

## Research Article

# The Equivalence of Learning Paths in Early Science Instruction

## Effects of Direct Instruction and Discovery Learning

David Klahr<sup>1</sup> and Milena Nigam<sup>2</sup><sup>1</sup>Department of Psychology, Carnegie Mellon University, and <sup>2</sup>Center for Biomedical Informatics, University of Pittsburgh

**ABSTRACT**—*In a study with 112 third- and fourth-grade children, we measured the relative effectiveness of discovery learning and direct instruction at two points in the learning process: (a) during the initial acquisition of the basic cognitive objective (a procedure for designing and interpreting simple, unconfounded experiments) and (b) during the subsequent transfer and application of this basic skill to more diffuse and authentic reasoning associated with the evaluation of science-fair posters. We found not only that many more children learned from direct instruction than from discovery learning, but also that when asked to make broader, richer scientific judgments, the many children who learned about experimental design from direct instruction performed as well as those few children who discovered the method on their own. These results challenge predictions derived from the presumed superiority of discovery approaches in teaching young children basic procedures for early scientific investigations.*

A widely accepted claim in the science- and mathematics-education community is the constructivist idea that discovery learning, as opposed to direct instruction, is the best way to get deep and lasting understanding of scientific phenomena and procedures, particularly for young children. “The premise of constructivism implies that the knowledge students construct on their own, for example, is more valuable than the knowledge modeled for them; told to them; or shown, demonstrated, or explained to them by a teacher” (Loveless, 1998, p. 285). Advocates of discovery learning concur with Piaget’s assertion that “each time one prematurely teaches a child something he could have discovered for himself, that child is kept from inventing it and consequently from understanding it completely” (Piaget, 1970,

p. 715). Moreover, they argue that children who acquire knowledge on their own are more likely to apply and extend that knowledge than those who receive direct instruction (Bredderman, 1983; McDaniel & Schlager, 1990; Schauble, 1996; Stohr-Hunt, 1996).

There are pragmatic, empirical, and theoretical grounds for questioning this position. Pragmatically, it is clear that most of what students (and teachers and scientists) know about science was taught to them, rather than discovered by them. Empirical challenges come from studies demonstrating that teacher-centered methods using direct instruction are highly effective (cf. Brophy & Good, 1986; Rosenshine & Stevens, 1986), particularly for teaching multistep procedures that students are unlikely to discover on their own, such as those involved in geometry, algebra, and computer programming (Anderson, Corbett, Koedinger, & Pelletier, 1995; Klahr & Carver, 1988). Finally, most developmental and cognitive theories predict that many of the phenomena associated with discovery learning would make it a relatively ineffective instructional method (Mayer, 2004). For example, children in discovery situations are more likely than those receiving direct instruction to encounter inconsistent or misleading feedback, to make encoding errors and causal misattributions, and to experience inadequate practice and elaboration. These impediments to learning may overwhelm benefits commonly attributed to discovery learning—such as “ownership” and “authenticity.”

However, our aim in this study was to go beyond a comparison of the immediate effectiveness of two radically different types of instruction. In addition, we tested the prediction that if children achieve mastery of a new procedure, then the way that they reached that mastery will not affect their ability to transfer what they have learned. This *path-independent transfer*, if supported, has implications for discovery learning, because one of its purported advantages is that it has long-term benefits on how children ultimately transfer what they have learned—benefits that justify its admittedly lower efficiency.

In order to evaluate this hypothesis, we had to create exemplars of both the discovery-learning and the direct-instruction approaches, expose children to them, and then challenge learners with an appro-

Address correspondence to David Klahr, Department of Psychology, Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15213-3890; e-mail: klahr@cmu.edu.

prate transfer task. However, at the outset, we faced a difficult definitional problem because nearly 100 years of research (cf. Winch, 1913) had yet to produce a consistent definition of *discovery learning*.<sup>1</sup> Therefore, we intentionally magnified the difference between the two instructional treatments in order to provide a strong test of the path-independent transfer hypothesis. In our discovery-learning condition, there was no teacher intervention beyond the suggestion of a learning objective; there were no guiding questions and no feedback about the quality of the child's selection of materials, explorations, or self-assessments. Correspondingly, we used an extreme type of direct instruction in which the goals, the materials, the examples, the explanations, and the pace of instruction were all teacher controlled.

The specific context in which we contrasted these two instructional approaches was an important elementary-school science objective known as 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 logical aspects of CVS include an understanding of the inherent indeterminacy of confounded experiments. In short, CVS is the basic procedure that enables children to design unconfounded experiments from which valid causal inferences can be made. Its acquisition is an important step in the development of scientific reasoning skills because it provides a strong constraint on search in the space of experiments (Klahr, 2000; Klahr & Simon, 1999). CVS mastery is considered a central instructional objective from a wide variety of educational perspectives (DeBoer, 1991; Duschl, 1990; Murnane & Raizen, 1988; National Research Council, 1995).

