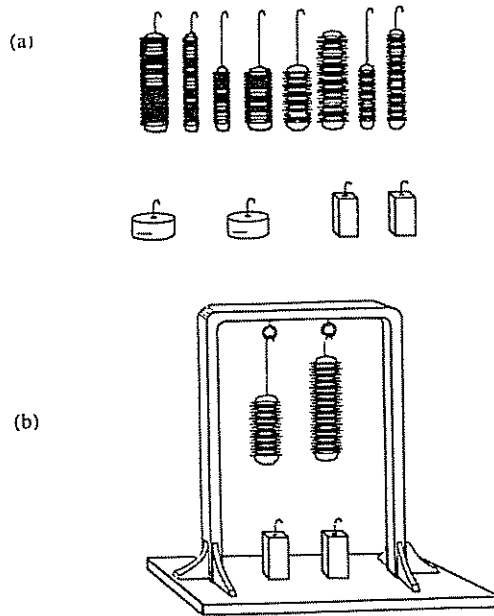


FIG. 3.1. The springs domain. (a) Set of eight springs varying in wire thickness, spring width, and spring length, and set of two heavy weights (cylinders) and light weights (cubes); (b) An unconfounded experiment in which length is varied, and all other factors are held constant.

Training Conditions

1. Explicit Training. Explicit training was provided in the *training/probe* condition. It included an explanation of the rationale behind controlling variables as well as examples of how to make unconfounded comparisons. Children in this condition also received probe questions before and after each comparison that they made. Before the experiment was executed, children were asked to explain and justify the design. After the experiment was executed, children were asked if they could “tell for sure” whether the variable they were testing made a difference and also why they were sure or not sure. The explicit instruction was provided following the exploration phase (see Procedure section below) in which children had designed a few experiments and pondered probe questions about those experiments.

2. Implicit Training. Implicit training was provided in the *no training/probe* condition. Here, children did not receive direct instruction, but—as in the explicit training condition—they did actively construct experiments and receive probe questions before and after each of them.

3. Discovery Learning. Discovery learning opportunities were provided to children in the *no training/no probe* condition. They received neither training nor probes but they did receive the same number of opportunities as children in the other conditions to actively construct experiments.

PARTICIPANTS, PROCEDURE, AND MEASURES USED IN THE LABORATORY STUDY

Eighty-seven second, third, and fourth graders from two private schools³ in an urban area were randomly assigned to one of the three different instructional methods. Each child worked with one of the three domains on their first day in the study (exploration and assessment phases) and then with two other domains on their second day (transfer-1 and transfer-2). Domain order was counterbalanced, as was the order of focal variables within each domain.

Procedure

Part I consisted of four phases: exploration, assessment, transfer-1, and transfer-2 (see Table 3.4). In each phase, children were asked to construct four different experimental contrasts from which they could make a valid inference about the causal status of some dimension of the domain. The exploration phase established an initial baseline of children’s ability to design unconfounded experiments in the first domain (e.g., springs). For the *training/probe* condition, the instructional session immediately followed the exploration phase. Then followed the assessment phase in which children were asked to design experiments on a different dimension but in the same domain. (Thus, if, in the exploration phase, the experiments focused on spring *length*, then the assessment phase would focus on

³In School A, we used only third and fourth graders.

TABLE 3.4
Time Line for Laboratory Study

Time	Part I			Part II	
	First day	A few days later	7 months later		
Phase	Exploration	Transfer-1	Transfer-2	Late training	Far transfer
Purpose	Establish base-line of children's ability to design unconfounded experiments in Domain A (e.g., springs)	Determine whether children can design an unconfounded test in Domain B (e.g., sinking objects)	Determine whether children can design an unconfounded test in Domain C (e.g., ramps)	Train the two "no training" groups in preparation for far transfer test	Determine whether children can identify good and bad experimental designs on paper and pencil test
Domain	A (springs)	B (sinking)	C (ramps)	C (ramps)	Nonphysical: cooking, plant growth, foot race, etc
	Training or additional exposure	Assessment			
	Explicit training (for training group only) or Additional experiments (for no-training groups) (Factor a; e.g., spring length)	Determine whether children can design an unconfounded test of Factor b (e.g., spring width)			

spring *width*.) Transfer-1 and transfer-2 took place a few days after exploration and assessment. Children returned to the lab and were asked to design unconfounded experiments in the other two domains (e.g., in the current example, they would do experiments with ramps and with sinking objects).⁴

Part II was a paper and pencil, experiment evaluation posttest, given 7 months after the individual interviews. This consisted of a set of 15 pairwise experimental comparisons in a variety of domains. The child's task was to examine the experimental setup and decide whether it was a good or a bad experiment. (This type of assessment was also used in the classroom study, and it is described in more detail later.)

RESULTS FROM THE LABORATORY TRAINING STUDY

Measures

Three measures used in the lab study that were also used in the classroom study⁵ were (a) *CVS score*—a simple performance measure based on children's use of CVS in designing experimental contrasts; (b) *robust use of CVS*—a more stringent measure based on both performance and verbal justifications (in responses to probes) about why children designed their experiments as they did; (c) *domain knowledge*—based on children's responses to questions about the effects of different causal variables in the domain.

CVS Score. Children's use of CVS was indexed by their use of valid experimental designs. For example, a valid design to test the effect of the length of a spring is that the pair differs only in the focal variable (i.e., length) while all other variables (coil width, wire size, and weight) are kept constant. Invalid designs included (a) noncontrastive comparisons in which the focal variable was not varied and one or more other variables were varied, and (b) confounded comparisons in which the focal variable as well as one or more other variables were varied. Each valid design was given a score of 1. All other types of design were given a score of 0.

⁴Following the transfer-1 and transfer-2 phases, children in the two no-training groups in School A were trained, in preparation for the far transfer test in Part II.

⁵A fourth measure—strategy similarity awareness—was based on children's responses to questions about the similarity across tasks. This is described in Chen and Klahr (1999).

Because children made four comparisons in each phase, the CVS scores for each phase could range from 0 to 4.

Robust Use of CVS. Children's responses to the probe questions "Why did you set up the comparison this way?" and "Can you tell for sure from this comparison?" were classified into four categories: (1) Explanations that included mentions of CVS (e.g., "You just need to make the surface different, but put the gates in the same places, set the ramps the same height, and use the same kind of balls"); (2) Explanations that included controlling some but not all of the other relevant variables (e.g., "Because they're both metal but one was round and one was square"); (3) Explanations that mentioned a comparison within the focal variable (e.g., "Because I had to make the surfaces different"); and (4) Explanations that were irrelevant to CVS.

Children received a robust CVS score of 1 only for those trials for which they produced an unconfounded design *and* provided an explanation or interpretation that mentioned the control of all other variables (i.e., a response fitting category 1, above). Other trials received a score of 0. Again, because children made four designs in each phase, the range of robust use scores was 0 to 4.

Domain Knowledge. Domain knowledge was assessed by asking children, both before and after they designed and implemented their tests, how they thought each variable would affect the outcome. Children's correct prediction/judgment of each variable was given a score of 1, and for incorrect prediction/judgment, a score of 0 was assigned.

Initial Performance in Using CVS

Children's initial performance was measured by the proportion of unconfounded comparisons they produced during the exploration phase. We found significant⁶ grade differences in this initial performance with 26%, 34%, and 48% in second, third, and fourth grade, respectively. Note that, even for second graders, these scores are significantly above chance.⁷

⁶Statistical detail has been suppressed throughout this chapter. Most of the effects reported as "significant" have *p* values less than .01, although a few are only .05, whereas "marginally significant" values are between .05 and .10. For more detail, see Chen and Klahr (1999) and Toth, Klahr, and Chen (2000).

⁷The chance probability of producing an unconfounded comparison is .083. See Chen and Klahr (1999) for a detailed explanation.

Thus, although continued exposure to science classes in each grade does lead to improvement in children's ability to design unconfounded experiments, their overall performance is far below ceiling.

Acquisition and Transfer of CVS

The three training conditions differed substantially in their effects. As indicated in Fig. 3.2, the frequency of CVS use in the training/probe condition increased immediately following training, and remained at a relatively high level. In contrast, for the no training conditions, the increase was slow (for no training/probe) and unsustainable (for no training/no probe). Statistical analysis revealed that, when averaged over all three grade levels, the only significant gains occurred in the training/probe condition.

