COGNITION AND INSTRUCTION, *21*(2), 149–173

Point and Click or Grab and Heft: Comparing the Influence of Physical and Virtual Instructional Materials on Elementary School Students' Ability to Design Experiments

Lara M. Triona and David Klahr
Carnegie Mellon University

The widespread availability of computers in elementary schools makes them an appealing option for presenting instructional materials in laboratory science. However, there are neither theoretical nor empirical grounds for predicting whether virtual or physical presentation of instructional materials will be more effective. The definition of "active manipulation" is poorly specified and there are few studies that directly compare the two approaches unaccompanied by other potential confounds. In this study, 4th- and 5th-grade children were taught how to design simple unconfounded experiments using 1 of 2 instructional methods differing only in whether children manipulated physical or virtual materials. The 2 types of materials were equally effective in achieving several instructional objectives, including the design of unconfounded experiments, the derivation of correct predictions from them, and explicit reference to the need for experiments to be unconfounded.

The increasingly widespread availability of computers in elementary schools makes them an appealing option for presenting instructional materials in laboratory science. U.S. public schools have an average of one instructional computer for every four students (Education Week, 2002). Of the more than 700 instructional software titles recommended by Technology and Learning (2002), hundreds are oriented specifically toward science education. There are many obvious advantages to the use of computer-based instructional materials, especially for

Requests for reprints should be sent to Lara M. Triona, Department of Psychology, Carnegie Mellon University, Pittsburgh, PA 15213-3890. E-mail: triona@cmu.edu

laboratory science instruction. These include portability, safety, cost-efficiency, minimization of error, amplification or reduction of temporal and spatial dimensions, and flexible, rapid, and dynamic data displays.

In contrast to both the popularity and potential advantages of computer-based instruction are claims that it is detrimental not only to the achievement of specific instructional objectives, but also to broader educational goals ranging from brain development to social development. In the popular media (e.g., Alliance for Childhood, 2000), one can find both pragmatic and theoretical arguments against the use of computer-based instruction (Armstrong & Casement, 1998; Healy, 1999). The claim is that such instruction is likely to be ineffective at best, and harmful at worst. Proponents of this anticomputer view claim that because manipulation of physical materials is essential for learning, moving instruction to computers deprives children of these essential experiences. Articles addressed to teachers often describe the benefits of concrete manipulation (e.g., Berk, 1999) or suggest new hands-on activities (e.g., Diem, 2001).

For all of its general support in the teaching community, questions about exactly how manipulatives influence learning have only begun to be addressed. For example, Chao, Stigler, and Woodward (2000) investigated whether structure or variability in physical manipulatives would have greater effectiveness in kindergartners' learning of number concepts. They found that each type of material had an impact on a different aspect of learning. However, their study did not contrast the effects of manipulation versus nonmanipulation of physical materials. Resnick (1998) suggested that it is manipulation, rather than physicality, as such, that may be the important aspect of instruction, and he described several types of "digital manipulatives"—a variety of programmable toys—that appear to engage children in the exploration and understanding of complex physical and mathematical phenomena. If it is the physicality of materials that increases instructional effectiveness, as the view against computer instruction suggests, then virtual materials would be less effective than physical materials.

Another perspective on the relative effectiveness of virtual versus physical materials is that the presentation medium makes no difference one way or another as long as the method of instruction is controlled (Clark, 1983, 1994). Clark's (1983, 1994) claim is based on a meta-analysis of instructional studies from the 1970s that compared comprehension of information conveyed through television, radio, books, and early computers. According to Clark, all studies that find learning effects of different media have failed to control instructional methods. However, Kozma (1991, 1994) challenged this claim because it did not account for recent cognitive research showing that information processing is highly dependent on the input modality, which varies among different media. In addition, Kozma argued that the unique learning opportunities provided by different media still remain to be fully explored. Whether Clark's (1983, 1994) claim that media effects will never be found is correct or not, his concern about the confounding of media with method remains relevant for much of the more recent research on the media effect on learning.

These two views suggest contrasting predictions about the effect of replacing physical materials with virtual presentations: (a) Presenting the materials on the computer will decrease learning because the physical interaction is an essential feature of the learning, or (b) Whether the materials are virtual or physical will make little difference as long as the method of instruction is preserved.

EMPIRICAL STUDIES

There are surprisingly few empirical studies that directly test the effect of using a computer or physical materials during instruction. Although many computer-based instructional systems from different domains have been evaluated—including some with impressive results (e.g., Koedinger, Anderson, Hadley, & Mark, 1997; White & Frederiksen, 1998)—such assessments typically, and intentionally, confound the computer versus noncomputer comparison with a host of other desired curricular changes, such as the content and sequencing of instructional components, the appearance of the materials, and the instructional goals themselves. The widely acclaimed intelligent tutor for algebra (Koedinger et al., 1997), which includes both the computer instruction and in-class activities, produced a standard deviation of improvement over traditional high school algebra instruction. Similarly, White and Frederiksen (1998) demonstrated an increase in elementary school students' knowledge of Newtonian physics nearly equivalent to 10 weeks of traditional instruction for high school students with a curriculum that combines hands-on and computer learning. However, neither of these two projects aimed to isolate the effect of computers in particular. Consequently, the role of the computer in the learning process—independent of other changes—remains obscure. It is this confounding of methods and media that Clark (1983, 1994) claimed was responsible for any media learning effects that have been found.

Similar issues arise in studies comparing the use of physical manipulatives to traditional instruction without physical materials. Although meta-analyses show moderate yet reliable effects of using manipulatives for periods of a semester or longer (Bredderman, 1983; Shymansky, Kyle, & Alport, 1983; Sowell, 1989; Stohr-Hunt, 1996), in these studies the comparison groups were taught using traditional instruction, whereas the manipulatives groups curricula differed in the activities as well as the addition of physical materials. To eliminate this confound, Sowell's (1989) meta-analysis separated pictorial manipulatives (e.g., flashcards) from physical manipulatives (e.g., wooden blocks), finding no reliable differences between pictorial manipulation and either the physical manipulatives or the traditional instruction conditions.

Research evaluating the use of concrete manipulatives covers both the mathematics and science domain. It is possible that the role of physical materials is influenced by the domain, but without research that compares only the instructional media while holding method constant, it is impossible to isolate the effect of physical manipulation from the method of instruction.

