The equivalence of learning paths in early science instruction: effects of direct instruction and discovery learning

(To appear in *Psychological Science*, 2004)

David Klahr[1]
Department of Psychology
Carnegie Mellon University

Milena Nigam
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 to 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 claim that "each time one prematurely teaches a child something he could have discovered for himself the 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 (Alliance for Childhood, 2000; 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 to the general enthusiasm for discovery learning come from studies demonstrating that teacher-centered methods using direct instruction are highly effective (c.f., Brophy & Good, 1986; Rosenshine & Stevens, 1986). In particular, direct instruction has been shown to be an efficient way to teach procedures that are difficult for students to discover on their own, such as those involved in geometry, algebra, and computer programming (Anderson, Corbett, Koedinger, & Pelletier, 1995; Klahr & Carver, 1988). Finally, the theoretical basis for the predicted superiority of discovery over direct instruction is vague and, in many cases, inconsistent with much of the literature on learning and memory. For example, in most cases children in discovery situations are more likely than those receiving direct instruction to encounter inconsistent or misleading feedback, encoding errors, causal misattributions, and inadequate practice and elaborations. These severe cognitive disadvantages may overwhelm the increased motivational aspects that are commonly attributed to discovery learning. In sum, we question the widely accepted view that discovery learning usually trumps direct instruction.

However, our aim in this study is to go beyond simple comparisons of whether children are more likely to learn from one type of instruction than from another. In addition to a direct comparison of the effectiveness of two radically different types of instruction, we test the prediction that *if* children achieve mastery of a new procedure -- via either instructional condition -- *then* the way that they reached that mastery has no effect on their ability to transfer what they have learned. This "path independent transfer" prediction, if supported, has implications for the discovery learning approach,

[1] Klahr@cmu.edu

because one of its purported advantages is that it has long-term benefits on how children ultimately apply what they have leaned via the discovery approach – benefits that justify its admittedly lower efficiency.

In order to evaluate this hypothesis, we must create exemplars of each approach, expose children to them, and then challenge learners with an appropriate transfer task. However, at the outset, we face a difficult definitional problem because nearly 100 years of research (cf., Winch, 1913) have yet to produce a consistent definition of *discovery learning*[2]. Therefore, in this study, 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 is no teacher intervention beyond the suggestion of a learning objective: no guiding questions and no feedback about the quality of the child's selection of materials, explorations, or self-assessments. Correspondingly, we use an extreme type of direct instruction in which the goals, the materials, the examples, the explanations, and the pace or instruction are all teacher controlled[3].

The specific context in which we contrasted these two instructional approaches is 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; NSES, 1995).

Chen & Klahr (1999) demonstrated that direct instruction on CVS led to statistically and educationally significant improvement in children's ability to design simple, unconfounded, experiments. Mean performance on CVS increased from 40% correct prior to instruction to 80% correct following instruction in the psychology lab and from 30% to 96% in a classroom setting (Toth, Klahr, & Chen, 2000). Moreover, students receiving Direct Instruction were superior to controls on a far transfer test of experimental design administered seven months later (Chen & Klahr, 1999).

However, the type of direct instruction used in the CVS studies cited above has been criticized with respect to both its content and its epistemology (Chinn & Malhotra, 2001). The content critique argues that CVS -- as taught in those studies, as well as similar investigations of children's experimental skill (e.g., Germann, Aram, & Burke, 1996; Metz, 1985; Palinscar, Anderson, & David, 1993; Schauble, 1990; Zohar, 1995) -- is a relatively circumscribed and inflexible procedure. The epistemological critique suggests 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, children taught via direct instruction will show little transfer beyond the specific context in which they are taught when faced with such complexities.