A more detailed analysis, in which we looked at each grade level separately, revealed that only the third and fourth graders in the training/probe condition showed significant gains after training that were

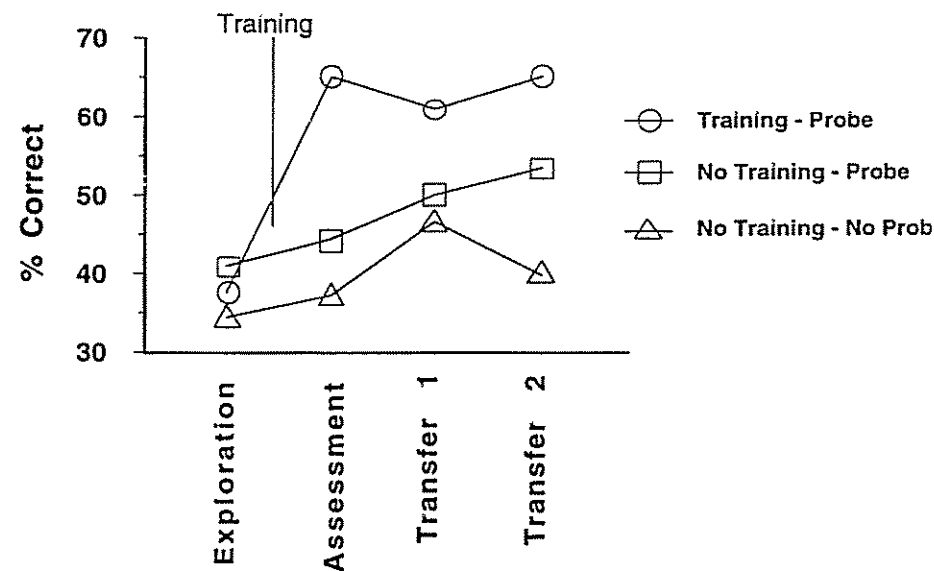


FIG. 3.2. Percentage of trials with correct use of CVS by phase and condition (lab study).

maintained into the transfer phases (see Fig. 3.3). For second graders in the training/probe condition, transfer performance was not significantly higher than the initial exploration performance.

In order to assess transfer in *individual* students, we defined a "good experimenter" as a child who produced at least 7 out of 8 unconfounded comparisons during transfer-1 and transfer-2, and then we computed the proportion of children who became good experimenters between exploration and transfer. There were substantial effects of condition: 44% of the children in the training/probe condition, 22% in the no training/probe

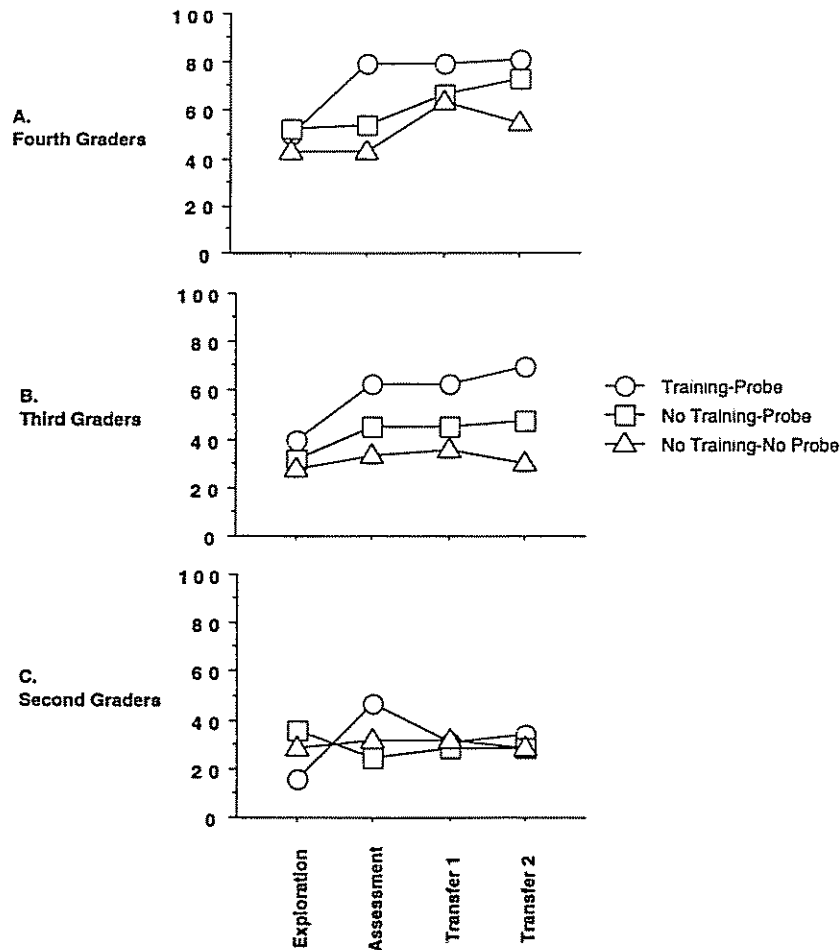


FIG. 3.3. Percentage of correct CVS usage by phase, grade, and condition (lab study).

condition, and 13% in the no training/no probe condition became good experimenters.

Relations Between the Use of CVS and Domain Knowledge

An important issue concerning the function of training in CVS is whether children's domain-specific knowledge—that is, their understanding of the effects of the variables associated with springs, ramps, and sinking—improved as a result of training. Because our primary goal was to examine elementary school children's ability to learn and transfer CVS, neither the training nor the probe questions were directed toward, or contingent on, the children's understanding of the *content* of the domains. However, any change in children's beliefs about the causal mechanisms in the three domains is of obvious interest, because the ultimate goal of good experimental design is to learn about the world. We found that only those children who were directly trained to design informative (i.e., unconfounded) comparisons showed an increase in their domain knowledge (see Fig. 3.4).

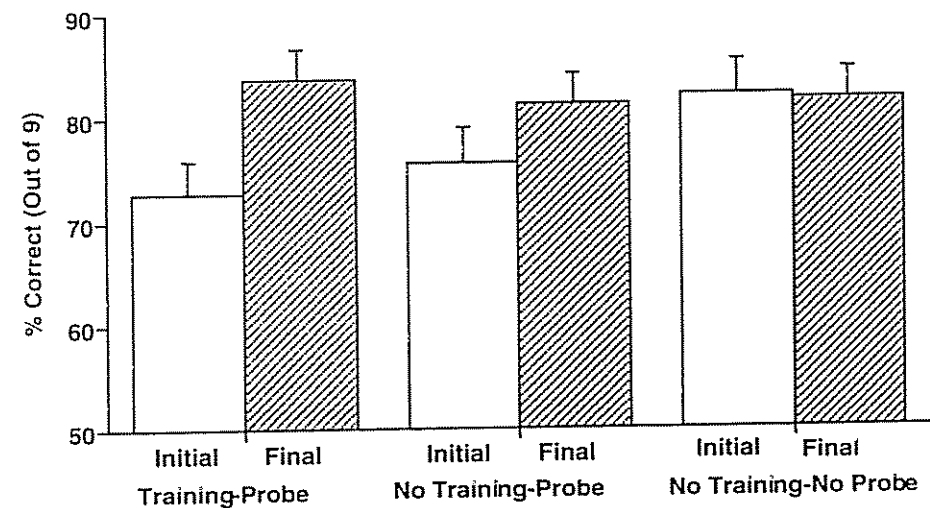


FIG. 3.4. Initial and final domain knowledge for each condition (lab study).

Posttest Performance

The posttest was designed to see whether children were able to transfer the learned strategy to remote problems after a long (7 months) delay. In School A, all children who participated in the hands-on interviews were trained in CVS, either early in the procedure or at the end of the hands-on study. Because they were all trained at some point, all of these School A children are now, for the purposes of the posttest analysis, considered the experimental group, whereas their classmates who did not participate make up the control group. Posttest data were collected only in School A and therefore only third and fourth graders were included.

Far transfer was indexed by the number of correct responses to the 15 posttest problems. A correct response was given a score of 1, an incorrect one, a score of 0. We found that fourth graders—but not third graders—in the experimental group outperformed those in the control group (see Fig. 3.5).