THEORETICAL PERSPECTIVES

Although active manipulation of physical materials plays an essential role in Piaget's theory of a concrete operational stage (Piaget, 1960), it is difficult to derive an unambiguous prediction from current major learning theories about the effect of physical manipulation on learning. Constructivist theory emphasizes the importance of children taking an active role in their own learning, but it does not specifically require physical manipulation. Cognitive theory focuses on the need for children to actively process information and practice the target skill. Clearly, when the target skill is perceptual-motor, then the practice of manipulating the actual items should benefit performance. However, for other types of tasks, active processing and practice do not require physical manipulation of materials. Thus, the general assumption that manipulation of physical materials enhances learning—particularly in young children—is not well-grounded in either constructivist or cognitive learning theory.

Why then is the emphasis on physical manipulation so prevalent? One possible reason is that instruction using physical materials tends to use methods that do result in better learning. In contrast to traditional instruction, which puts students in a passive role, instruction with physical materials typically incorporates active learning. The use of concrete materials builds on students' extensive experience in the physical world, whereas traditional instruction is often abstract and unrelated to children's knowledge base. In addition, physical materials can render relevant information as highly salient. All of these factors may contribute to learners' ability to understand and manipulate important concepts in comparisons of physical materials to more traditional instruction. Clements (1999) discussed the need to redefine "concrete" manipulatives to include digital manipulatives that preserve many of these features of physical manipulatives. In fact, many studies confound the comparison of physical materials with these other potentially influential factors (Baxter, 1995; Char, 1991; Ferguson & Hegerty, 1995; Fuller, May, & Butts, 1979; Gabel & Sherwood, 1980; Glasson, 1989; Moore, 1993; Richardson, Montello, & Hegarty, 1999; Riley, 1979; Thompson, 1992). Thus, the question remains about whether each of those methods would be equally effective if instantiated by virtual rather than physical materials. In this study, we contrasted only the presentation medium—virtual or physical—used in teaching children how to design unconfounded experiments, and we controlled the method of instruction such that the other factors that may influence learning were the same for both types of materials.

THIS EXPERIMENT

Our experiment compares the effectiveness of two instructional conditions that differ only in the medium of presentation. In one condition, elementary school children were taught using physical, manipulable materials, and in the other condition, they were taught using virtual, computer-based materials that were otherwise identical to the physical materials. All other important variables, such as the teacher, lesson plan, instructional script, time on task, number and type of examples, types of questions and probes from the teacher, and, perhaps most importantly, learners' selection and choice of how to set up different experiments, were the same in both conditions. Only the medium of presentation—virtual or physical—varied between conditions.

The topic of our instruction was how to design a simple unconfounded experiment. We taught fourth- and fifth-graders the control of variables strategy (CVS). CVS includes both the rationale and the procedure for setting up simple experimental contrasts such that two levels of a target variable are contrasted while all other variables are held constant. Figure 1 shows an example of an unconfounded experiment for the target variable of length of a spring in which all other variables are set to the same level. Figure 2 shows a confounded experiment with ramps, in which all variable levels differ for the comparison.

Although most elementary science curricula include some lessons on experimental design, many investigations have found that the majority of elementary

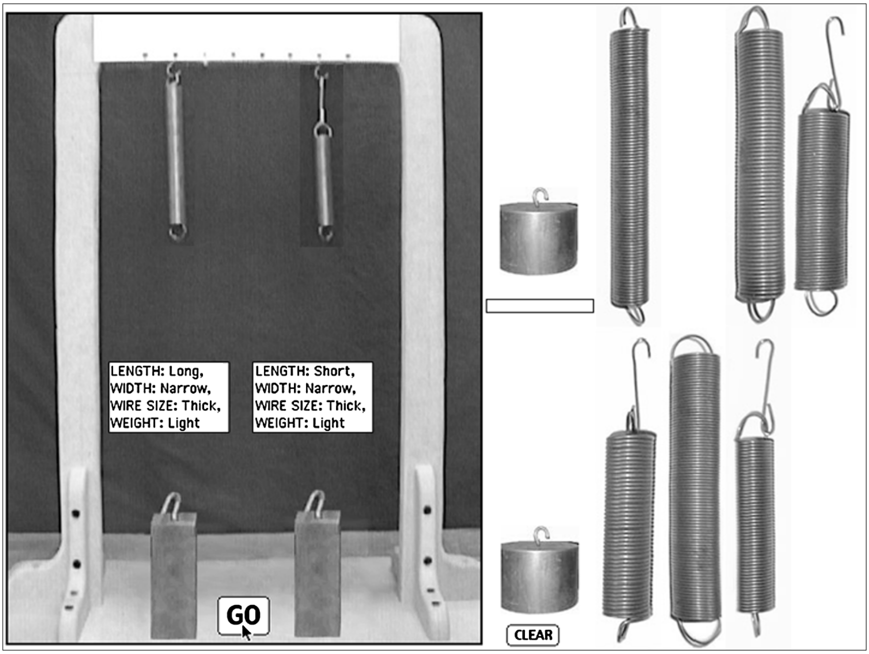

FIGURE 1 Screen shot of the spring selection screen of virtual materials showing an example of an unconfounded experiment of length of spring. Children click on a spring or weight from the right portion of the screen and then click on one side of the hanging rack to select an object for their experiment. After selecting a pair of springs and weights, children click on "go," and the display shows a dynamic video of the springs stretching with the weights.

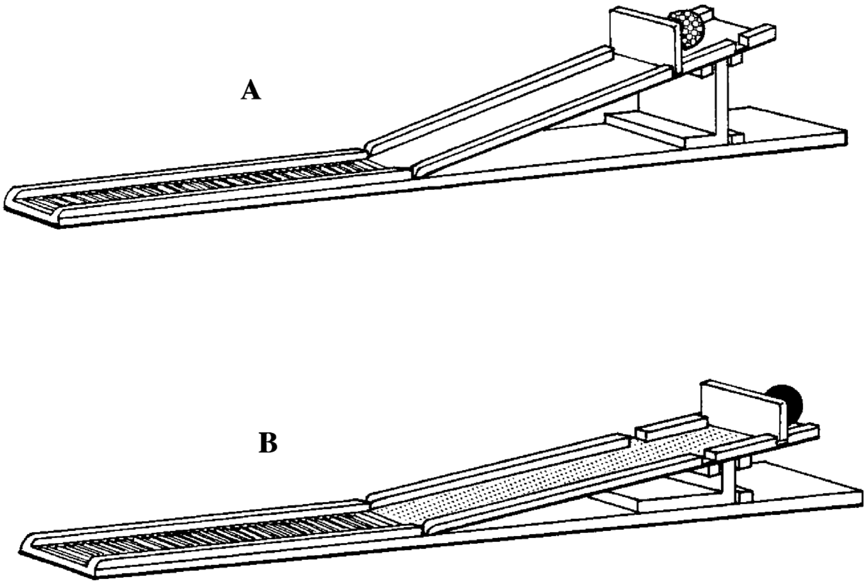

FIGURE 2 Diagram of the ramp materials used during the transfer phase. On each of the two slopes, children can vary the angle of the slope, the surface of the ramp, the length of the ramp, and the type of ball. The confounded experiment depicted here contrasts (A) the golf ball on the steep, smooth, short ramp with (B) the rubber ball on a shallow, rough, long ramp. From "All other things being equal: Acquisition and transfer of the control of variables strategy," by Chen and Klahr, 1999, *Child Development, 70,* p. 1103. Copyright 1999 by Society for Research Development. Reprinted with permission.