The aim of the present study is 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 measure of "authentic scientific reasoning" is based on 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 teach students how to evaluate the soundness of others' scientific endeavors (NSES, 1995). While only a small proportion of students will become practicing scientists, it is important for citizens to acquire the skills necessary to assess the validity of 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

---

2 See Hammer, 1997, p. 489 for a reasonable, but ambiguous, definition.

3 The term "direct instruction" is sometimes used to refer to a specific pedagogical approach to early reading education advocated by S. Engelmann and his colleagues (Adams & Engelmann, 1996). We use the term in its broader, generic sense, exemplified by the procedures described later in this paper.

research, including assessment of the design, measurements, data analysis, alternative hypotheses, and conclusions depicted in their posters.

Our three hypotheses are as follows:

1. Direct instruction will be more effective than discovery learning in the "local" acquisition of CVS procedures. This is simply a prediction that we will replicate earlier comparisons of different instructional approaches to CVS training (Chen & Klahr, 1999, Klahr, Chen & Toth, 2001; Toth, Klahr, & Chen, 2000).

2. When evaluating science fair posters, children who have mastered CVS will outperform those who have not. This prediction 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.

3. What is learned is more important than how it is taught. This hypothesis leads to the prediction that the association between CVS mastery and high scores on the poster assessment task will be independent of the "learning path", that is, whether that mastery was achieved under the Direct Instruction condition or the Discovery Learning condition. Conversely, poor CVS scores are predicted to be associated with poor poster evaluation sores, independent of the training condition under which children fail 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.

Procedure and Method

Participants included 112 children from third and fourth grades 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 yrs. to 10.2 yrs.) and 54 fourth-graders (12 boys and 42 girls, mean age 10 years, range 9.3 yrs. to 10.6 yrs.). 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[4]. At the outset of the Exploration phase, prior to any instruction, children's baseline CVS competence was determined. Next, children in the Direct Instruction condition were provided with detailed instruction about how to design an unconfounded experiment while children in the Discovery Learning condition were allowed to design experiments on their own, without any guidance or feedback from the experimenter. The Assessment Phase followed immediately.

Throughout both phases and in both conditions, children worked with the 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. 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 set up using these materials.

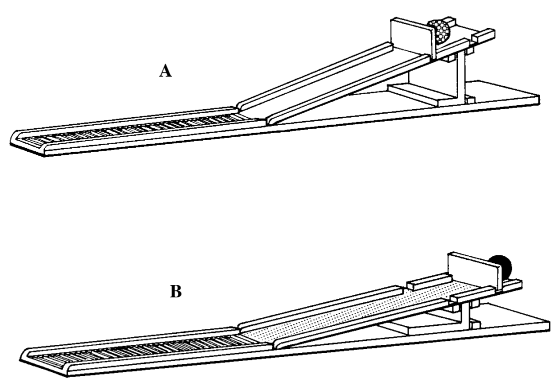

Figure 1. The Ramps used during Phase 1. On each of the two ramps, children can vary the angle of the ramp, the surface of the ramp, the length of the ramp, and the type of ball. The confounded experiment depicted here contrasts (a) the golf ball on the steep, smooth, short ramp with (b) the rubber ball on a shallow, rough, long ramp

At the beginning of the Exploration phase, the ramp apparatus was described, and children were asked to set up four experiments: two to determine the

4 The design, materials, and procedure for Day 1 are nearly identical to those described in our earlier work (Chen and Klahr, 1999; Toth, Klahr, & Chen, ,2000)