Another measure of remote transfer involved the percentage of “good reasoners” in the experimental and control groups. Children who made 13

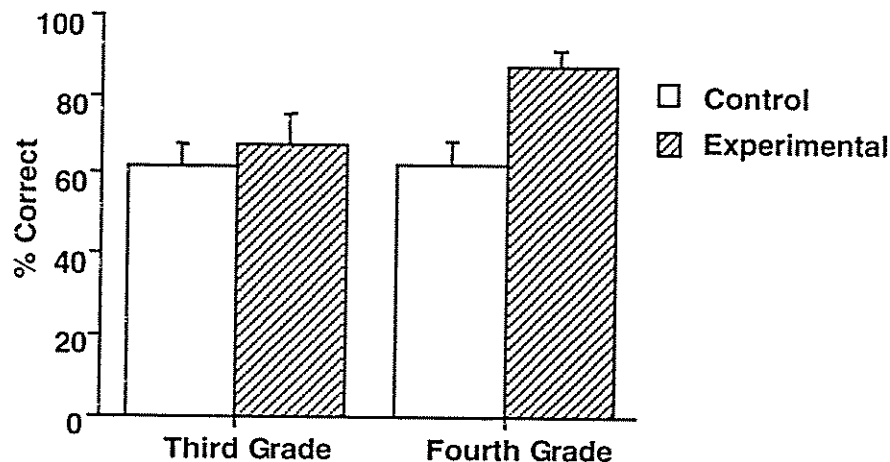


FIG. 3.5. Percentage of correct posttest answers by grade and condition (lab study).

or more correct judgments out of a total of 15 problems were considered good reasoners. Forty percent of the third and 79% of the fourth graders in the experimental group were categorized as good reasoners, compared to 22% of the third and 15% of the fourth graders in the control group. This difference was significant only for the fourth graders.

MAIN FINDINGS FROM THE LABORATORY STUDY

To summarize, the key results from our laboratory study are that absent direct instruction, children did not discover CVS on their own, even when they had repeated opportunities to work with hands-on materials; that brief direct instruction on CVS, combined with active participation in experimental setups and execution, was sufficient to promote substantial gains in CVS performance; and that these gains transferred to both near and (for fourth graders) far domains. These results gave us confidence that we were ready to move to the classroom and we began planning to recruit a few elementary school science teachers to let us implement the instructional method that had worked so well in our lab in their classrooms. Thus began the second phase of our project.

MOVING FROM THE LAB TO THE CLASSROOM

Although our lab study demonstrated that participation in a brief session of direct instruction about CVS produced substantial and long-lasting learning in fourth graders, it was clear to us that the type of one-on-one instruction and assessment used in a typical psychology experiment requiring strict adherence to a carefully crafted script would be impractical for everyday classroom use. Furthermore, we became increasingly aware that our lab study had a relatively narrow focus when compared to the multiple goals and pragmatic constraints that classroom teachers usually have when teaching about experimental design. Thus, we formulated the goal of translating, adapting, and enriching this procedure so that it could be used as a lesson plan for a classroom unit, that is, engineering a classroom learning environment (Brown, 1992; Collins, 1992). In addition, because we wanted to study the effectiveness of this translation process, we recognized the need to include a variety of assessment

procedures—assessments that would serve the dual purpose of enhancing students' learning while informing us about the relative effectiveness of our instruction.

With this as background, we began to craft a lesson plan based on our initial laboratory script. In designing the lesson plan and its associated assessments, we addressed the following questions: (a) Can fourth graders learn and transfer CVS when exposed to direct classroom instruction combined with hands-on experimentation? (b) Does the classroom introduce any new issues or difficulties in learning CVS? (c) Will instruction that is focused on the design and justification of students' own experiments increase their ability to evaluate experiments designed by others? (d) What is the relation between students' experimentation skills and the acquisition of domain knowledge?⁸

Throughout this process, we conceptualized our task in terms of differences and similarities between lab and classroom with respect to instructional objectives and pragmatic constraints, and types of assessments. These are summarized in Table 3.5. For a minimalist but still effective intervention, we maintained both the instructional objective (teaching CVS) and the proven instructional strategy (direct instruction interspersed with hands-on experimentation) from the earlier laboratory study. In addition, we attempted, insofar as possible, to make all necessary modifications in terms of our theoretical orientation that the mechanism of transfer from one domain to another was analogical processing. Within these constraints, there were several important differences between the laboratory script and the classroom lesson.

Pragmatic differences were extensive. Because the teacher could not keep track of the experimental setups of all of the groups, we transferred this responsibility to the students. They were instructed in how to record, for each of their experiments, the way that they had set up their pair of ramps. We provided students with worksheets that they completed after each experiment. The worksheets included a preformatted table representation to record ramp setups (see Appendix). The methods for filling out this table and the rest of the questions on the worksheet were discussed before experimentation. Thus, although students had to record the way in which they set up each pair of ramps, they did not have the additional responsibility of devising an external representation for the physical setup.

TABLE 3.5
Comparison of Pragmatics and Instructional Methods in Laboratory and Classroom Study

	<i>Laboratory Study</i>	<i>Classroom Study</i>
Instruction		
Instructional objective	Mastery of CVS	Mastery of CVS
Instructional strategy	Didactic instruction of one student Active construction, execution, and evaluation of experiments by solo student	Didactic instruction — group of students Active construction, execution, and evaluation of experiments by group (unequal participation possible)
Materials	Ramps of springs or sinking	Only ramps during classroom work. (Springs and sinking during individual pre and posttest interviews)
Cognitive mechanism targeted	Analogical transfer	Analogical transfer Representational transfer with interpretive use of experimenter-provided representation
Pragmatic Constraints		
Timing	Two 45-minute sessions, during or after school	Four 45-minute science classes
Teacher	Outside experimenter	Regular science teacher
Student grouping	Individual students	Entire classroom, organized into five groups of 3–4 students
Teacher–student ratio	1 to 1	1 to 20
Record keeping	By experimenter, not available for students	By students in experimenter-designed data sheets
Assessment		
	Domain knowledge test	Domain knowledge test
	Experimenter's written record of comparisons made by students during individual interviews	Experimenter's written record of comparisons made by students during individual interviews
	Videotaped record of students' answers to questions about comparisons during individual interviews with subset of subjects	Videotaped record of students' answers to questions about comparisons during individual interviews with subset of subjects Students' written records of comparisons made and responses given during classroom work Paper and pencil pre and posttests for all students in participating classes

⁸This transition from the lab to the classroom also involved a variety of practical, conceptual, organizational, and interpersonal processes that are described elsewhere (Toth, Klahr, & Chen, 2000).

However, they did have to negotiate the mapping between physical and tabular representations, and they received detailed instruction on how to do this. During the classroom work, only the ramps domain was used so that the other two domains (sinking and springs) could be used for the individual assessment interviews preceding and following the classroom work. Instead of a single student working with an experimenter, students in the classroom worked in groups of three to five people per pair of ramps. They made joint decisions about how to set up the pair of ramps, but then proceeded to individually record both their setup and the experimental outcome in their laboratory worksheets. (This is explained in more detail later.)

Assessment methods in the classroom were derived from assessments developed for the laboratory study. In both environments, students were tested for their domain knowledge prior to and after instruction. In the laboratory work, this happened in a dialog format between the experimenter and the individual student, whereas in the classroom, each student completed a paper and pencil forced-choice test. Students' ability to compose correct experiments was measured in both situations from the experimental comparisons they made with the set of two ramps. In the classroom study, a paper and pencil experiment evaluation test—similar to the far transfer test used in the lab study—was given before and after instruction.

METHOD

Research Design for Classroom Study

The research design for the classroom study included a set of nested preinstruction and postinstruction measures (see Fig. 3.6). The "inner" set of evaluations—depicted inside the solid box in Fig. 3.6—used several assessment methods, including an in-class paper and pencil test for evaluating experiments that was identical in form to the remote posttest used in the lab study. The full set of assessments was designed to measure students' hands-on experimentation performance as well as their ability to evaluate experiments designed by others. These evaluations were administered by the teacher to all students, in class, immediately before and after the instructional sessions. The "outer" set of individual one-on-one interviews used the same scoring procedures used in Part I of the lab study. For half of the individual interviews, the pretest domain was springs and the posttest domain was sinking objects, and for the other half, the order was reversed.