Chen and Klahr (1999) demonstrated that direct instruction on CVS led to statistically and educationally significant improvement in children's ability to design simple, unconfounded experiments. For children receiving direct instruction, mean performance on CVS increased in the psychology lab from 40% correct prior to instruction to 80% correct following instruction, whereas children in the discovery-learning condition showed no significant improvement. Moreover, students receiving direct instruction were superior to control students on a far-transfer test of experimental design administered 7 months later. When the direct-instruction procedure was adapted from an experimental script to a lesson plan implemented in a classroom setting, mean CVS performance increased from 30% prior to instruction to 96% following instruction (Toth, Klahr, & Chen, 2000).

However, the type of direct instruction used in these CVS studies has been criticized with respect to both its content and its epistemology (Chinn & Malhotra, 2001). The content critique is that CVS—as taught in these studies, as well as similar investigations of children's experimental skill (e.g., Germann, Aram, & Burke, 1996; Metz, 1985; Palincsar, Anderson, & David, 1993; Schauble, 1990; Zohar, 1995)—is a relatively circumscribed and inflexible procedure. The epistemological critique is that direct instruction in CVS, although apparently effective in improving students' CVS scores, does not provide them with a basis for exploring broader issues surrounding “authentic” scientific inquiry, such as detecting potential confounds and other validity challenges in complex experimental situations. According to these critiques, when children taught via direct instruction are faced with such complexities, they will show little transfer beyond the specific context in which they were taught.

The aim of the present study was to evaluate these critiques by investigating the relative effectiveness of direct instruction and discovery learning not only with respect to the acquisition of CVS, but also with respect to children's ability to reason in more authentic contexts. Our operational definition of authentic scientific reasoning is based on measures of children's ability to evaluate science-fair posters created by other children. We chose this transfer task because one important aim of science education—from elementary school through graduate training—is to increase students' ability to evaluate the soundness of other people's scientific endeavors (National Research Council, 1995) and to assess the validity of their scientific claims (Zimmerman, Bisanz, & Bisanz, 1998). The poster-evaluation task used in this study challenged children to reason about the quality of other children's research—to evaluate the design, measurements, data analysis, alternative hypotheses, and conclusions depicted in their posters.

We tested three hypotheses:

- *Direct instruction is more effective than discovery learning in teaching children CVS.* That is, we expected to replicate earlier comparisons of different instructional approaches to CVS training (Chen & Klahr, 1999; Klahr, Chen, & Toth, 2001; Toth et al., 2000).
- *When evaluating science-fair posters, children who have mastered CVS outperform those who have not.* This hypothesis is based on the expectation that mastery of the procedural and conceptual rationale for controlling variables in experimental settings provides children with a basis for generating a rich array of critiques about not only experimental design, but also other aspects of “good science,” such as the adequacy of the design, manipulations, measurements, inferences, and conclusions associated with an experimental study.
- *What is learned is more important than how it is taught.* This hypothesis led to the prediction that the association between CVS mastery and good performance on the poster-evaluation task would be independent of the learning path, that is, independent of whether that mastery was achieved under the direct-instruction condition or the discovery-learning condition. Conversely, we predicted that the association between poor CVS scores and poor poster-evaluation scores would be independent of the instructional condition under which children failed to achieve CVS mastery. The importance of this hypothesis is that it runs counter to one of the primary purported benefits of discovery learning: that it has a positive influence on long-term transfer.