school students design confounded experiments (e.g., Kuhn, Garcia-Mila, Zohar, & Andersen, 1995; Schauble, 1996). Our instructional procedure was based on a training study by Chen and Klahr (1999) in which second-, third-, and fourth-grade children were taught CVS in the context of three domains: ramps, springs, and sinking objects. Prior to instruction, approximately 25% of children's experiments were unconfounded, but after a brief explicit instructional session using physical materials, this percentage rose to more than 75%. Moreover, on a far transfer test administered 7 months later, fourth-graders who had received instruction did significantly better than their untrained classmates. In addition to increasing children's CVS performance, the instruction also improved their knowledge about the effects of the variables in the domain that they explored without any explicit instruction about these effects. Toth, Klahr, and Chen (2000) extended the Chen and Klahr (1999) method for teaching CVS to a classroom context in which students' regular science teachers presented the instruction. Toth et al. (2000) found similar levels of learning despite an increase in the student/teacher ratio from 1:1 in Chen and

Klahr's (1999) laboratory study to approximately 20:1 in the Toth et al. (2000) classroom study.

These studies used physical materials: adjustable wooden ramps, springs of various dimensions, and various objects dropped into water-filled cylinders. Children handled, hefted, and manipulated all of these objects as they designed and interpreted their experiments during the training and assessment phases of the studies.

In this study, we compared the original Chen and Klahr (1999) CVS training procedure to one in which the materials were manipulated in a computer simulation. The virtual materials training condition replicated the Chen and Klahr procedures with one exception: Instead of physical materials, a computer simulation showed video depictions of the materials. During the session, children actively designed and set up their experiments by clicking on color photographic images of each of the elements to be compared from a layout similar to, and with the same amount of detail as, the physical materials. When the children ran their virtual experiments, they saw color video of the same physical objects used in the physical materials condition.

Although the move from physical to virtual materials carried with it the opportunity to change features of the Chen and Klahr (1999) instructional procedure, we avoided all such changes except for the medium of presentation. For example, we used a human instructor in both cases, rather than using a human in the physical materials condition and having the computer present a step-by-step tutorial in the virtual materials condition.

Because learning effects due to instructional medium might be specific to only certain types of measures, it is important to use multiple measures of student learning. Furthermore, the possibility that the two instructional media are equally effective puts us in the unenviable position of attempting to prove the null hypothesis. Therefore, in addition to measuring the proportion of trials in which children created unconfounded experiments, we also examined children's explicit knowledge about CVS as indicated by their explanations of their experimental designs and the conclusions they drew from the outcomes of their experiments. We also assessed the extent to which other, subtler, effects of the two training conditions could be found in children's confidence in their experimental conclusions and in the specific domain knowledge that they gained by running experiments.

We wanted to verify that children taught CVS using virtual materials could transfer this knowledge to experiments with physical materials. Thus, all children were assessed a week after training to determine whether they could transfer their CVS knowledge—acquired in the first domain—to the design of unconfounded experiments in another domain. For this transfer assessment, children in both training conditions worked with physical materials. Note that this assessment requires farther transfer for the children trained with virtual materials because they had to transfer their CVS knowledge not only to a different domain but also to a different medium, whereas children trained using physical materials needed only to transfer to a different domain in the same medium in which they were trained. Because our main purpose was to determine whether children using virtual materials learned as much as those using physical materials, testing whether students trained with physical materials could transfer their knowledge to virtual materials was deemed unnecessary, although such a test would have provided a symmetric design for the transfer phase of the study. As a further evaluation of learning, after children completed both sessions, we asked them about the similarities between the training and transfer sessions to determine if they could explicitly map their CVS knowledge across domains.

METHOD

Participants

Participants were 92 fourth- and fifth-graders (*M* age = 10.6, 51 girls and 41 boys) from two parochial elementary schools in an urban area of southwestern Pennsylvania. Participants were recruited with notices sent to parents. Children were randomly assigned to a training session that used either physical or virtual materials.

Design

We used a 2 (condition: physical vs. virtual materials) × 3 (phase: pretest and training, posttest, and transfer) factorial design with phase as a within-participant factor. During each of three phases—pretest and training, posttest, and transfer—children designed four simple paired-comparison experiments.

The two conditions differed only during the first two phases in the medium of materials. In the physical materials condition, children used real springs and weights to set up and conduct their experiments. In the virtual materials condition, these same springs and weights were depicted on a computer screen, and children clicked using a mouse to design their experiments and then observed dynamic digital video depictions of the springs stretching. In the third phase (transfer), children in both conditions worked with physical materials in the new domain of ramps.

Materials

Springs. Children first worked with the spring domain. There were both physical and virtual versions of this task. (See Figure 1 for a picture of the spring selection screen from the virtual condition.) In both conditions, the children selected from eight different springs that varied on three dimensions, each having two levels: length (long or short), width (wide or narrow), and wire size (thin or thick). The fourth variable was the mass of the weight that could be attached to each spring (heavy or light).[1] The goal of each experiment was to determine whether a specified variable mattered (e.g., "make a comparison that shows whether the length of a spring makes a difference in how far it stretches"). To select springs for a comparison, children in the physical materials condition grabbed a spring or a weight and then placed them onto a hanging frame. Children in the virtual materials condition used a mouse click to select a spring or weight and then clicked one side of the frame to place it. Because the differences in wire size were not easily discernable in either condition, written labels showing the value of all the dimensions were provided for the springs and weights. After designing the experiment (i.e., selecting a pair of springs and weight combinations), children in the physical materials condition put the weights on the selected springs to see the two springs stretch. Children using virtual materials clicked a "go" button and saw simultaneous color video segments of their chosen springs stretching as the selected weights were attached. The video segments were of the same springs and weights used in the physical materials condition.

Ramps. For the transfer task, all children worked with physical ramps. Materials consisted of two wooden ramps approximately 24 in. long that could be adjusted in several ways. (See Figure 2 for a depiction of the physical materials.) The steepness of the slope could be changed to a high or low position. The length the ball rolled down the ramp could be long or short. The surface of the ramp could be either smooth or rough. In addition, children chose whether a golf ball or a rubber ball would roll down the ramp. All children physically manipulated the ramps to set up their experiments and released starting gates to see which ball rolled farther on a stepped surface after leaving the downhill ramp.