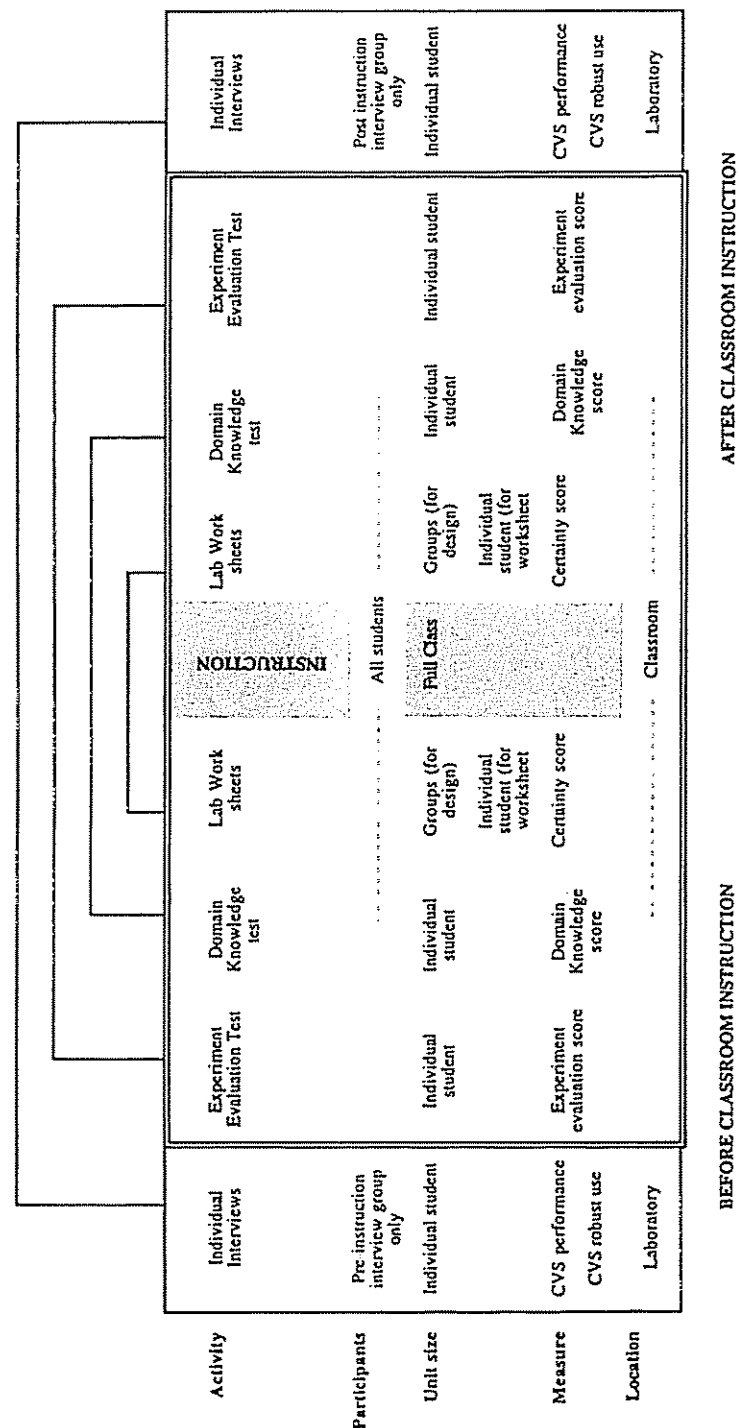


FIG 3 6 Schedule of various assessments before and after classroom insiruc-

Participants

Seventy-seven students from 4 fourth-grade classrooms in two demographically similar private elementary schools in southwestern Pennsylvania participated. Neither school had participated in the earlier lab study. Schools were selected from among the set of schools represented in our small teacher network on the basis of several pragmatic factors, including permission of school authorities, teacher interest and available time, and the fit between the CVS topic and the normal progression of topics through the fourth-grade science curriculum. From these four classrooms, we recruited volunteers for pre and postinstruction interviews. Of the 77 students participating in the study, 43 students volunteered to be individually interviewed.

Procedure

Individual Interviews

The initial and final assessments were based on individual interviews that were essentially identical to those used throughout the lab study. The pragmatics of conducting research in schools shaped the design of this outer evaluation, because we could only conduct the individual interviews with "volunteers" for whom we had received parental permission.⁹ Because we wanted to avoid any potential reactivity between the individual assessments and students' response to the classroom instruction, we included only half of the "permission" students on the individual lab pretest and the other half on the individual lab posttest. Twenty-one of the 43 volunteer students were randomly assigned to the preinstructional interview group and were individually interviewed before the classroom activities began. The rest were assigned to the postinstructional interview group and were individually interviewed after the classroom activities had been completed. The assumption was that in each case, these students were representative of the full classroom and that there would be no reactivity. Subsequent analyses supported both assumptions.

These individual pre and postinstructional interviews—conducted out of the classroom in a separate room—included students' hands-on design of valid experiments as well as verbal justifications for their experiments and the conclusions they drew from them. The interviewer followed the

⁹This permission is required by our Institutional Review Board for all experimental work, but not for classroom interventions that do not depart substantially from normal classroom instruction. (This constraint is just one more example of the complexity of applied work.)

same script used in the lab study, except that now students were asked to design and conduct nine experiments: three with each of three variables. After designing each of their experiments, students were asked to justify them. They were also asked to indicate how certain they were about the role of the focal variable, based on the outcome of the experiment. They were asked: "Can you tell for sure from this comparison whether ___ makes a difference? Why are you sure/not sure?" The entire session was recorded on videotape.

Classroom Activities and Assessments

Experiment Evaluation Assessment. At the start of the first day of the classroom work, all students individually completed a paper and pencil, experiment evaluation test on which they judged preconstructed experiments to be good or bad. Students were presented with 10-page test booklets on which each page displayed a pair of airplanes representing an experimental comparison to test a given variable. For each airplane, there were three variables considered: length of wings, shape of body and size of tailfin. Figure 3.7 depicts one of the types of comparisons that were used. Four different types of experiments were presented:

The engineers wanted to compare two planes to figure out whether the length of the wings makes a difference in how fast a model plane flies. Picture A shows one plane they built, and picture B shows the other plane they built.

- They built plane A with a thick body, and they built plane B with a narrow body.
- They built plane A with long wings, and they built plane B with short wings.
- They built plane A with a big tail, and they built plane B with a big tail.

Look at these two pictures carefully. If you think these two pictures show a good way to test if the length of the wings makes a difference, circle the words "Good Test" below. If you think it is a bad way, circle "Bad Test."

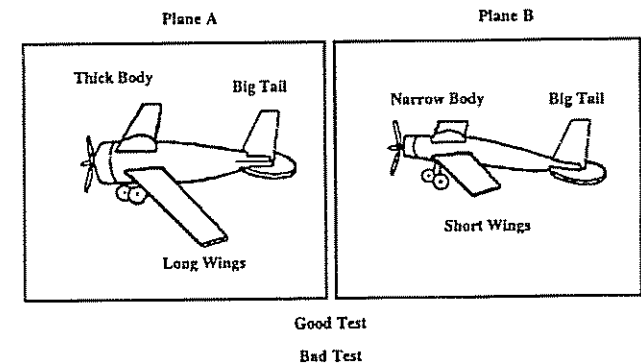


FIG. 3.7. Sample page from experiment evaluation assessment booklet used in classroom study (airplanes test). This example has a single confound because the body type is confounded with the focal variable (wing length).

(1) unconfounded comparisons, (2) singly confounded comparisons, (3) multiply confounded comparisons, and (4) noncontrastive comparisons. Students were asked to evaluate these comparisons—that is, to judge whether each picture pair showed a valid experiment to test the focal variable—by circling the word “bad” or “good.” (Only unconfounded comparisons are good tests; all others are bad.) This assessment was repeated, using a different set of materials, after classroom instruction (see Fig. 3.6).

Classroom Instruction. Classroom instruction began with a short demonstration of the different ways the ramps can be set up and an explanation of how to map these ramp setups into preformatted tables on the students’ individual laboratory worksheets. Following the demonstration, there was a short domain knowledge test, to assess students’ prior beliefs about the role of different variables on the ramps. The next phase of classroom work was comprised of three parts: (1) exploratory experiments conducted in small groups,¹⁰ (2) direct instruction for the whole classroom, and (3) application experiments conducted in small groups.

Exploratory Experiments. Students were asked to conduct four different experiments—two to test each of two different variables. Students were required to individually record their experimental setups and data into preformatted worksheets. These worksheets had two sections (see Appendix). The first section asked students to map their ramp setup into a table representation and the second section included questions about the outcome of each experiment and about whether the students were sure or unsure from this experiment about the focal variable’s influence on the experimental outcome.

For example, they were asked: “Does the X (the focal variable) of the ramp make a difference? Circle your answer: Yes / No. Think about this carefully, can you tell for sure from this comparison whether X (the current, focal variable) of the ramp makes a difference? Circle your answer: Yes / No.” The students were not asked to provide a rationale for their answer. These four experiments conducted in the first stage of classroom work were later analyzed for students’ preinstruction knowledge of CVS.

The process whereby these worksheets were designed illustrates some of the complexities of the lab-to-classroom transition. Our initial concep-

¹⁰Assignment of students to groups was determined by the teacher’s judgment of the ability of the different students to work together.

tion was very simple: Because we could not simultaneously observe what each group was doing, we needed some way to keep track of their experimental setups. Having each student record them seemed like the most obvious way to do this. In our collaboration with teachers prior to the classroom study, we considered several forms for this worksheet, and finally converged on the one illustrated in the Appendix. Although it seems fairly straightforward, we have no rigorous basis for claiming that it is optimal, or ideally suited for all students or for all CVS instruction. At present, this form simply represents an educated guess of the kind that permeates many transitions between basic research and applied contexts. We return to this issue at the end of the chapter.