effect of steepness and two to determine the effect of run length on how far a ball rolls. At the end of the Exploration phase, children in the Direct Instruction condition observed as the Experimenter designed several additional experiments – some confounded, and some unconfounded -- to make these determinations. 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 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 the difference between our operationalization of Direct Instruction and Discovery Learning is not 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, children in both conditions were asked to design four additional experiments: one to determine the effect of a factor that had been investigated earlier (run length) and one 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 suggestion 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 -- depicted in Appendix A – bore 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 texts describing the hypothesis, procedure, materials, results (presented graphically) and conclusions from the study. An important feature of these posters is that they each 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 the child, the experimenter read aloud all information on the poster, pointing out each section (title, hypothesis, materials/subjects and procedure, results, and conclusion) and explaining the graphical representation of the results. Following this general overview of the poster, a structured interview with increasingly specific probes asked children to critique the poster, first at a general level and then with regard to specific elements depicted on the poster (subjects and materials, procedure, results, and conclusions). The main topics covered during the interview are listed in Appendix B. The number of valid critiques was an open-ended measure, and 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 audio tape and coded independently by two coders. 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 conclusion (see Appendix B). Probes about specific elements of experimentation were used in order to provide children with multiple opportunities to critique a poster, 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 is to determine whether there are differences in the ability of children who do or do not acquire CVS, via either direct instruction or discovery, to reason about the broader and more diffuse domain of evaluating science fair posters. There are two parts to the analysis. First we examine the extent to which children in each condition learned CVS on Day 1. Second, we look at the ability of children to assess the posters, a week later, on Day 2.

Day 1. CVS Exploration scores and Assessment scores

In both the Exploration and the Assessment phases, children had four opportunities to design an unconfounded experiment. Thus, their scores in each phase could range from zero to four. (We defined CVS "Experts" as children who produced four out of four unconfounded experiments during the Exploration phase and excluded them from the Day 1 analysis. There were eight such Experts, so the analysis presented here is based on the remaining 104 participants.) A 2 (training condition) x 2 (grade) x 2 (Phase: Exploration & Assessment) ANOVA on CVS scores, with phase as a repeated measure, revealed main effects for training condition $\underline{F}(1,100) = 26.4$, $p < .0001$, $d = .711$, and phase, $\underline{F}(1,100) = 128.5$, $p < .0001$, $d = 1.11$, and a phase by 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 Figure 2.) Scores for children in the Discovery condition also increased, but much less: from 0.6 to 1.5, $t(51) = 5.3$, $p < .0001$, $d = .725$

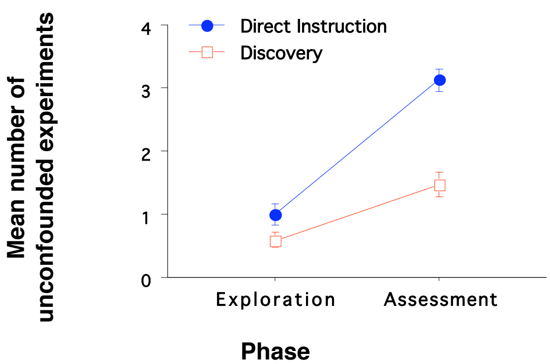

Figure 2. Mean CVS scores in Exploration and Assessment phases for Direct Instruction group and Discovery group

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 3 out of 4 unconfounded experiments during the Assessment phase. As shown in Figure 3a, Direct Instruction produced many more Masters than did Discovery Learning. Forty of the 52 Direct Instruction children (77%) became Masters, while only 12 of the 52 Discovery children (23%) did so, Chi Sq. (1df) = 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 Figure 3b): 69% of the 35 children in the Direct Instruction condition who received such low scores during the Exploration phase became Masters by the Assessment phase, compared to only 15% of the 41 children in the Discovery condition with equally low initial CVS scores (Chi Sq. (1df) = 20.8, p < .0001, with Yates correction).

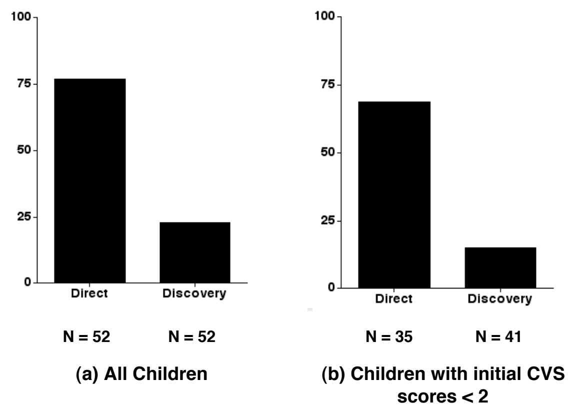

Figure 3. Percent of children in each training condition who achieved Master level performance (at least 3 out of 4 unconfounded experiments) in the Assessment Phase. (a) All children. (b) Only children with lowest scores during Exploration Phase.