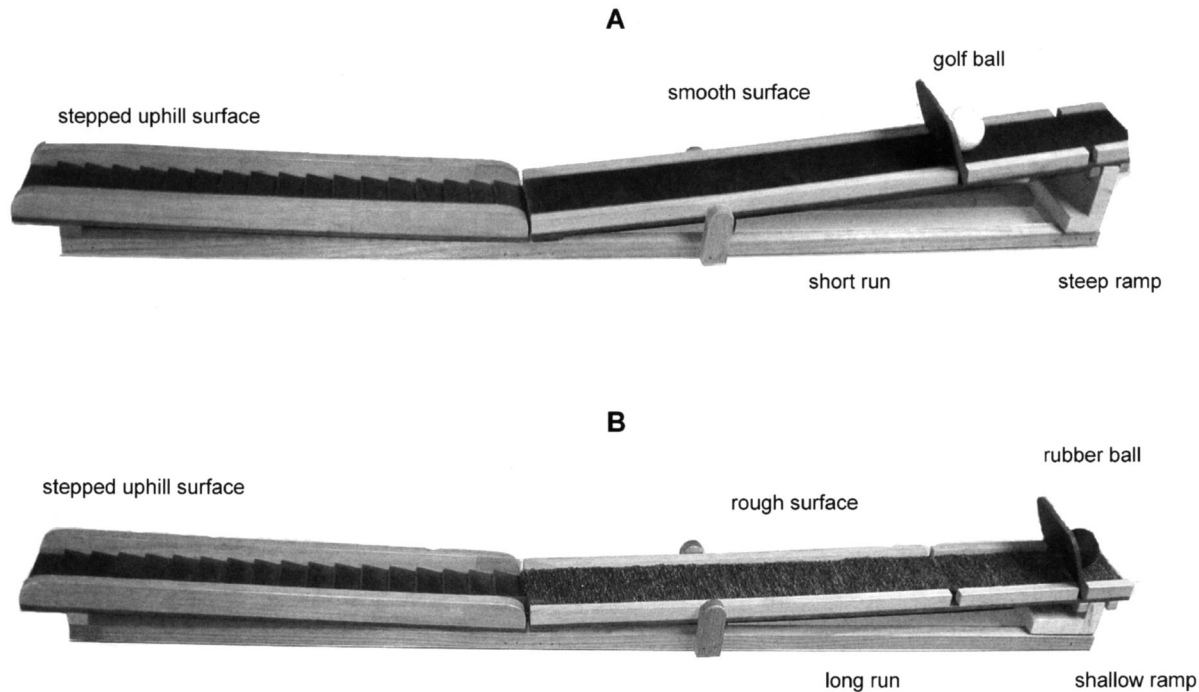
## METHOD

Participants were 112 third- and fourth-grade children in four different elementary schools, one of which was an all-girls school. There were 58 third graders (21 boys and 37 girls; mean age = 9 years, range: 8.4 years to 10.2 years) and 54 fourth graders (12 boys and 42 girls; mean age = 10 years, range: 9.3 years to 10.6 years). Children from both grades at each school were randomly assigned to either the direct-instruction condition or the discovery-learning condition.

Day 1 had two phases: exploration and assessment.<sup>2</sup> Throughout both phases and in both conditions, the children worked with the

<sup>1</sup>See Hammer (1997, p. 489) for a reasonable, but ambiguous, definition.

<sup>2</sup>The design, materials, and procedure for Day 1 are nearly identical to those described in our earlier work (Chen & Klahr, 1999; Toth et al., 2000).



**Fig. 1.** The ramps used during the exploration and assessment phases. On each of the two ramps, children could vary the steepness, surface, and length of the ramp, as well as the type of ball. The confounded experiment depicted here contrasts (a) a golf ball on a steep, smooth, short ramp with (b) a rubber ball on a shallow, rough, long ramp.

apparatus depicted in Figure 1. Materials included two wooden ramps, each with an adjustable downhill side and a slightly uphill, stepped surface on the other side, and two kinds of balls. The children could set the steepness of each ramp (high or low), the surface of the ramps (rough or smooth), and the length of the downhill run (long or short), and they could choose which type of ball (a rubber ball or a golf ball) to roll down each ramp. They were asked to make comparisons to determine how different variables affected the distance that balls rolled after leaving the downhill ramp. Figure 1 depicts a (confounded) experimental setup using these materials.

At the beginning of the exploration phase, the ramp apparatus was described, and then the children's baseline competence was assessed. They were asked to set up four experiments: two to determine the effect of steepness and two to determine the effect of run length on how far a ball rolls. Each child received a score indicating the number of unconfounded experiments he or she designed during this first part of the exploration phase.

What happened next depended on the child's training condition. Children in the direct-instruction condition observed as the experimenter designed several additional experiments—some confounded, and some unconfounded—to determine the effects of steepness and run length. For each experiment, the instructor asked the children whether or not they thought the design would allow them to “tell for sure” whether a variable had an effect on the outcome. Then the instructor explained why each of the unconfounded experiments uniquely identified the factor that affected the outcome, and why each confounded experiment did not. Children in the discovery condition instead continued to design their own experiments, focused on the same two variables that the direct-instruction children were focusing on, but without any instruction on CVS or any feedback from the experimenter.

It is important to note that in our operationalization, the difference between direct instruction and discovery learning does not involve a difference between “active” and “passive” learning. In both conditions, students were actively engaged in the design of their experiments and the physical manipulation of the apparatus. The main distinction is that in direct instruction, the instructor provided good and bad examples of CVS, explained what the differences were between them, and told the students how and why CVS worked, whereas in the discovery condition, there were no examples and no explanations, even though there was an equivalent amount of design and manipulation of materials.