Another important difference that emerged in the design of this form—and the classroom process in general—has to do with experimental error. In the course of rolling real balls down real ramps, a variety of errors can occur (even in unconfounded experimental designs). For example, the ball might bump into the side of the ramp, the experimenter (or student) might unintentionally accelerate or impede one of the balls, and so forth. In the lab study, any such anomalous experiments were simply corrected by the experimenter, with a minor comment (e.g., “oops, let’s run that one again”). In this way, the experimenter could maintain control over the good and bad executions of each experiment and dismiss any effects of random error. However, in the classroom context, this rigid and artificial control over error is neither possible nor desirable. It is not possible, because—as indicated by the extensive discussions of potential error sources by the children in Lehrer, Schauble, Strom, and Pligge’s chapter (chap. 2)—the issue of variation in outcomes becomes a topic of inherent interest when groups of students are running experiments. It is not desirable because students’ conceptions of error and their understanding of the distinction between a design error (i.e., a confounded experiment) and other types of error (random error and measurement error) are typically among the instructional objectives of science teachers who—like the teachers in our study—insist that children always do multiple trials for the same experimental setup. Thus, the issue of how students understand error, absent from our lab study, arose for the first time as we moved to the classroom. (As we explain later, it became one of the issues on our list of questions that flowed from the classroom study back to a future lab study.) But this discussion of error is based on hindsight. At the time we designed the worksheet, we encapsulated the entire issue into a simple question about students’ certainty about the conclusion they could draw from their experiments.

Direct Instruction. The second stage included about 20 minutes of direct instruction to the entire class on how to create valid experiments. The students' regular science teacher followed these six steps (see Toth, Klahr, & Chen, 2000, for details):

1. *Initiate reflective discussion* based on a "bad comparison"—a multiply confounded comparison between two ramps. After setting up this bad test the teacher asked students whether it was a good or bad test and then provided time and opportunity for students' different views and—often conflicting—explanations. The teacher asked the students to point out what variables were different between the two ramps and asked whether they would be able to tell for sure from this comparison whether the focal variable made a difference in the outcome.
2. *Resolve students' opposing points of view by modeling correct thinking.* After a number of conflicting opinions were heard, the teacher proceeded to reveal that the example was not a good comparison. She explained that other variables, in addition to the focal variable, were different in this comparison, and thus if there was a difference in the outcome, one could *not* tell for sure which variable had caused it. The teacher proceeded to make a good comparison to contrast with the bad one and continued a classroom discussion to determine why the comparison was good.
3. *Test understanding* with another bad comparison. Next, the teacher tested the students' understanding with another bad comparison and asked a similar set of questions.
4. *Reinforce learning* by pointing out the error in the bad comparison.
5. *Summarize the rationale* for CVS. The teacher reinforced her teaching by providing a detailed account of the confounds in the bad test. Next, the teacher created another good comparison and used the same method of classroom discussion as before to review why this test allowed one to tell for sure whether the studied variable makes a difference.
6. Finally, the teacher provided an overall *conceptual justification* for CVS with the following words:

Now you know that if you are going to see whether something about the ramps makes a difference in how far the balls roll you need to make two ramps that are different only in the one thing that you are testing. Only when you make those kinds of comparisons can you really tell for sure if that thing makes a difference.

Application Experiments. The third phase of the classroom work was created to allow students to apply their newly learned strategy during experimentation. The students' activity in this phase was very similar to what they did in Phase 1, with the exception that during the first application experiment, they tested the effect of a variable they had not tested previously. In the second application, they tested the same variable they had tested in Phase 1.

Measures

We used five measures, similar to those used in the lab study, designed to capture both the procedural and logical components of the control of variables strategy (CVS). They included: (1) *CVS performance score.* We measured students' CVS performance by scoring the experiments students conducted. Each valid (unconfounded) comparison was scored 1, and all other, invalid comparisons were scored 0. (2) *Robust CVS use score.* During individual interviews, a score of 1 was assigned to a valid experiment accompanied by a correct rationale. (3) *Certainty measure.* Probe questions asked students whether they were certain about their conclusion about the role of the focal variable. This question was stated in both the individual interviews and in the classroom worksheets. The certainty measure—not used in the lab study—was intended to capture some of the additional complexity of the type of knowledge students extract from classroom experiences. (4) *Experiment evaluation score.* Correctly indicating whether a given experimental comparison was good or bad gained students a score of 1 and incorrect evaluations were scored 0. (5) *Domain knowledge score.* Correct prediction of the effect for each variable was scored as 1 and incorrect prediction as 0.

RESULTS FROM THE CLASSROOM STUDY

First, we present the results on procedural knowledge, that is, knowledge about CVS based on all instruments: individual interviews, classroom worksheets, and pre/post experiment evaluation tests. Next, we describe students' domain knowledge, that is, knowledge about which values of the variables make a ball roll farther—before and after classroom instruction. Finally, we report on changes in students' ability to discriminate between

good and bad experiments created by others. For each measure, we provide pre and postinstructional comparisons (corresponding to the pairs of connected columns in Fig. 3.6).

Analysis of CVS Performance and Certainty Based on Individual Interviews

CVS performance scores on the individual interviews increased dramatically following instruction, from a mean score of 30% prior to instruction to a mean score of 96% after instruction. With respect to individual students, we defined a CVS "expert" as a student who correctly used CVS on at least eight of the nine trials in the individual interviews. Only one of the 21 children taking the individual pretest interviews was an expert, whereas 20 of the 22 children in the posttest individual interviews exhibited such near-perfect performance.

A similar analysis on robust use (designing an unconfounded experiment *and* providing a CVS rationale) revealed an increase in the mean score from 6% on the pretest to 78% on the posttest. Prior to instruction, none of the 21 students in the preinstructional interview group was a robust expert (i.e., robust use on more than eight of nine trials), whereas after instruction, 12 of the 22 (55%) in the postinstructional interview group were experts. Interestingly, of the 20 CVS use experts, only 12 were robust CVS use experts. That is, a substantial proportion of children who could do CVS failed to explain it adequately.

The analysis of certainty scores also revealed a large improvement in children's confidence about the conclusions they could draw from well-designed experiments. Prior to instruction, children exhibited little differentiation in the confidence with which they drew conclusions from confounded versus unconfounded experiments (see Fig. 3.8). When they designed good experiments, they said they were certain about the effects for 70% of the trials, and when they designed bad experiments, they said they were certain for 60% of the trials. That is, they were only 17% more likely ($.70/.60 = 1.17$) to be certain about a conclusion drawn from an unconfounded experiment as from a confounded one. In contrast, after instruction, children showed a high degree of differentiation between these two classes of experiments. When they designed good tests, they were certain for 84% of the trials, and when they designed bad tests, they were certain only for 46% of the trials. That is, they were nearly twice as

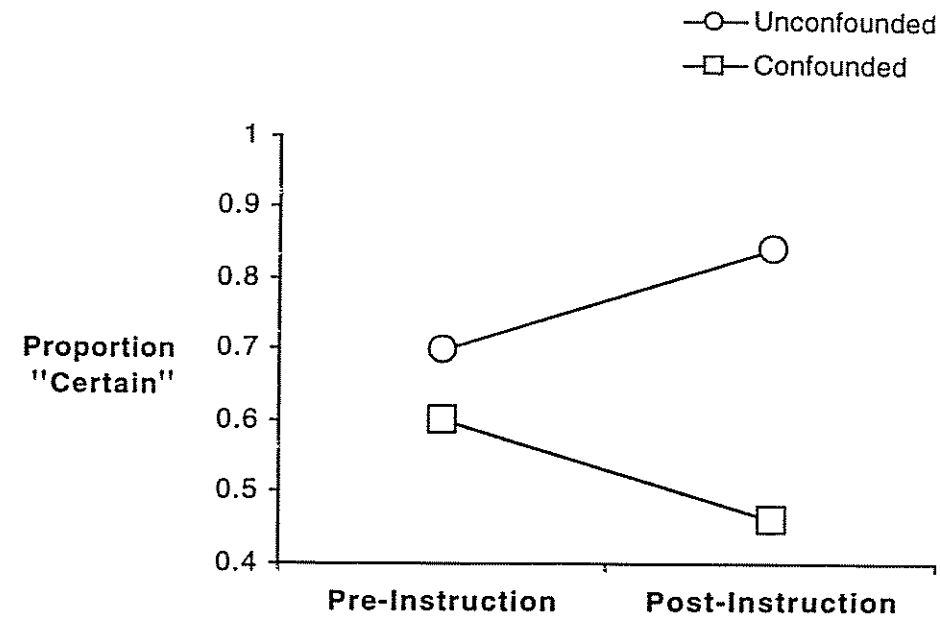


FIG. 3.8. Proportion of student responses before and after classroom instruction indicating certainty about the conclusions that could be drawn from unconfounded and confounded experiments (based on individual lab interviews).

likely to be confident about the conclusions they could draw from an unconfounded experiment as from a confounded one.