Confidence scale. For some of the questions, children were asked to indicate how sure they were of their answers. Children were asked whether they were *totally sure, pretty sure, kind of sure*, or *not so sure*. The scale was presented on paper to the children and was always available for reference as they responded to the confidence questions. Responses were coded on a 4-point Likert-type scale ranging from 4 (*totally sure*) to 1 (*not so sure*).

Procedure

All children participated in three phases: (a) pretest and training, (b) posttest, and (c) transfer. The first two phases were completed during a 45-min session, and the

[1]Mass was arbitrarily correlated with height and shape: the heavy weights were short cylinders and the light weights were tall rectangular cubes.

transfer phase was completed during a 20-min follow-up session about a week later. Children were interviewed individually in a quiet room, and all of the children's activities, including their experimental designs and their verbal explanations, were videotaped for later coding and analysis.

Pretest and training. The pretest assessed children's initial knowledge about the spring domain and their initial knowledge about the control of variables strategy. An instructor[2] presented the spring materials (either the physical or virtual materials, depending on condition) and introduced the four variables (length, width, wire size, and weight). Children verified their understanding by identifying each of the springs by each variable level (e.g., "go through all the springs and tell me if it has thick or thin wires"). In addition, children were introduced to the confidence scale and given a practice question to ensure they understood how to respond about their confidence. Once the variables and the confidence scale were introduced, children's knowledge about the spring domain was assessed by asking them to indicate which level of each variable would make the spring stretch farther (e.g., "which stretches farther, a long spring or a short spring?").

Next, children designed and ran experiments to show whether a particular target variable made a difference in how far springs stretch. Children designed four experiments during pretest: two that focused on the effect of spring length and two that focused on the effect of spring width. After selecting the variable levels, but before seeing the results (i.e., before actually running the experiment), children responded to the instructor's question about their design: "Why did you choose those two springs and weights?" After seeing the springs stretch, children indicated their confidence in the effect of the target variable by using the 4-point scale (e.g., "How sure are you from this comparison that the length of a spring makes a difference?"), and explained their response (e.g., "What makes you pretty sure?") to the instructor.

Once these preliminary assessments were completed, the instructor trained the child on the control of variables strategy with a brief session (approximately 10 min) of didactic instruction. Children were told about the goal of designing unconfounded experiments and were presented with both positive and negative examples. First, they observed a confounded experiment in which all the variable levels differed. They were asked whether that experiment was a "smart way" to determine whether the length of a spring makes a difference and asked to explain their answer. Then the instructor modeled the appropriate response, explaining that you could not make an inference about the effect of spring length from the

[2]We use instructor rather than the more conventional experimenter in this description to avoid confusion between the child experimenters and the adult who is conducting the study, presenting the materials, asking questions, and, during training, providing explanations and examples.

confounded experiment because the two springs differed in several ways in addition to the length. Next, children were shown an unconfounded experiment about the effect of spring length. They were told that this design was a better experiment and asked to explain why. Then the instructor explained that, because length was the only thing different between the two springs, if one spring stretches farther, it must be caused by the difference in length. The instructor went through two additional examples using width as the target variable, one example with a single confound and one that was unconfounded. As a final step, the instructor summarized that, to figure out if something makes a difference, comparisons need to differ in only one way. Children did not see the results of the experiments during training.

Posttest. The immediate effectiveness of training was evaluated in the posttest phase. Children designed four more experiments: two for wire size and two for width.[3] As in the pretest, for each experiment, children explained their design, observed the outcome, and indicated their confidence about whether the target variable made a difference. After the four experiments were completed, children were asked the same questions as in the pretest about the effects of the variables to assess changes in domain knowledge.

Transfer. The transfer phase occurred about a week after the first session. All the children worked with physical ramps. After the instructor described the domain and the four variables, children predicted the effects of the variables to assess their initial knowledge of ramps. Children then designed four experiments: two to determine the effect of steepness on how far the balls rolled and two to determine the effect of the length of the run. As before, children explained why they picked the design, observed the outcome, and then indicated their confidence about whether the target variable made a difference. After all four experiments were completed, children's knowledge about the effects of the variables in the ramp domain was reassessed. Finally, children were asked if they noticed any similarities between the spring and ramp domains with the following questions: (a) "Did the problem with the ramps this week remind you of the problem last week?" (b) "Were today's problem and last week's problem pretty much alike or pretty much different?" (c) "Did the problem you worked on last week help you figure out the problem this week?" By assessing whether children mentioned controlling extraneous variables as a similarity between two sessions, we could determine if they explicitly mapped their knowledge from the training domain (springs) to the transfer domain (ramps).

[3]Although the spring domain has four binary variables, because the effect of the mass of the weight used on the spring is not visually discernible, it was not used as a target variable. Instead children designed experiments to test the effect of spring width both before and after training.

RESULTS

Our analysis addresses four possible effects of physical versus virtual instructional materials: (a) learning of CVS, as reflected in both experimental designs and justifications for them; (b) confidence in conclusions from the experiments; (c) changes in knowledge of the domain variable effects; and (d) explicit awareness of CVS as a similarity between the training and transfer contexts.

Learning and Transfer of CVS

Our primary question was whether there is a difference in how much children learn about designing good experiments from instruction using physical or virtual materials. Our first analysis focused on the number of good designs children made in each phase. The mean proportion of unconfounded experiments in each phase is shown in Figure 3. A 2 (training condition) × 3 (phase) multiple analyses of variance (MANOVA), with phase as a within-participant factor, showed a main effect for phase, $F(2, 89) = 65.93$, $p < .001$, with no main effect of condition, $F(1, 90) = .01$, *ns*, and no interaction between phase and condition, $F(2, 89) = 0.73$, *ns*.[4] The proportion of unconfounded experiments increased significantly from pretest, $M = .20$, $SD = .26$, to posttest, $M = .65$, $SD = .41$, *95% CI of diff*: (.35, .55)[5] and from pretest to transfer, $M = .64$, $SD = .41$, *95% CI of diff*: (.34, .54), and there was no significant difference between posttest and transfer, *95% CI of diff*: (–.06, .08). The two types of training were equally effective in teaching children how to design unconfounded experiments. Moreover, children trained with virtual materials designed as many unconfounded experiments with the physical materials during the transfer phase as children who used physical materials for all phases.