In the assessment phase, which started immediately after the exploration phase, children in both conditions were asked to design four additional experiments: two to determine the effect of a factor that had been investigated earlier (run length) and two to determine the effect of a factor that had not been investigated earlier (surface). During the assessment phase, the experimenter did not provide any feedback in either condition.

The evaluation of science-fair posters took place on Day 2, about a week later. A different experimenter (blind to training condition) asked all children to evaluate two science-fair posters (based on real posters generated by sixth graders from another school) by making comments and suggestions that would help to make the poster “good enough to enter in a state-level science fair.” One poster explored the effect of the number of holes in a Ping-Pong ball on how far the ball traveled when launched from a catapult, and the other poster compared the short-term memory of boys and girls for a set of common objects. Both posters—one of which is depicted in Figure 2<sup>3</sup>—bore

<sup>3</sup>Digital images of both posters are available from the first author upon request.

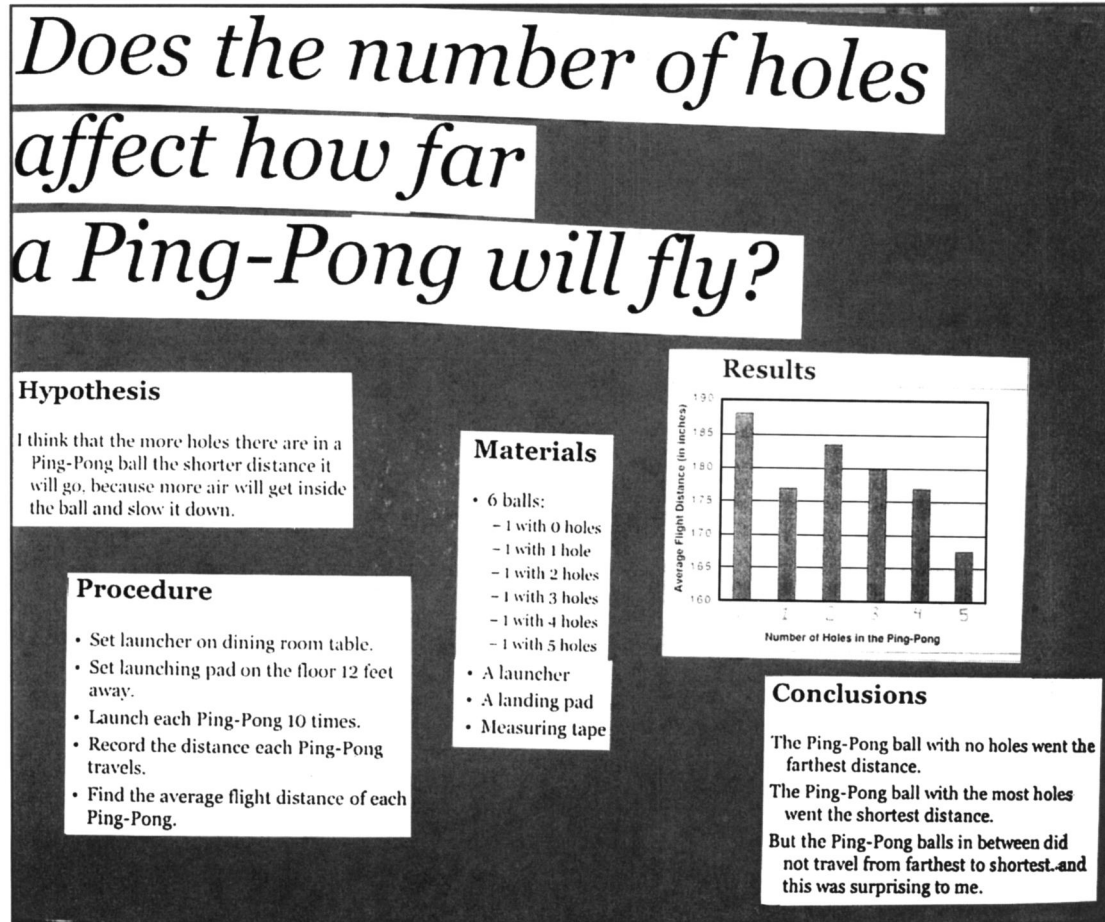


Fig. 2. Ping-Pong poster that children evaluated on Day 2.

titles stating the research question (i.e., “Who has a better memory? Boys or girls?”; “Does the number of holes affect how far a Ping-Pong will fly?”) and displayed brief descriptions of the hypothesis, procedure, materials, results (presented graphically), and conclusions. An important feature of these posters is that they described highly imperfect experiments, thus affording opportunities for wide-ranging evaluations.