Despite this improvement, we were puzzled by why the proportion of certain conclusions from unconfounded experiments was only 84%, rather than much closer to 100%. We conjecture that there are at least two quite different reasons for this lack of certainty about how to interpret the outcome of what is formally an unconfounded (and therefore unambiguous) experiment. First, the experiment might have been fortuitously unconfounded, even though the student didn't fully understand the logic of CVS. Such uncertainty about the design of an experiment would be likely to be reflected in a low certainty score for the outcome of the experiment. Second, it is possible that some aspect of the execution of an experiment whose design was unconfounded would lead to uncertainty about the outcome. (We return to the different types of error later in the chapter.)

Analysis of CVS Performance and Certainty From Classroom Worksheets

The nested design used in this study allowed us to measure several of the same constructs in both the lab and the classroom (see Fig. 3.6). In this section, we describe the results of the “inner” pairs of pre and postmeasures. As noted earlier, during classroom activities, students worked in small groups. Although the students made their ramp setup decisions and built experimental comparisons together, they each individually filled out a laboratory worksheet. Mean CVS performance scores derived from these worksheets increased from 61% before instruction to 97% after instruction. However, here too, students remained uncertain about the effect of the focal variable on approximately 20% of these experiments.

Analysis of Domain Knowledge Test

Recall that at no point was there any direct instruction regarding the role of the causal variables in the ramps domain. Nevertheless, there was a significant pre to postinstructional increase in domain knowledge. Whereas 79% of the students provided correct answers to all three questions on the domain knowledge test prior to CVS instruction, all students correctly answered all three domain knowledge questions after instruction. Unfortunately, we cannot attribute this gain entirely to the CVS training, because the classroom study had no control group of students who had an equivalent amount of experience in setting up and running experiments without the direct classroom instruction.

Analysis of Experiment Evaluation Scores

We found a similar increase in students’ ability to evaluate experiments designed by others. The mean experiment evaluation (airplanes comparison) scores increased from 61% correct on the initial test to 97% correct on the final test. The percentage of students who were evaluation experts—that is, who could correctly evaluate at least 9 of the 10 comparisons—increased from 28% prior to instruction to 75% after instruction. Thus, a brief period of direct instruction interspersed with hands-on experimentation significantly increased students’ ability to evaluate the validity of experiments designed by others.

DISCUSSION (CLASSROOM STUDY)

The main goal of the classroom study was to determine whether an instructional procedure that produced substantial and long-lasting learning in the psychology lab could be transformed to an effective instructional unit for everyday classroom use. The laboratory instruction involved one-on-one direct instruction coupled with hands-on experimentation. The classroom teaching also used direct instruction, but now it was instruction directed at several teams of students conducting hands-on experimentation. As indicated by a series of independent but converging measures, the classroom instruction was overwhelmingly successful, not only in terms of statistical significance, but more importantly, with respect to absolute levels of performance. Overall, students’ ability to design and assess unconfounded experiments, which ranged from 30% to 60% on various pretests, now was close to perfect.¹¹ Students learned to do CVS, to explain CVS, and to distinguish between CVS and non-CVS experiments designed by others. Contrary to previous suggestions that early elementary school students are developmentally unable to conduct controlled experiments because it requires formal operational thinking, our results indicate that students’ procedural and conceptual knowledge significantly increased after a short, succinct direct instruction session combined with hands-on experimentation.¹²

SHIPS AND BEACONS REVISITED

Our goal in this chapter has been to describe a case in which the intersection between psychological research and classroom practice is not the empty set. In this concluding section, we discuss some of the possible reasons for the success of our lab-to-classroom transition, and then we discuss some of the issues related to the classroom-to-lab aspect of the “mutual illumination” notion.

¹¹See Toth, Klahr, and Chen (2000) for a more detailed analysis of the classroom study.

¹²One reviewer of this chapter pointed out the fact that we had very “clean” classrooms: private schools, cooperative teachers, well-behaved students. However, these factors still produced average CVS scores of around 50% prior to our intervention.

Why Did Our Instruction Work?

To what can we attribute the success of our classroom instruction? Why was our simple procedure so effective, when, as indicated earlier, even fourth graders who had been exposed to high-quality science curricula for several years could design unconfounded experiments on fewer than one third of their initial attempts, and could correctly explain their designs on fewer than 10% of their attempts (based on the individual interviews)? In order to give a complete answer to this question, we would have to carefully examine these children's prior exposure to CVS instruction and activities. However, in lieu of such information, we can look at a few examples of "model" practice, and contrast our procedures with what we find there.

Our examination of dozens of texts and monographs about science teaching and learning suggests that, at best, CVS receives only a few paragraphs in a handful of books. And in the rare instances when it is taught, instruction is both brief and confusing, as illustrated by the following two examples.

Example 1. A Middle-School Science Text. Consider the excerpt shown in Table 3.6, taken from the opening chapter of a widely used middle-school science text. (Although students do not use texts extensively in K-4 science instruction, teachers get advice and suggestions from a variety of such books and other teacher-oriented materials; for example, the journal *Science and Children*. The content of such articles tends to be similar to the illustrative example used here.)

At first glance, the excerpt seems like a reasonable treatment of many of the issues involved in CVS. However, a careful examination reveals several potential difficulties. Consider, for example, the presumption that the student would believe that amount of light is a growth factor for mold (sent. 2). To a student lacking any domain knowledge about the growth of mold, this might seem as plausible (or implausible) a factor as type of container, or food source, or day of the week. Indeed, as the student discovers later (sent. 12), the light variable is irrelevant.

Next, while intending to provide a control for another possible causal variable (water), the passage introduces an unnecessary quantification of that variable (i.e., Why 10 drops each? Why not 12 drops each? Is it the "10" that's important here, or the "each"?) and it adds a procedure whose impact is not explicit (Is "cap tightly" important?).

TABLE 3.6

Example of a Description of a Control of Variables Strategy
(from HBJ Science - Nova Edition, 1989)

1. A simple question occurs to you.
2. Will mold grow better in the light or the dark? This, you decide, calls for an investigation.
3. So you divide a slice of bread in half and place each in a jar.
4. You add ten drops of water to each jar and cap tightly.
5. You put one in a dark closet.
6. You keep the other one in the light.
7. In a few days, you make an observation.
8. You observe that the mold in the light is growing better than the mold in the dark.
9. From this observation, you infer that mold grows better in the light.
10. You are sure of your answer.
11. After all, the evidence seems clear.
12. Yet, in truth, scientists know that light has no effect on the growth of mold.
13. Why, then, did the investigation make it seem as though mold grows better in the light?
14. Think a moment.
15. Because the amount of light varied, we say that light was a **variable**.
16. Since light was the only variable considered, you assumed that it was the light that affected the growth of the mold.
17. Was light the only variable you changed? Or could you have changed another variable without realizing it?
18. What about temperature? Suppose that the temperature of the mold in the light was higher.
19. Then it is possible that the higher temperature, not the light itself, caused the growth.
20. Whenever you do an investigation, there may be several possible variables.
21. If you wish to see the effects of changing one variable, such as the amount of light, then you must make sure all the other possible variables remain the same.
22. That is, you must **control** the other variables.
23. In this investigation, to control the variable temperature, you must keep the temperature the same for the mold in the light and the mold in the dark.
24. If you had done so, you would have discovered that the mold grew just as well in the dark closet.
25. However, even then you could not be sure of your conclusion.
26. The investigation must be repeated many times.
27. After all, what happens once can be an accident.
28. Scientists don't base their conclusions on one trial.
29. They repeat the investigation again and again and again.
30. The rest of this year you will be an apprentice.
31. You will be an investigator.
32. You may even decide one day to become a scientist.

Note: Emphasis in original. Sentence numbers added.

Following the quantitative variable, the passage introduces categorical variables (sent. 5–6), but without any explicit statement that the experiment will be run in terms of categorical variables, rather than continuous variables. (And this is in contrast to the unnecessarily specific quantification of the amount of water.)