Day 2: Science fair poster evaluations

<u>Relation between training condition, CVS skill level, and poster evaluation score.</u>

Having established that there was a very strong effect of training condition on the immediately tested CVS scores (in the Assessment phase on Day 1), we next addressed the central question of this paper: do children's Grand Poster Scores depend on their learning paths? We defined a learning path by crossing training condition (Direct Instruction or Discovery) with the Master or Non-Master classification. This produced four types of learning paths: "Master (Discovery)," "Non-Master (Discovery)," "Master (Direct Instruction)," and "Non-Master (Direct Instruction)." For this analysis we also included, as a fifth category, the eight children who were "natural" Experts in the Exploration

Table 1 Poster evaluation scores for children following different learning paths

Learning Path	(a) Grand Poster Score		(b) Grand Poster Score – CVS Comments Score		
	Mean	Std. Err.	Mean	Std. Err.	N
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
Non-Master (Discovery)	9.1	0.7	5.8	0.5	40
Non-Master (Direct Instruction)	8.2	1.3	5.4	1.0	12

phase.[5] A one-way ANOVA with learning path as the independent variable and Grand Poster Score as the dependent variable yielded a main effect for learner type, F (4, 107) = 5.2, p = .0007. Table 1a shows that Masters – as well as Experts – outperformed Non-masters, regardless of the learning path, *d* = .83[6]. Moreover, there were no statistical differences among the poster evaluation scores for Discovery Masters, Direct Instruction Masters, and Experts. Post-hoc Fisher PLSD tests produced significant pair-wise differences (p < .015) between the Non-Master paths of both types and all the other categories, and no differences among the Experts and the two types of Masters, nor between the two types of Non-Masters[7]. In other words, the Grand Poster Scores for the three groups that had achieved high levels of CVS performance via different paths were indistinguishable. Similarly, the Grand Posters Scores for the two groups who failed to master CVS were also indistinguishable, but significantly lower than those of the three high-CVS scoring groups.

The Grand Poster Score includes, as a component, the sub-scores based on children's comments about CVS-related issues. Thus, although it 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 that excludes all such comments. Therefore we repeated the analysis based on the Grand Score minus the CVS component scores. A one-way ANOVA with learning path as the independent variable and Grand Poster Score – CVS Related Score as the dependent variable yielded a main effect for learner type, F (4, 107) = 3.1, p = .017. The results, shown in Table 1b, replicate the pattern described above for the Grand Poster score: Experts and Masters outperform Non-Masters, regardless of the learning paths.

5 Experts continued to design four out of four unconfounded experiments in the Assessment phase as well.

6 For the difference in mean grand poster scores for the 52 children in the non-master groups *vs* the 60 children in the Expert and Master groups.

7 A more conservative post-hoc test – Games-Howell (Games & Howell, 1976) – which is sensitive to both unequal N's and variances produced the same pattern of results, although two of the six significant pair-wise contrasts dropped just below significance.

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 children's acquisition of CVS. This was true not only for the full sample of children in this study, but also for those who were initially the most in need of instruction.

While a substantial proportion of children in the Direct Instruction condition (77%) learned CVS, these results also show that non-trivial proportion (23%) of those in the Discovery condition became CVS Masters (see Figure 3a). And even among the children receiving the lowest scores in the Exploration phase, 15% became Masters in the Discovery condition (see Figure 3b). Thus, two questions for further research are (a) whether this proportion is stable across other populations of learners and (b) whether there are specific features of some learners that render Discovery Learning effective. Not only might specific learner characteristics favor either direct instruction or discovery learning, but also the nature of the learning goal is likely to interact with instructional type. In the case of a multi-step procedure like CVS, discovery learning is at a clear disadvantage, because there is no feedback from a confounded experiment about which step(s) led to the confound. CVS is but one of several types of procedural knowledge -- or "process skills"-- that are deemed central to early science instruction, and is important to see whether other procedures are also better taught via direct instruction than by discovery.