Although the most direct measure of children's CVS knowledge—whether or not their design was unconfounded—showed no effect of training condition, it is possible that there was a difference between conditions in children's acquisition of explicit knowledge of CVS. Such knowledge could be revealed by their explanations of their designs and confidence in the conclusions of the experiments. To measure explicit CVS knowledge, we categorized children's responses to the questions about their design as to whether the children mentioned controlling all variables besides the target variable (e.g., "You just need to make the length of the spring different, but use springs with the same wire size and both wide, and the same kind of weights"). It is

[4]The results are aggregated over grade and sex of participants because a MANOVA including these factors found no significant effects of them.

[5]Sheffé post hoc critical values are used for all 95% confidence intervals of difference. For 2 × 3 MANOVAs, $F_{critical} = 2.02 * F_{.05}(2, 89)$.

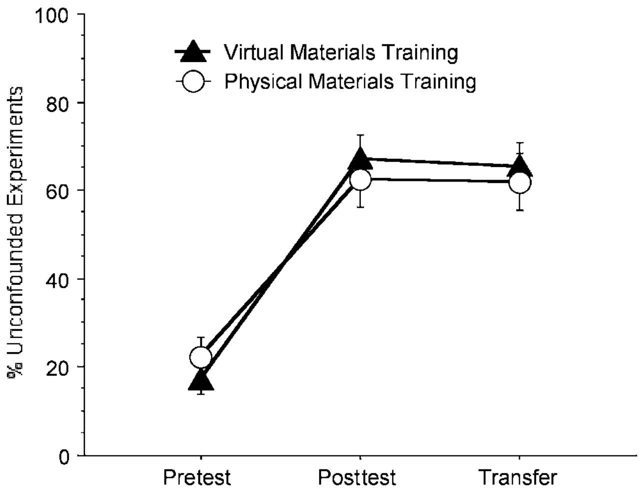

FIGURE 3 Mean proportion of unconfounded experiments for each phase separated by training condition with standard error bars.

important to emphasize that children did not simply give a verbatim repetition of the terminology they heard during instruction. Children had to use different contrastive dimensions during the posttest phase than those that were used during training (e.g., they had to talk about wire size rather than spring length). During the transfer phase, children had to go well beyond the training terminology to make the correct mapping of the CVS procedure to the new variable names in the ramp domain. The proportion of trials for which children produced an unconfounded design and mentioned the control of all the other variables is presented in Figure 4.[6] A 2 (training condition) × 3 (phase) MANOVA, with phase as a within-participant factor, showed a main effect of phase, $F(2, 88) = 62.44$, $p < .001$, no main effect of condition, $F(1, 89) = 0.14$, *ns*, and no interaction, $F(2, 88) = 0.71$, *ns*. Children revealed explicit knowledge of CVS more often during posttest, $M = .52$, $SD = .42$, and transfer, $M = .46$, $SD = .39$, than during pretest, $M = .07$, $SD = .17$, *95% CI of diff*: (.35, .55), and *95% CI of diff*: (.29, .48), respectively. The difference between posttest and transfer was not significant, *95% CI of diff*: (–.01, .14). The main effect of phase suggests that children were more likely to express explicit knowledge of CVS after training than before, and the lack of an interaction suggests that this effect did not vary by condition. On approximately half of the experiments in the posttest and transfer phases, children explicitly mentioned the need to control all but the target variable when describing their

[6]One child is not included in the analyses of explicit CVS knowledge because of missing data for the transfer phase.

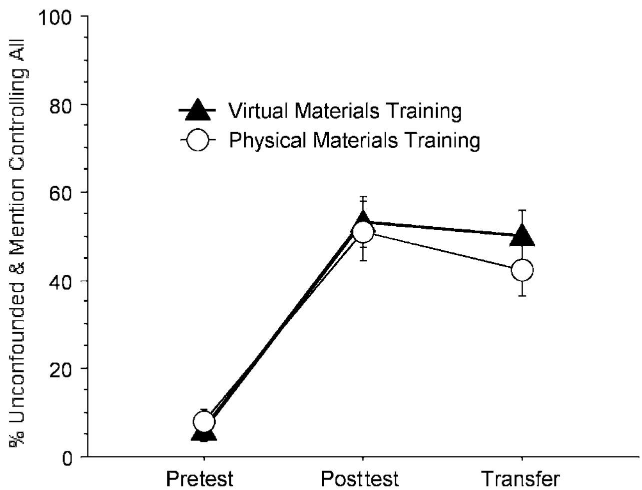

FIGURE 4 Mean proportion of trials in which children designed an unconfounded experiment *and* mentioned controlling all other variables for each phase separated by training condition with standard error bars.

unconfounded designs. These results provide convincing evidence that children in both conditions learned the control of variables strategy.

The analyses up to this point have been based on averages over all children in each condition. To assess learning and transfer by individual children, we defined a *good experimenter* as a child who designed at least seven out of eight unconfounded experiments during the four posttest and four transfer trials (i.e., after training). Fifty percent (23/46) of the children trained with physical materials and 52% (24/46) of the children trained with virtual materials were classified as good experimenters. Children in both conditions had the same likelihood of becoming good experimenters, $\chi^2(1, N = 92) = .04$, *ns*. The lack of difference between conditions remains even if we modify the definition of good experimenter to require mentioning controlling all variables in addition to designing good experiments (Physical: 28% or 13/46, Virtual: 31% or 14/45), $\chi^2(1, N = 91) = .09$, *ns* (see footnote 6). About 50% of the children designed seven or eight unconfounded experiments after training, with more than a quarter of all children explicitly referring to CVS on at least seven of the eight trials after training, regardless of training condition.

Although the number of children classified as good experimenters was the same in both conditions, it is possible that training condition could affect the proportion of *nonlearners*—children who designed less than two unconfounded experiments during the four posttest and four transfer trials. There were no significant differences in the proportion for whom training was ineffective, $\chi^2(1, N = 91) = .60$, *ns*, with 24% (11/46) of the children trained with virtual materials and 17% (8/46) of the children trained with physical materials classified as nonlearners.

The criteria used to categorize both good experimenters and nonlearners are somewhat arbitrary, but, as Figure 5 shows, the distribution of children by number of unconfounded experiments designed after training is similar for the two conditions regardless of the particular criteria chosen.

To examine learning and transfer over all three phases by individual children, we set a criterion of excelling in each phase. Children who designed at least three out of four unconfounded experiments in a phase were considered to have excelled in this phase. We then compared the proportion of children who excelled in each phase under the two training conditions (see Table 1). Fisher Exact tests (Siegel & Castellan, 1988) revealed no effects of condition for any of the phases. These analyses show the proportion of children excelling in each phase was not influenced by type of instructional materials.