Children were assessed individually, and the order of poster presentation was counterbalanced. As each poster was presented to a child, the experimenter read aloud all information on the poster, pointing out each section (title, hypothesis, materials, procedure, results, and conclusions) and explaining the graphical representation of the results. Following this general overview of the poster, the experimenter conducted a structured interview with increasingly specific probes asking the child to critique the poster, first at a general level and then with regard to specific elements depicted (materials, procedure, results, and conclusions). The main topics covered during the interview are listed in Table 1. The number of valid critiques was an open-ended measure, and the children were encouraged to say as much as they could think of, in addition to being asked to respond to the specific questions in the structured interview.

Children’s responses were transcribed from audiotape and coded independently by two coders. (There was 83% agreement between coders on what constituted a valid critique.) Scientific reasoning skills

were indexed according to the extent to which children commented about basic elements of experimentation—including research design, theoretical explanation, control of variables, measurement, statistical inferences, or conclusions (see Table 1). Probes about specific elements of experimentation were used in order to provide children with multiple opportunities to critique the posters, rather than to assess distinct knowledge. Therefore, the primary poster-evaluation score for each child was based on the total number of valid critiques about any aspect of either poster (i.e., a *grand poster score*).

## RESULTS

The overall aim of the analysis was to determine whether there were differences in the ability of children who did or did not acquire CVS, via either direct instruction or discovery, to reason about the broader and more diffuse domain of evaluating science-fair posters. There were two parts to the analysis. First, we examined the extent to which children in each condition learned CVS on Day 1. Second, we looked at the ability of children to assess the posters a week later, on Day 2.

### Day 1: CVS Exploration and Assessment

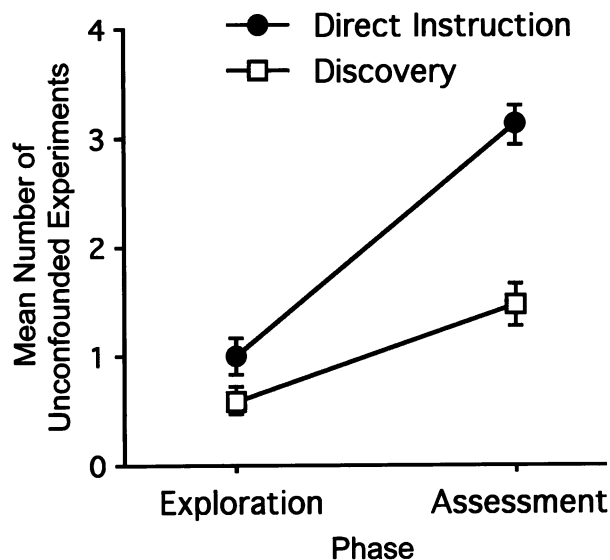
In both the exploration and the assessment phases, the children had four opportunities to design an unconfounded experiment. Thus, their

**TABLE 1**  
*Taxonomy for Scoring the Poster Evaluations*

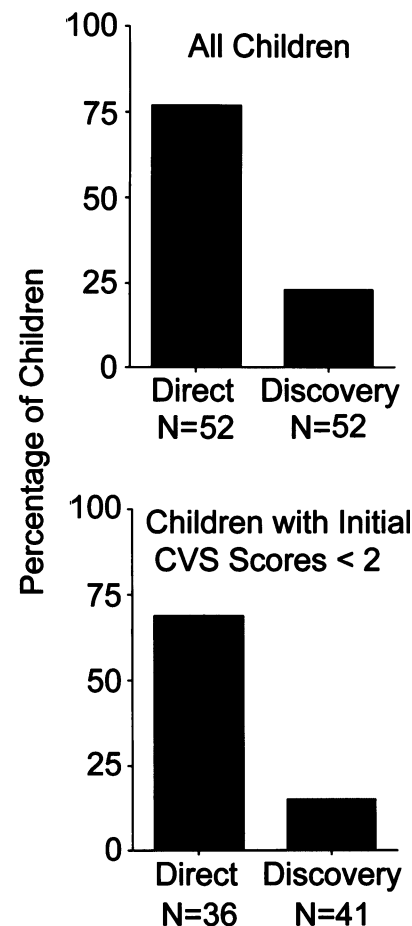
1. Adequacy of research design
Does the design support the question?
2. Theoretical explanation
Is a potential mechanism proposed?
Are mediating variables considered?
Is a plausible explanation for nonconforming data offered?
3. Controlling for confounding variables, elimination of other causes
Could the inference about the relationship between independent and dependent variables be incorrect?
Could factors other than those mentioned be causal?
4. Measurement
Are the measurements appropriate to the design and the research question?
5. Statistical inferences
Are the conclusions appropriate to the measures and the analysis?
6. Completeness of conclusion
Does the conclusion address the original research question?
Is the conclusion correctly based on the empirical results?