The goal of the next section (sent. 9–18) appears to be to demonstrate that it is easy to forget potentially important variables. More specifically, the intent is to show the student that it is important to consider causal variables other than light, and that temperature is one such possibility. But this example is problematic because it confounds the domain-general notion of CVS with the domain-specific knowledge that temperature might be a causal variable in this domain. Thus, as stated here, the example might convey the mistaken notion that the logic of the control of variables strategy is flawed in some way. Moreover, the example takes the student down a garden path to a pattern of evidence that is particularly difficult to interpret. Young children are easily misled when they are faced with a single piece of positive evidence, and several remaining sources of unknown evidence (Fay & Klahr, 1996; Piéraud-Le Bonniec, 1980). They tend to believe that the single instance renders the situation determinate, even though they will acknowledge that additional evidence might change that decision.

In closing, the passage abruptly introduces the notion of error variance (sent. 25), but it does so in a way that suggests that just when you think you are sure, you are really not sure. It is not surprising that students come away from such examples believing that their subjective opinions are as valid as the results of scientific investigations.

Finally, as brief as it is, the example represents the *only* explicit attempt in the entire book to teach the principles of good experimental design, the logic of rival hypothesis testing, and the distinctions between valid and invalid inferences or determinate and indeterminate situations.

In summary, the example attempts to cover too many things at once, and it confounds issues pertaining to the abstract logic of unconfounded experimentation and valid inference with other issues having to do with domain-specific knowledge and plausible hypotheses about causal variables.

Example 2. Model Inquiry Unit From NSES. As noted earlier, the middle-school science text example is used only to illustrate the complexity and subtlety of the topic. More relevant to our point about the inadequacy of the material available to K through 4 teachers who would like to teach CVS is the model inquiry unit entitled “Earthworms” pro-

vided in the NSES. The full unit carefully and sensitively elaborates a process whereby third graders interested in creating a habitat for earthworms could learn about the earthworm life cycle, needs, structure, and function. However, its treatment of CVS is extremely sparse:

Two groups were investigating what kind of environment the earthworms liked best. Both were struggling with several variables at once—moisture, light, and temperature. Ms. F. planned to let groups struggle before suggesting that students focus on one variable at a time. She hoped they might come to this idea on their own. (NSES, 1995, pp. 34–35)

This brief treatment of CVS—consisting of only about 50 of the 1,000 words in the Earthworms inquiry unit—provides virtually no guidance to the teacher on how to present the rationale of CVS, how to provide positive and negative instances, how to draw children’s attention to the variables and the design of tests, or how to guide students in interpreting results from controlled or uncontrolled experiments.

Contrasts Between Examples and Our Procedure

These examples suggest several potentially important contrasts between our approach and what children had experienced earlier. First, it is clear that, in typical classroom situations, the detailed components of CVS are not adequately isolated and emphasized either directly, as decontextualized domain-general principles or indirectly, as contextualized skill in a specific domain. In contrast, our instruction avoided, insofar as possible, potential confusion between CVS logic errors and inadequate domain knowledge by making it very clear exactly what dimensions were under consideration at all times. It also used positive and negative examples of CVS designs, and it presented students with both a mechanistic procedure and a conceptual justification for why the procedure worked. We presented several examples of each, so that students could construct an internal representation of CVS that could be transferred, via analogical mapping, to new, but structurally similar situations.

Second, in contrast to the second example’s suggestion that children “might come to this idea on their own,” we made instruction highly explicit and direct. Prior to the results of our lab study, this focus on explicit instruction was based partly on our own intuitions and partly on the results of other studies of the power of detailed and direct instruction

about complex procedures (e.g., Klahr & Carver, 1988). The results of our lab study further demonstrated, in the specific context of CVS, that children find it very difficult to discover CVS on their own. As we have argued several times, we believe that there is too much here for students to acquire on their own, via discovery, so we opted instead for explicit and direct instruction about each of these aspects of CVS.

NEW ISSUES IN BASIC RESEARCH ON SCIENTIFIC THINKING

Although we have argued for a bidirectional flow between lab and classroom, thus far, we have mainly emphasized only one side of that flow: the lab to classroom transition. In this concluding section, we illustrate the other side of that flow by elaborating on some of the new issues that the classroom study has raised—issues that we will continue to investigate in the laboratory.

Representation

Recall that the classroom study used a worksheet on which students recorded the way that they set up their ramps. At the time we prepared these worksheets, we were working under the usual kinds of scheduling deadlines that surround any attempt to intervene in an ongoing “live” classroom during the school year, and so we did not have the luxury of carefully considering just what was involved in asking students to carry out this “trivial” task. However, as psychologists, we are well aware that such mappings entail a complex set of procedures for establishing correspondence between the physical setup of the ramps and a set of marks on paper that represent that setup. Indeed, the ability to move between various equivalent representations for both apparatus and data is one of the commonly stated goals of science educators, and represents the main focus of the chapter by Lehrer, Schauble, Strom, and Pligge (chap. 2). Only with systematic study in a laboratory context on just this issue could we claim that the particular form that we used is the best way to capture students’ representations of their experiments. Thus, the issue of the most effective type of representation for this and similar types of classroom exercises has become a topic not only on our research agenda for further laboratory studies, but it has also become a potential instructional objective for subsequent classroom work by the teachers in our schools.

Although the lab studies of representation remain in the planning stages, we have already begun to explore the representation issue in the classroom by introducing a very simple change in our procedure. Rather than providing each group with a pair of ramps, we provide only one ramp. This requires students to set up, execute, and record the effect of one combination of variables, and then follow with another setup, execution, and recording. We believe that this will challenge students to consider the important role of “inscriptions” (Lehrer, Schauble, Strom, & Pligge, chap. 2) as permanent, inspectable representations of transient events, and that it will motivate them to more accurately record and better interpret such external representations and their role in science.

Certainty and Error

Although our classroom instruction produced substantial increases both in students’ ability to design unconfounded experiments and in their certainty about the conclusions they could draw from them, there remained a non-trivial proportion of valid experiments from which students were unwilling to draw unambiguous conclusions. Recall that, even after instruction, students indicated uncertainty about the conclusions they could draw from approximately 15% of their valid experiments (see Fig. 3.8). This occurred for both the interview and laboratory assessments. Because *all* the variables in the ramps domain influenced the outcome measure—that is, the distance a ball rolled down a ramp—this finding was, at first, puzzling to us.

Further consideration led to the following conjecture. We believe that children found it difficult to distinguish between a logical error (such as a confounded experiment) and other types of error (such as random error in the execution of the experiment or measurement error), and that they were unsure about which of several replications of the same setup was the “true” result. Although these are important aspects of a rich understanding of experimentation, we did not include them in our highly focused instructional goals. Thus, two additional topics for both detailed lab research and further classroom instruction arose: the distinction among various types of errors involved in scientific experimentation, and a better understanding of how children extract a general conclusion from several replications of the same experiment.

These two particular issues—representational competence and understanding of error—certainly do not exhaust the list of questions arising during our classroom study that could be further investigated in the psychology lab. But they serve to illustrate how complex issues that arise in

the authentic and complex setting of the science classroom can be returned to the psychology lab for further controlled investigation. More generally, they provide concrete examples of the two-way influence of relevance and importance between the lab and the classroom. To return to our opening metaphor, we believe that the work described in this chapter, as well as in many of the other chapters in this volume, support the position that the two ships of science education and cognitive development need not pass in darkness, but can, instead, be mutually illuminating.

ACKNOWLEDGMENTS

Portions of the work described in this chapter were supported in part by grants from NICHD (HD25211) and from the James S. McDonnell Foundation (96-37). We thank Jen Schnakenberg, Sharon Roque, Anne Siegel, and Rose Russo for their assistance with data collection and preliminary analysis, and John Anderson, Milena Koziol Nigam, Amy Masnick, and especially Sharon Carver for their comments and suggestions on earlier drafts of this paper. This project could not have proceeded without the generous cooperation of the parents and administrators at The Ellis School, Winchester-Thurston School, Shady Side Academy, and The Carlow College Campus School, as well as the active involvement of the master teachers who allowed us into their classrooms and their teaching practice, Linn Cline, Patricia Cooper, Cheryl Little, and Dr. Mary Wactlar. Finally, our thanks to the children at all of these schools who participated with enthusiasm and excitement.