However, it is possible that the equivalent proportions in the posttest and transfer phases from the two training conditions are made up of different types of distributions: (a) children who maintained the same performance across the phases, or (b) a combination of children whose performance improved and children whose performance declined. Because in the posttest phase children used different materials according to condition, whereas in the transfer phase all children used physical materials, separate McNemar chi-square tests (Agresti, 1996) for each training

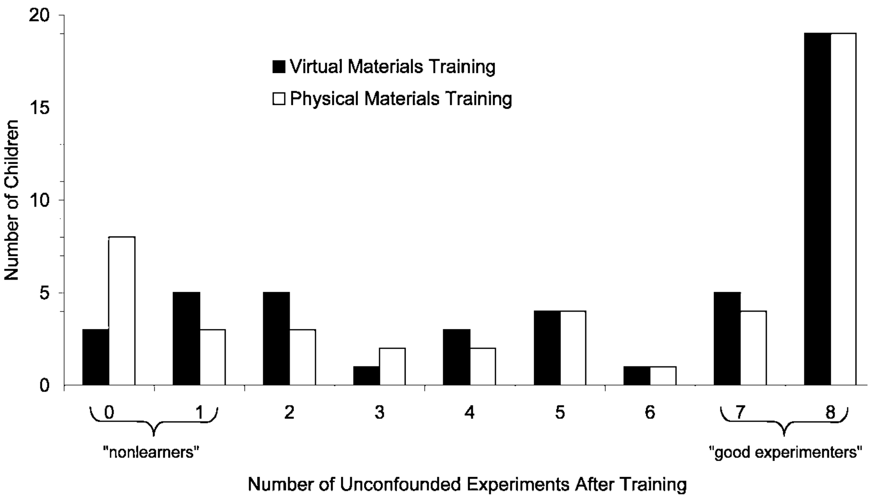

FIGURE 5 Distribution of children separated by condition according to number of unconfounded experiments designed after training (the eight comparisons in the posttest and transfer phases) to show that, although the criteria used to compare the number of good experimenters and nonlearners are arbitrary, the conditions produced similar distributions.

TABLE 1
Proportion and Number of Children With Three out of Four Unconfounded Experiments in Each Phase and Training Condition

		Phase					
Training Type	*n*	*Pretest*	*%*	*Posttest*	*%*	*Transfer*[a]	*%*
Criterion: Unconfounded experiments on at least three out of four trials							
Physical materials training	46	11	5_a	57	26_b	59	27_b
Virtual materials training	46	4	2_a	61	28_b	61	28_b
Criterion: Unconfounded experiment and mention CVS on at least three out of four trials							
Physical materials training	46	2	1_c	50	23_d	35	16_d
Virtual materials training	45	2	1_c	52	24_d	47	21_d

Note. None of the differences between training conditions are significant using Fisher Exact test $p < .05$. Mean differences in the same row that do not share subscripts differ at $p < .02$ using McNemar chi-square tests. CVS = control of variables strategy.

[a]Both the physical and virtual materials training conditions worked with physical materials during the transfer phase.

condition compared the proportion of children who changed from posttest to transfer in whether they excelled. These analyses found no significant differences in the number of children changing excellence classification for both types of training materials, Physical: $\chi^2(1, N = 46) = .20$, *ns*; Virtual: $\chi^2(1, N = 46) = .00$, *ns*. Of the children who used physical materials, 3 began excelling in the transfer phase, 2 excelled only in the transfer phase, and the other children maintained their posttest excellence classification into the transfer phase. For the virtual materials condition, 4 children improved and 4 children declined in their classification. The conditions did not differ in changes between posttest and transfer on CVS performance, and the low numbers of children changing performance is further evidence that children maintained the CVS knowledge they learned during training.

Somewhat different results were obtained if the CVS performance classification requires that children design an unconfounded experiment and mention controlling extraneous variables. Children trained with virtual materials did not differ in changes between posttest and transfer, $\chi^2(1, N = 45) = .82$, *ns*, with 4 children improving and 7 children failing to transfer their good posttest performance. In contrast, because no children trained with physical materials began excelling in the transfer phase, but 7 children's performance declined, there was a significant difference between the number of children in the physical materials condition who declined compared to those improving between phases in explicitly mentioning CVS, $\chi^2(1, N = 46) = 7.00$, $p < .01$. However, a similar proportion of children in both conditions maintained their good posttest performance into the transfer phase (Physical: 16/46; Virtual: 17/45) and there were no significant condition differences in excellence classification for any of the phases. The difference between conditions is that, whereas a few children trained with virtual materials began mentioning CVS three or more times during the transfer phase, no children trained with physical materials made this transition. Overall, these results suggest that the two training conditions were equally effective in producing substantial and stable improvements in the performance of individual children from posttest and transfer.

In summary, these analyses—at both aggregate and individual levels—failed to reveal any differences of type of training on children's ability to either correctly execute the CVS procedures or explicitly describe their knowledge about CVS.

Confidence About Conclusions

Although children appear to learn and transfer CVS equally well in both training conditions, it is possible that the training medium influences their confidence in the conclusions they can draw from their experiments. Recall that, after seeing the results of an experiment, children were asked to indicate their confidence—based on the experiment—that the target variable made a difference (e.g., "How sure are you from this comparison that length of a spring makes a difference?"). Children indicated their confidence levels on a 4-point scale ranging from 4 (*totally sure*) to 1 (*not so sure*). Figure 6 depicts these responses averaged over the four trials in each phase.[7] A 2 (training condition) × 3 (phase) MANOVA on children's mean confidence, with phase as a within-participant factor, did not find a main effect of training condition, $F(1, 90) = 1.21$, $p = .27$. There was a main effect of phase, $F(2, 89) = 12.68$, $p < .001$, such that differences were significant between pretest ($M = 2.8$, $SD = .60$) and posttest ($M = 3.2$, $SD = .63$), *95% CI of diff*: (.17, .51), and between pretest and transfer ($M = 3.1$, $SD = .63$), *95% CI of diff*: (.02, .42), but not between posttest and transfer, *95% CI of diff*: (–.03, .26). The interaction between phase and condition approached significance, $F(2, 89) = 2.88$, $p = .06$. The largest difference in confidence between conditions was during the pretest (physical materials: $M = 3.0$, $SD = .61$; virtual materials: $M = 2.7$, $SD = .56$), $t(90) = 2.35$, $p = .02$, (*95% CI of diff*: .04, .53). No significant differences between conditions were found for posttest or transfer. Overall, these results reveal significant increases in children's confidence levels—increases that did not differ between conditions after training.

The increase in children's confidence in the conclusions of their experiments from pretest to posttest and transfer is most likely due to the increase in their use of

[7]To control for individual variation in use of the 4-point confidence scale, all analyses were also done with a standardized confidence measure. The nontransformed results are reported here because the analyses with standardized confidence replicated the effects with the original 4-point scale.