**Note.** Only the main categories are listed here. The complete coding protocol is available upon request. This taxonomy is adapted from Korpan et al. (1994).

scores in each phase could range from 0 to 4. (We categorized the 8 children who produced four out of four unconfounded experiments during the exploration phase as “CVS experts” and excluded them from the Day 1 analysis. The analysis presented here is based on the remaining 104 participants.) A 2 (training condition)  $\times$  2 (grade)  $\times$  2 (phase: exploration vs. assessment) analysis of variance (ANOVA) on CVS scores, with phase as a repeated measure, revealed main effects for training condition,  $F(1, 100) = 26.4, p < .0001, d = 0.711$ , and phase,  $F(1, 100) = 128.5, p < .0001, d = 1.11$ , and a phase-by-



**Fig. 3.** Mean number of unconfounded experiments in the exploration and assessment phases. Results are shown separately for the direct-instruction condition and the discovery-learning condition.



**Fig. 4.** Percentage of children (out of the total *ns* indicated) in the direct-instruction and discovery-learning conditions who achieved master-level performance (at least 3 unconfounded experiments out of 4) in the assessment phase. The top graph shows results for all children (excluding experts), and the bottom graph shows results for only those children with control-of-variables-strategy (CVS) scores less than 2 (out of 4) during the exploration phase.

training condition interaction,  $F(1, 100) = 23.0, p < .0001$ . There were no other main effects or interactions. CVS scores for direct-instruction children increased dramatically from exploration to assessment: from 1.0 to 3.1,  $t(51) = 10.5, p < .0001, d = 1.73$  (see Fig. 3). Scores for children in the discovery condition also increased, but much less: from 0.6 to 1.5,  $t(51) = 5.3, p < .0001, d = 0.725$ .

In order to examine the impact of the two training conditions on individual children, we classified each child according to his or her CVS score following training. We defined a CVS “master” as a child who designed at least three unconfounded experiments (out of four experiments) during the assessment phase. As shown in the top panel of Figure 4, direct instruction produced many more masters than did discovery learning. Forty of the 52 direct-instruction children (77%) became masters, whereas only 12 of the 52 discovery children (23%) did so,  $\chi^2(1, N = 104) = 28.0, p < .0001$  (with Yates correction). The superiority of direct instruction over discovery was maintained when we looked only at the children who started out with the lowest initial CVS scores (0 or 1) in the exploration phase (see Fig. 4, bottom panel): Sixty-nine percent of the 36 children in the direct-instruction

**TABLE 2**  
**Poster-Evaluation Scores for Children Following Different Learning Paths**

Learning path	Grand poster score		Modified grand poster score		n
	Mean	SE	Mean	SE	
Expert	15.1	1.4	8.1	1.4	8
Master (discovery)	13.8	1.2	9.3	0.9	12
Master (direct instruction)	12.5	1.0	7.5	0.6	40
Nonmaster (discovery)	9.1	0.7	5.8	0.5	40
Nonmaster (direct instruction)	8.2	1.3	5.4	1.0	12

**Note.** The modified grand poster score is the grand poster score with comments related to the control-of-variables strategy removed.

condition who received such low scores during the exploration phase became masters by the assessment phase, compared with only 15% of the 41 children in the discovery condition with equally low initial CVS scores,  $\chi^2(1, N = 76) = 20.8, p < .0001$  (with Yates correction).

#### Day 2: Poster Evaluations

Having established that there was a very strong immediate effect of training condition on CVS scores in the assessment phase on Day 1, we next addressed our central question: Did children's grand poster scores depend on their learning paths? We defined learning paths by crossing training condition (direct instruction or discovery) with the master or nonmaster classification. This produced four learning paths: "master (discovery)," "nonmaster (discovery)," "master (direct instruction)," and "nonmaster (direct instruction)." For this analysis, we also included, as a fifth category, the 8 children who were "natural" experts in the exploration phase.<sup>4</sup>

A one-way ANOVA with learning path as the independent variable and grand poster score as the dependent variable yielded a main effect for learning path,  $F(4, 107) = 5.2, p = .0007$ . Table 2 shows that experts, discovery masters, and direct-instruction masters had higher grand poster scores than discovery nonmasters and direct-instruction nonmasters,  $d = 0.83$ .<sup>5</sup> Moreover, there were no statistical differences among the grand poster scores for experts, discovery masters, and direct-instruction masters. Post hoc Fisher PLSD (protected least significant difference) tests produced significant pair-wise differences ( $p < .015$ ) between the two nonmaster paths and the other three paths, and no differences among the experts and the two master paths, or between the two nonmaster paths.<sup>6</sup> In other words, the grand poster scores of the three groups that achieved high levels of CVS performance via different paths were indistinguishable. The grand poster scores of the two groups who failed to master CVS were also indis-

tinguishable, but significantly lower than the scores of the three high-CVS-score groups.