We believe that the most important result of this study is the relationship between learning paths and transfer. Children who became CVS Masters via Direct Instruction were as skilled at evaluating science fair posters as were Discovery Masters and Experts. Similarly, children who failed to become Masters did equally poorly on the poster evaluation task, regardless of training condition. That is, with respect to poster evaluation scores, children's learning paths were irrelevant. In other words, 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 Masters (and Experts) when subsequently asked to demonstrate richer, more authentic, scientific judgments.

These results suggest a re-examination of the long-standing claim that the limitations of direct instruction, as well as the advantages of discovery methods, will manifest themselves in 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 mixes and the most suitable matches between topic and student for different types of pedagogy. Such results could provide evidence-based guidance to teachers for achieving a balanced portfolio of instructional approaches to early science instruction.

Author note

This work was supported, in part, by grants to the first author from the McDonnell Foundation's CSEP program and from NICHHD (HD25211) and the National Science 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

Adams, G. L., & Engelmann, S. (1996). *Research on direct instruction: 25 years beyond DISTAR*. Seattle: Educational Achievement Systems.

Alliance for Childhood. (2000). *Fool's gold: A critical look at computers in childhood.* Retrieved from http://www.allianceforchildhood.net/projects/computers/computers_reports.htm

American Association for the Advancement of Science. (2002). Project 2061, Update 2001-2002. Retrieved from: http://www.project2061.org/newsletter/default.htm#update.

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 Wittrock, M. C. (Ed.), *Handbook of research on teaching*. 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*. 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* (4), 485-529.

Kennedy, D. (2001). College science: Pass, no credit. *Science 293*, 1557.

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. & Simon, H. A. (1999) Studies of scientific discovery: Complementary approaches and convergent findings. *Psychological Bulletin, 125*(5), 524-543.

Klahr, D, Chen, Z., & Toth, E. E. (2001) From cognition to instruction to cognition: A case study in elementary school science instruction . In Crowley, K., Schunn, C.D., & Okada, T. (Eds.), *Designing for science: implications from professional, instructional, and everyday science*. Mahwah, NJ: Erlbaum

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. Report 94-1). Edmonton, Alberta, CA: University of Alberta, Centre for Research in Child Development.

Loveless, T. (1998) The use and misuse of research in educational reform. In Ravitch, D. (Ed.), *Education Policy* (pp. 285-286) Washington, DC: Brookings Institution Press.

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, D. C.: National Academy Press.

National Science Education Standards (NSES). (1995). 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 Mussen, P. (Ed.) *Carmichael's manual of child psychology* (vol. 1). New York: John Wiley & Sons

Rosenshine, B. & Stevens, R. (1986) Teaching functions. In Wittrock, M. C. (Ed.), *Handbook of research on teaching*. 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.

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

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*(4), 423-459.

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, XLIV*, 188-207.

Winch, W. A. (1913) *Inductive versus deductive methods of teaching.* Baltimore: Warwick and York.

Zohar, A. (1995). Reasoning about interactions between variables. *Journal of Research in Science Teaching, 32*, 1039-1063

Appendix A: Posters used on Day 2 (Poster Evaluation Phase)