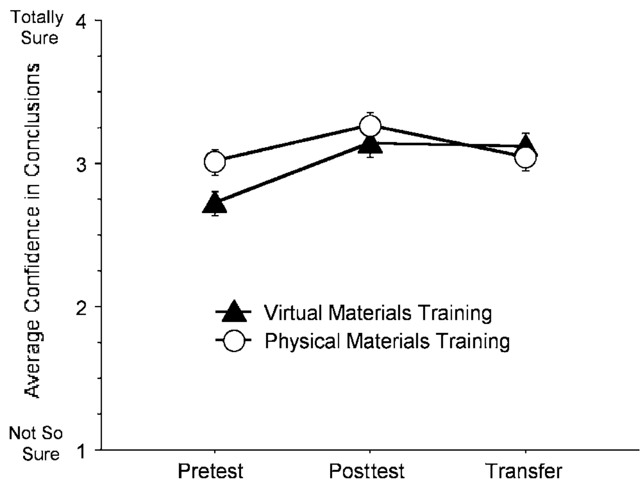

FIGURE 6 Mean of children's confidence in experimental conclusions for each of the phases separated by training condition with standard error bars.

unconfounded experiments. To test this hypothesis, we compared children's average confidence after running unconfounded experiments to their average confidence after running confounded experiments—over all three phases—using a paired *t* test. Eleven children were excluded from this analysis because they created either all unconfounded ($n = 3$) or all confounded ($n = 8$) experiments. The difference in children's confidence between unconfounded, $M = 3.2$, $SD = .56$, and confounded experiments, $M = 2.8$, $SD = .63$, was significant, $t(80) = 4.86$, $p < .001$, *95% CI of diff*: (.24, .56). This difference was not affected by training condition, $t(79) = 0.38$, *ns, 95% CI of diff*: (–.39, .27). Across all three phases, children were more confident in the conclusions they based on their unconfounded experiments than on their confounded experiments, regardless of training condition.

Because children conducted two experiments about a particular target variable, it is likely their confidence was influenced by whether they were conducting their first or second experiment on a particular target variable. A paired *t* test comparing children's average confidence of their first experiment to their second experiment on the same target variable revealed a small but significant increase in confidence (First: $M = 3.0$, $SD = .52$; Second: $M = 3.1$, $SD = .55$), $t(91) = 2.46$, $p < .02$, *95% CI of diff*: (.03, .19). This increase in confidence between the first and second experiments on a target variable did not vary between training conditions, $t(90) = 0.95$, *ns*. Children became more confident in the conclusions of experiments after doing two experiments on the same target variable in both training conditions.

In summary, children's confidence in experiments increased from pretest to posttest and transfer, after designing unconfounded experiments, and after completing a second experiment on a target variable. None of these changes in confidence was influenced by the training condition.

Domain Knowledge

Although we did not find an effect of training medium on children's learning of CVS or on their confidence in their conclusions, it is possible that training medium has subtler effects on the amount of domain knowledge that children indirectly acquire as they learn from their experimental outcomes. We examined children's domain knowledge before and after they completed their experiments in the spring domain and in the transfer domain of ramps. Figure 7 shows the proportion of children's correct responses about the effects of all four binary variables (e.g., "which stretches farther, a wide spring or a narrow spring?"). Separate 2 (condition) × 2 (time of assessment) MANOVAs for each of the domains, with time as a within-participant factor, showed a significant main effect on domain knowledge for time of assessment, Springs: $F(1, 90) = 13.99$, $p < .001$;

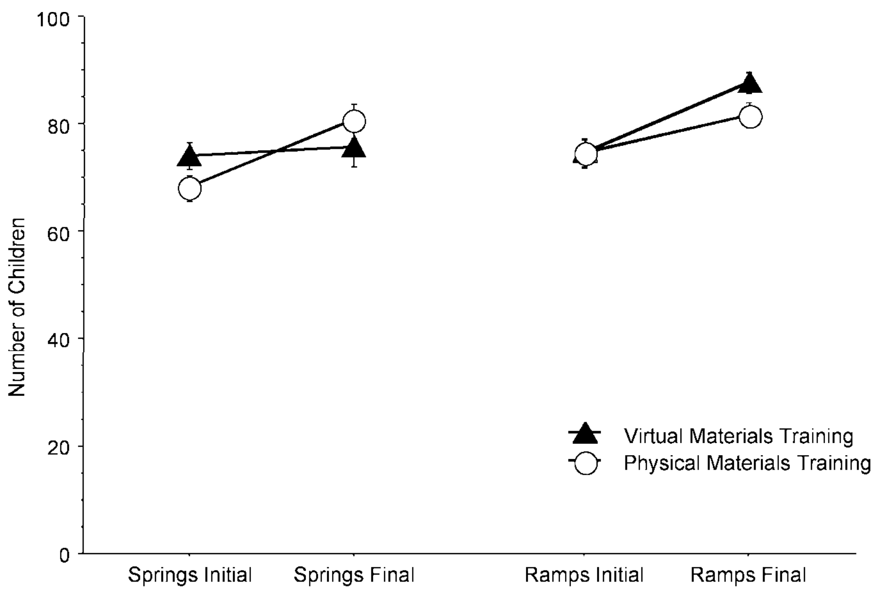

FIGURE 7 Mean proportion of correct descriptions of variable effects for both the training (springs) and transfer (ramps) domains separated by training condition with standard error bars.

Ramps: $F(1, 90) = 17.44$, $p < .001$. Children knew more about the effects of the variables after designing and running experiments (Springs: $M = .80$, $SD = .22$; Ramps: $M = .85$, $SD = .15$) than before (Springs: $M = .71$, $SD = .17$, *95% CI of diff*: .05, .13; Ramps: $M = .75$, $SD = .18$, *95% CI of diff*: .06, .14). There was not a main effect of condition for the spring domain, $F(1, 90) < 0.01$, *ns*, or for the ramp domain, $F(1, 90) = 1.46$, $p = .23$. There was also no interaction between condition and time of assessment for the ramp domain, $F(1, 90) = 1.54$, $p = .22$. However, for the spring domain there was a significant interaction between condition and time of assessment, $F(1, 90) = 6.61$, $p = .01$, suggesting that children in each condition might learn domain knowledge at a different rate. But, as Figure 7 shows, initial domain knowledge was slightly higher for children in the virtual materials training condition than children in the physical materials condition. This initial difference failed to reach significance, $t(90) = 1.69$, $p = .09$, *95% CI of diff*: (–.13, .01), and, by the end of the training, there were no significant differences in domain knowledge, $t(90) = 1.31$, $p = .19$, *95% CI of diff*: (–.03, .15). It is unclear how to interpret this interaction because of the lack of pair-wise differences.