The grand poster score was based on all valid critiques, including those involving CVS-related issues. Thus, although this score provides an overall picture of children's performance on the poster-evaluation task, it is not as stringent a measure of transfer as a poster score excluding all CVS-related comments. Therefore, in another analysis, we examined transfer by subtracting subscores for CVS-related comments from the grand poster scores. A one-way ANOVA with learning path as the independent variable and grand poster score minus CVS-related score as the dependent variable yielded a main effect for learning path,  $F(4, 107) = 3.1, p = .017$ . The results, shown in Table 2, replicated the pattern for the grand poster score: Experts and masters outperformed nonmasters, regardless of the learning paths.

#### DISCUSSION

With respect to the focal skill of designing unconfounded experiments in simple contexts, these results replicate other studies in which direct instruction was clearly superior to discovery learning in facilitating children's acquisition of CVS (Chen & Klahr, 1999; Klahr et al., 2001). These results also indicate that discovery learning does produce significant—albeit much smaller—gains: A nontrivial proportion (23%) of discovery-learning children became CVS masters (see Fig. 4, top panel). And even among the discovery-learning children receiving the lowest scores in the exploration phase, 15% became masters (see Fig. 4, bottom panel). Thus, two questions for further research are (a) whether these proportions are stable across different populations of learners and (b) whether there are specific features of some learners that render discovery learning effective.

The most important result of this study is the relationship between learning paths and transfer. Children who became masters via direct instruction were as skilled at evaluating science-fair posters as were discovery-learning masters and experts. Similarly, children who failed to become masters did equally poorly on the poster-evaluation task regardless of training condition. That is, the focused, explicit, and didactic training in the direct-instruction condition produced a high proportion of CVS masters who were as proficient as the few discovery-learning masters (and experts) when subsequently asked to demonstrate richer, more authentic, scientific judgments.

These results suggest the need to reexamine the long-standing claim that the limitations of direct instruction, as well as the advantages of discovery methods, will invariably manifest themselves in tasks requiring broad transfer to authentic contexts (e.g., "... learning under external reinforcement ... produces either very little change in logical thinking or a striking momentary change with no real comprehension"—Piaget, 1970, p. 714). The aim of such an analysis would be to generate an empirically sound basis for determining the most effective matches between topic, student, and type of pedagogy. Such results could provide evidence-based guidance to teachers for achieving a balanced portfolio of instructional approaches to early science instruction.

**Acknowledgments**—This work was supported, in part, by grants to the first author from the McDonnell Foundation's Cognitive Studies for Educational Practice program and from the National Institute of Child Health and Human Development (HD25211) and the National Sci-

<sup>4</sup>In the assessment phase, experts again developed unconfounded designs for all four experiments.

<sup>5</sup>This effect size is based on the difference in mean grand poster scores for the 52 children in the nonmaster groups versus the 60 children in the expert and master groups.

<sup>6</sup>A more conservative post hoc test—Games-Howell (Games & Howell, 1976), which is sensitive to both unequal *ns* and unequal variances—produced the same pattern of results, although two of the six significant pair-wise contrasts dropped just below significance.

ence Foundation. We thank the following colleagues for their contributions at various stages of the project, from conception, to execution, to commenting on earlier drafts: Sharon Carver, Norma Chang, Zhe Chen, Kevin Dunbar, Amy Masnick, Brad Morris, Audrey Russo, Jen Schnakenberg, Chris Schunn, Anne Siegel, Lara Triona, Jolene Watson, and Corrine Zimmerman.