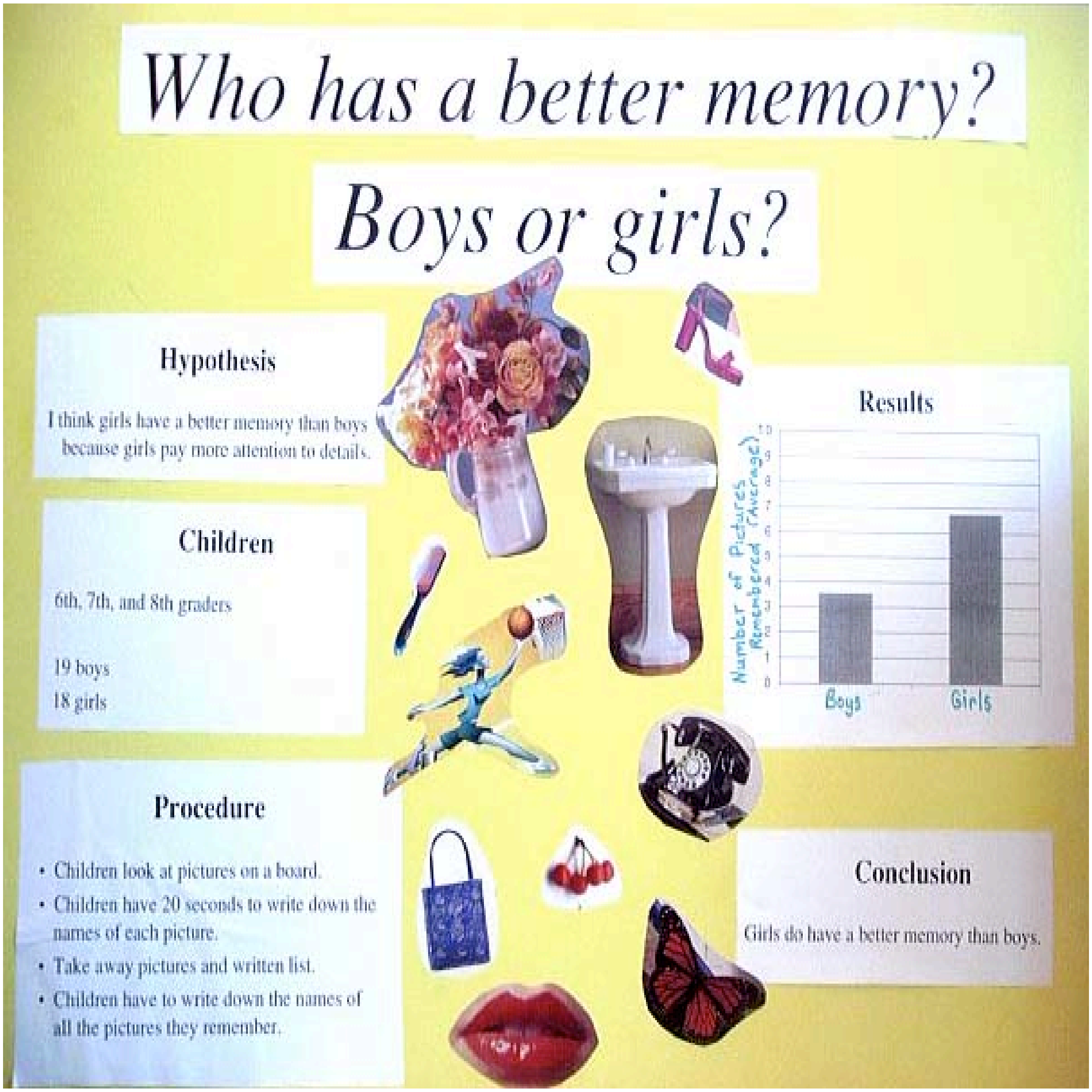

Figure A-1 Memory Poster used on Day 2.