REFERENCES

- Amsel, E., & Brock, S. (1996). The development of evidence evaluation skills. *Cognitive Development, 11*(4), 523-550.
- Brown, A. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *The Journal of the Learning Sciences, 2*, 141-178.
- Brown, A. (1997). Transforming schools into communities of thinking and learning about serious matters. *American Psychologist, 52*, 399-413.
- Bruer, J. T. (1993). *Schools for thought: A science of learning in the classroom*. Cambridge, MA: MIT Press.
- Bruner, J. S., Goodnow, J. J., & Austin, G. A. (1956). *A study of thinking*. New York: Science Editions.
- Bullock, M., Ziegler, A., & Martin, S. (1993). Scientific thinking. In F. E. Weinert & W. Schneider (Eds.), *LOGIC Report 9: Assessment Procedures and Results of Wave 6* (pp. 66-110). New York: Wiley.

- Bullock, M., & Ziegler, A. (1999). Scientific reasoning: developmental and individual differences. In F. E. Weinert & W. Schneider (Eds.), *Individual development from 3 to 12: Findings from the Munich Longitudinal Study* (pp. 309-336). Munich: Max Plank Institute for Psychological Research.
- Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: MIT Press.
- Case, R. (1974). Structures and structures: Some functional limitations on the course of cognitive growth. *Cognitive Psychology, 6*, 544-573.
- Chen, Z., & Klahr, D. (1999). All other things being equal: Children's acquisition of the Control of Variables Strategy. *Child Development, 70* (5), 1098-1120.
- Christensen, C. A., & Cooper, T. J. (1991). The effectiveness of instruction in cognitive strategies in developing proficiency in single-digit addition. *Cognition and Instruction, 8*, 363-371.
- Collins, A. (1992). Toward a design science of education. In E. Scanlon & T. O'Shea (Eds.), *New directions in educational technology* (pp. 15-22). New York: Springer-Verlag.
- Das-Small, E. A., Klapwijk, M. J., & van der Leij, A. (1996). Training of perceptual unit processing in children with a reading disability. *Cognition and Instruction, 14* (2), 221-250.
- Dunbar, K. (1993). Concept discovery in a scientific domain. *Cognitive Science, 17*, 397-434.
- Fay, A. L., & Klahr, D. (1996). Knowing about guessing and guessing about knowing: Preschoolers' understanding of indeterminacy. *Child Development, 67*, 689-716.
- Fennema, E., Carpenter, T. P., Franke, M. L., Levi, L., Jacobs, V. R., & Empson, S. B. (1996). A longitudinal study of learning to use children's thinking in mathematics instruction. *Journal for Research in Mathematics Education, 27*, 403-434.
- HBJ Science—Nova Edition*. (1989). Orlando, FL: Harcourt Brace.
- Jacoby, J. (1978). On interpreting the effects of repetition: Solving a problem versus remembering a solution. *Journal of Verbal Learning and Verbal Behavior, 17*, 649-667.
- Klahr, D. (Ed.). (1976). *Cognition and instruction*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Klahr, D. (2000). *Exploring science: The cognition and development of discovery processes*. Cambridge, MA: MIT Press.
- Klahr, D., & Carver, S. M. (1988). Cognitive objectives in a LOGO debugging curriculum: Instruction, learning, and transfer. *Cognitive Psychology, 20*, 362-404.
- Klahr, D., & Carver, S. M. (1995). Scientific thinking about scientific thinking. *Monographs of the Society for Research in Child Development, 245* (60, serial no. 4), 137-151.
- Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science, 12*, 1-55.
- Klahr, D., Fay, A. L., & Dunbar, K. (1993). Heuristics for scientific experimentation: A developmental study. *Cognitive Psychology, 24* (1), 111-146.
- Koslowski, B. (1996). *Theory and evidence: The development of scientific reasoning*. Cambridge, MA: MIT Press.
- Koslowski, B., & Okagaki, L. (1986). Non-Humean indices of causation in problem-solving situations: Causal mechanisms, analogous effects, and the status of rival alternative accounts. *Child Development, 57*, 1100-1108.
- Koslowski, B., Okagaki, L., Lorenz, C., & Umbach, D. (1989). When covariation is not enough: The role of causal mechanism, sampling method, and sample size in causal reasoning. *Child Development, 60*, 1316-1327.
- Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review, 96*, 674-689.
- Kuhn, D., & Angelev, J. (1976). An experimental study of the development of formal operational thought. *Child Development, 47*, 697-706.
- Kuhn, D., Amsel, E., & O'Loughlin, M. (1988). *The development of scientific reasoning skills*. Orlando, FL: Academic Press.
- Kuhn, D., Schauble, L., & Garcia-Mila, M. (1992). Cross-domain development of scientific reasoning. *Cognition and Instruction, 9* (4), 285-327.
- Kuhn, D., Garcia-Mila, M., Zohar, A., & Andersen, C. (1995). Strategies of knowledge acquisition. *Monographs of the Society for Research in Child Development, 60* (4, Serial No. 245, pp. 1-128).

- McCloskey, M. (1983). Naïve theories of motion. In D. Gentner & A. L. Stevens (Eds.), *Mental Models* (pp. 299-324). Hillsdale, NJ: Lawrence Erlbaum Associates.
- McDaniel, M. A., & Schlager, M. S. (1990). Discovery learning and transfer of problem-solving skills. *Cognition and Instruction*, 7, 129-159.
- McGilly, K. (Ed.). (1994). *Classroom lessons: Integrating cognitive theory and classroom practice*. Cambridge, MA: MIT Press.
- National Science Education Standards (NSES). (1995). Washington, DC: National Academy Press.
- Piéraut-Le Bonniec, G. (1980). *The development of modal reasoning: The genesis of necessity and possibility notions*. New York: Academic Press.
- Ross, A. J. (1988). Controlling variables: A meta-analysis of training studies. *Review of Educational Research*, 58 (4), 405-437.
- Ruffman, T., Permer, J., Olsen, D. R., & Doherty, M. (1993). Reflecting on scientific thinking: Children's understanding of the hypothesis-evidence relation. *Child Development*, 64, 1617-1636.
- Schauble, L. (1990). Belief revision in children: The role of prior knowledge and strategies for generating evidence. *Journal of Experimental Child Psychology*, 49, 31-57.
- Schauble, L. (1996). The development of scientific reasoning in knowledge-rich contexts. *Developmental Psychology*, 32, 102-109.
- Schauble, L., Glaser, R., Raghavan, K., & Reiner, M. (1991). Causal models and experimentation strategies in scientific reasoning. *The Journal of the Learning Sciences*, 1, 201-238.
- Shaklee, H., & Paszek, D. (1985). Covariation judgment: Systematic rule use in middle childhood. *Childhood Development*, 56, 1229-1240.
- Siegler, R. S. (1976). Three aspects of cognitive development. *Cognitive Psychology*, 8, 481-520.
- Siegler, R. S., & Liebert, R. M. (1975). Acquisition of formal scientific reasoning by 10- and 13-year-olds: Designing a factorial experiment. *Developmental Psychology*, 10, 401-402.
- Sigel, I. E., & Renninger, K. A. (Eds.). (1998). *Handbook of child psychology, Vol 4. Child psychology in practice*. New York: Wiley.
- Strauss, S. (1998). Cognitive development and science education: Toward a middle level model. In I. Sigel & K. A. Renninger (Eds.), *Handbook of child psychology, Vol. 4. Child psychology in practice* (pp. 357-400). New York: Wiley.
- Toth, E. E., Klahr, D., & Chen, Z. (2000). Bridging research and practice: A research-based classroom intervention for teaching experimentation skills to elementary school children. *Cognition and Instruction*, 18 (4), 423-459.
- Tschirgi, J. E. (1980). Sensible reasoning: A hypothesis about hypotheses. *Child Development*, 51, 1-10.
- 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.
- White, B. Y., & Frederiksen, J. R. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and Instruction*, 16, 3-118.

Appendix: Sample Page from Experiment Recording Sheets Used in Classroom

Does the surface make a difference?

I. FIRST COMPARISON FOR SURFACE:

How your ramp was set up:

Teacher reads the table aloud to students and instructs them on how to fill it out: circle answer corresponding to team's ramp setup.

VARIABLES	RAMP A	RAMP B
Surface	Smooth or Rough	Smooth or Rough
Steepness	High or Low	High or Low
Length of run	Long or Short	Long or Short
Type of ball	Golf ball or Rubber ball	Golf ball or Rubber ball

What happened after you rolled the balls down:

1. On which ramp did the ball roll farther most of the time? Circle your answer.
RAMP A or RAMP B

Teacher tells students : "Think about why you set up the ramps the way you did."

2. Does the surface of the ramp make a difference? Circle your answer.

YES or NO

3. Think about this carefully, can you tell for sure from this comparison whether the surface of the ramp makes a difference? Circle your answer.

VERY SURE or NOT SO SURE