## REFERENCES

- Anderson, J.R., Corbett, A.T., Koedinger, K., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *Journal of the Learning Sciences*, 4, 167–207.
- Bredderman, T. (1983). Effects of activity-based elementary science on student outcomes: A quantitative synthesis. *Review of Educational Research*, 53, 499–518.
- Brophy, J., & Good, T.L. (1986). Teacher behavior and student achievement. In M.C. Wittrock (Ed.), *Handbook of research on teaching* (pp. 328–375). New York: Macmillan.
- Chen, Z., & Klahr, D. (1999). All other things being equal: Children's acquisition of the control of variables strategy. *Child Development*, 70, 1098–1120.
- Chinn, C.A., & Malhotra, B.A. (2001). Epistemologically authentic scientific reasoning. In K. Crowley, C.D. Schunn, & T. Okada (Eds.), *Designing for science: Implications from everyday, classroom, and professional settings* (pp. 351–392). Mahwah, NJ: Erlbaum.
- DeBoer, G.E. (1991). *A history of ideas in science education: Implications for practice*. New York: Teachers College Press.
- Duschl, R.A. (1990). *Restructuring science education*. New York: Teachers College Press.
- Games, P.A., & Howell, J.F. (1976). Pairwise multiple comparison procedures with unequal n's and/or variances: A Monte Carlo study. *Journal of Educational Statistics*, 1, 113–125.
- Germann, P.J., Aram, R., & Burke, G. (1996). Identifying patterns and relationships among the responses of seventh-grade students to the science process skill of designing experiments. *Journal of Research in Science Teaching*, 33, 79–99.
- Hammer, D. (1997). Discovery learning and discovery teaching. *Cognition and Instruction*, 15, 485–529.
- 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., Chen, Z., & Toth, E.E. (2001). From cognition to instruction to cognition: A case study in elementary school science instruction. In K. Crowley, C.D. Schunn, & T. Okada (Eds.), *Designing for science: Implications from everyday, classroom, and professional settings* (pp. 209–250). Mahwah, NJ: Erlbaum.
- Klahr, D., & Simon, H.A. (1999). Studies of scientific discovery: Complementary approaches and convergent findings. *Psychological Bulletin*, 125, 524–543.
- Korpan, C.A., Bisanz, G., Dukewich, T., Robinson, K., Bisanz, J., Thibodeau, M., Hubbard, K., & Leighton, J. (1994). *Assessing scientific literacy: A taxonomy for classifying questions and knowledge about scientific research* (Tech. Rep. 94-1). Edmonton, Alberta, Canada: University of Alberta, Centre for Research in Child Development.
- Loveless, T. (1998). The use and misuse of research in educational reform. In D. Ravitch (Ed.), *Education policy* (pp. 285–286). Washington, DC: Brookings Institution Press.
- Mayer, R.E. (2004). Should there be a three-strikes rule against pure discovery learning? *American Psychologist*, 59, 14–19.
- McDaniel, M.A., & Schlager, M.S. (1990). Discovery learning and transfer of problem-solving skills. *Cognition and Instruction*, 7, 129–159.
- Metz, K.E. (1985). The development of children's problem solving in a gears task: A problem space perspective. *Cognitive Science*, 9, 431–471.
- Murnane, R.J., & Raizen, S.A. (Eds.). (1988). *Improving indicators of the quality of science and mathematics education in grades K-12*. Washington, DC: National Academy Press.
- National Research Council. (1995). *National science education standards*. Washington, DC: National Academy Press.
- Palincsar, A.S., Anderson, C., & David, Y.M. (1993). Pursuing scientific literacy in the middle grades through collaborative problem solving. *Elementary School Journal*, 93, 643–658.
- Piaget, J. (1970). Piaget's theory. In P. Mussen (Ed.), *Carmichael's manual of child psychology* (Vol. 1, pp. 703–772). New York: John Wiley & Sons.
- Rosenshine, B., & Stevens, R. (1986). Teaching functions. In M.C. Wittrock (Ed.), *Handbook of research on teaching* (pp. 376–391). New York: Macmillan.
- 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, 102–119.
- Stohr-Hunt, P.M. (1996). An analysis of frequency of hands-on experience and science achievement. *Journal of Research in Science Teaching*, 33, 101–109.
- Toth, E.E., Klahr, D., & Chen, Z. (2000). Bridging research and practice: A cognitively-based classroom intervention for teaching experimentation skills to elementary school children. *Cognition and Instruction*, 18, 423–459.
- Winch, W.A. (1913). *Inductive versus deductive methods of teaching*. Baltimore: Warwick and York.
- Zimmerman, C., Bisanz, G.L., & Bisanz, J. (1998). Everyday scientific literacy: Do students use information about the social context and methods of research to evaluate news briefs about science? *Alberta Journal of Educational Research*, 44, 188–207.
- Zohar, A. (1995). Reasoning about interactions between variables. *Journal of Research in Science Teaching*, 32, 1039–1063.

(RECEIVED 8/5/03; REVISION ACCEPTED 11/12/03)