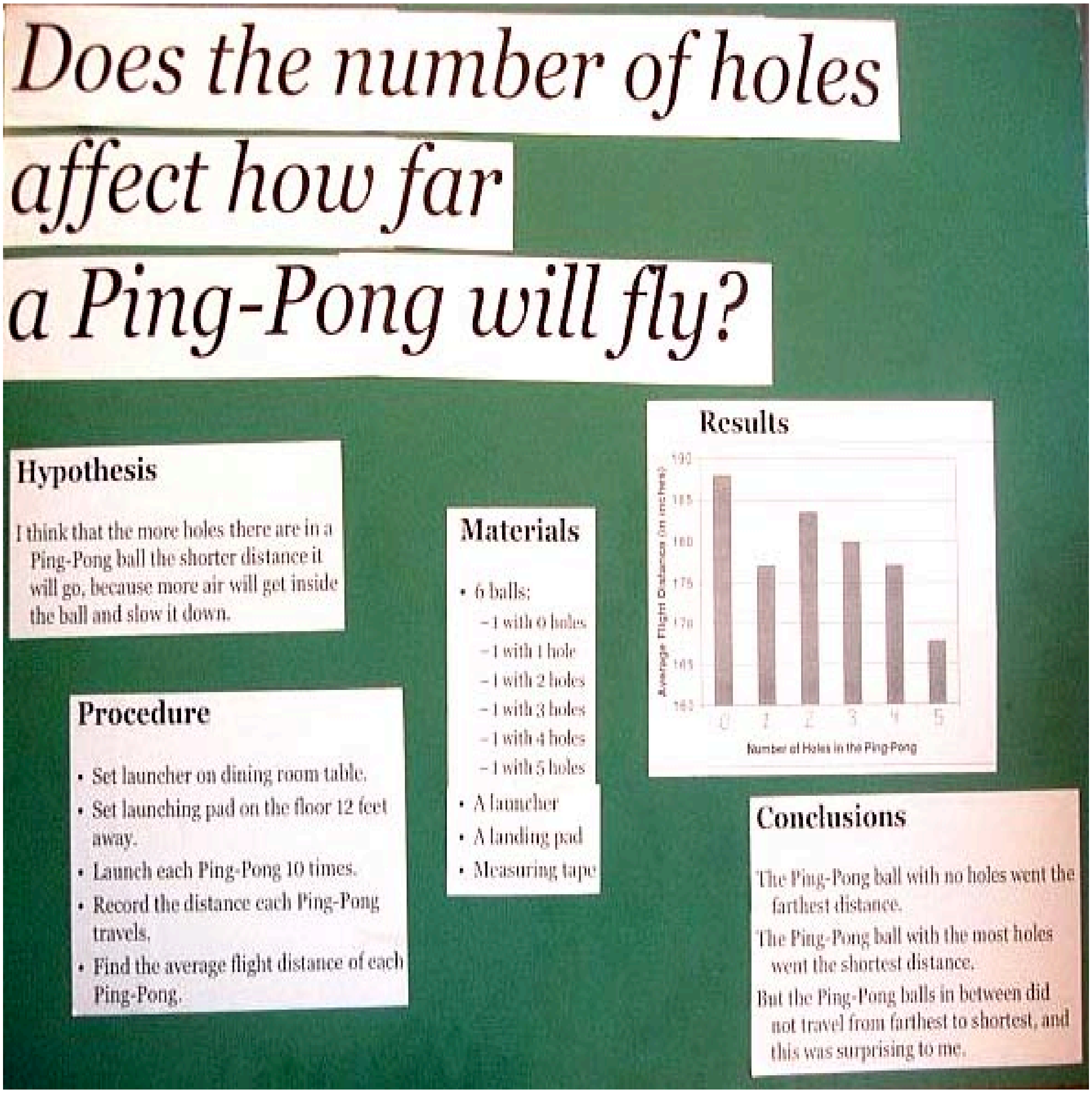

Figure A-2 Ping-Pong Poster used on Day 2.

Appendix B Poster Evaluation Scoring Taxonomy[8]

1. Adequacy of research design

Does the design support the question?

2. Theoretical explanation

Seeks to establish description of mechanism; asks “why;” mediating variables. Offers plausible explanation for nonconforming data.

3. Controlling for confounding variables/elimination of other causes

Inferences about relationship between agent and outcome could be wrong; other factors could cause outcome.

4. Measurement

Reflects that measures are not always appropriate.

5. Statistical inferences

6. Completeness of conclusion

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