In addition to examining children's aggregate performance, we explored changes in children's knowledge separately for each of the domain variables for two reasons: (a) to determine if children gained knowledge about all the variables or just a subset of the variables, and (b) to evaluate whether the interaction with condition was limited to a subset of the variables. Separate McNemar chi-square tests (Agresti, 1996) compared the proportion of children changing their responses about the effects of each variable from initial to final assessments. Children's changes in domain knowledge were significant for four of the eight domain variables: spring width, $\chi^2(1, N = 92) = 38.4$, $p < .001$, spring weight, $\chi^2(1, N = 92) = 12.8$, $p < .001$, ramp run length, $\chi^2(1, N = 92) = 11.6$, $p < .001$, and ramp ball type, $\chi^2(1, N = 92) = 5.5$, $p < .02$. More children changed from an incorrect to a correct response about the variable effect than from correct to incorrect for three of the domain variables: spring width, 60% compared to 7%, ramp run length, 22% compared to 4%, and ramp ball type, 30% compared to 14%. However, for weight, more children changed from correct to incorrect (20%) than from incorrect to correct (2%). Children's accuracy in describing the effect of spring length, spring wire size, ramp steepness, and ramp surface was quite high initially (80, 86, 94, and 96%, respectively) and did not change significantly between initial and final assessments. These analyses on separate variables suggest that the main effect of time for the spring domain was primarily due to changes in children's responses about width. Significant changes in responses about weight on the spring reduced the magnitude of the effect. For the ramp domain, the aggregate main effect of assessment time was due to changes in children's responses about the run length and ball type. It may be surprising that the only target variable on which children showed significant change in domain knowledge was spring width.

However, this was probably due to ceiling effects because the only spring variable that children had incorrect prior knowledge about was width: only 20% of the children were initially correct about the effect. Ball type was the only other variable that prior to using the materials children were below chance in predicting the effect: only 34% were correct initially.

To examine how the change in children's knowledge of the variable effects varied by condition, the McNemar chi-square tests (Agresti, 1996) were separated by training condition. Both conditions had significant changes from incorrect to correct for spring width, but children trained using physical materials had a greater change, $\chi^2(1, N = 46) = 24.03, p < .001$, with 70% becoming correct compared to 7% becoming incorrect, in contrast to children trained using virtual materials, with 48% improving compared to 7% deteriorating, $\chi^2(1, N = 46) = 14.44, p < .001$. The only other McNemar test that showed differential effects by training condition was ball type. Children trained with virtual materials were more likely to change from incorrect to correct (33%) than from correct to incorrect (7%), $\chi^2(1, N = 92) = 8.00, p < .005$, whereas children trained with physical materials were just as likely to change from incorrect to correct (28%) as from correct to incorrect (22%). Because all children used physical ramps and never tested ball type as the target variable, it is unclear why this variable would show differential change for the different training conditions.

In summary, more children had correct domain knowledge after experimentation than before, but a majority of this change is accounted for by four of the eight variables with responses to one variable (spring weight) becoming less accurate while the responses to the other three variables (spring width, run length, and ball type) became more accurate. Two of these variables showed differential effects for different conditions, but for one of them (spring width) children in the physical materials training condition learned more than virtual materials while the other variable (ball type) showed the opposite effect. Overall, these results suggest that children learned indirectly about the effects of some of the variables and there were no strong differences in that learning between training conditions.

Similarities Awareness Between Training and Transfer

At the end of the transfer phase during the second session, children were asked a few questions that compared their earlier experience with the springs to their current problems with the ramps.[8] Children's responses were analyzed to determine if any of their responses referred to controlling all extraneous variables as a similarity between the domains. Similarity awareness was shown by 38% (17/45) of the

[8]Two children are not included in the analysis on explicit similarity due to missing data.

children trained with physical materials and 33% (15/45) of those trained with virtual materials, $\chi^2(1, N = 90) = .19$, *ns*. Children in both conditions were equally likely to extract the deep structure of CVS and explicitly mention the similarity between training and transfer domains, despite children trained with virtual materials changing mediums in addition to domains.

DISCUSSION

The results of this study suggest that fourth- and fifth-graders learned to design unconfounded experiments equally well when taught with either virtual or physical materials. Children in both training conditions made large gains in their experimental design performance as well as in their ability to justify that performance verbally. Moreover, in the transfer phase with physical materials in a new domain, children who were trained with virtual materials, and thus had to transfer not only from one domain to another but also from one instructional medium to another, performed just as well as children who had to only transfer to a new domain in a familiar medium. In addition, there was no difference between conditions for most of the other measures, such as children's confidence in conclusions or their recognition of similarity between the two domains. The only significant interaction with training condition was due to differential gain in knowledge about the effect of spring width, but there were no significant pair-wise differences in domain knowledge between conditions. These results suggest that simply replacing the physical materials with virtual materials does not affect the amount of learning or transfer when other aspects of the instruction are preserved.

The equivalent effectiveness of physical and virtual materials challenges the general assumption that physical manipulation improves learning. The results of this study suggest that, because the virtual instruction condition captured the important features of the instruction, the physical interaction with the materials was unnecessary in this learning context.

Will instructional media ever influence learning? Clark (1983, 1994) insisted that the answer to this question is no, but we feel there are two issues that need to be addressed before it can be answered: (a) Are there differential effects of media for different types of tasks, and (b) Would the computer have the same learning effect without a human instructor?

The instructional objective for this study was experimental design (in our terms, the CVS). This domain-general skill is a central objective for elementary school science (National Research Council, 1996). The results of this study suggest that the essential procedural and conceptual knowledge for CVS does not require physical interaction with materials, and thus the substitution of virtual for physical materials had no negative effects on learning and transfer. Clearly, if instructional goals include mastery of a specific perceptual-motor skill, such as making precise incisions during dissections or careful mixing and pouring of liquids, then it is likely that there would be an effect of medium because the different movements required for manipulating a keyboard and a mouse rather than an apparatus would interfere with the acquisition of particular perceptual-motor skills needed in the target domain.

Physical materials are often used in mathematics instruction to support the acquisition of abstract concepts and manipulation instead of the symbolic representations that are difficult for novices to use. Within the science domain, in addition to experimentation, physical materials are often used to create models, such as a model of the solar system made of styrofoam. It remains to be seen what other domains, even those where physical materials typically have a large effect, will show an equivalence between the instructional efficacy of physical and virtual materials when methods are carefully controlled.

In this study, we used a human instructor for both training conditions. It is unknown whether similar learning would occur if students were trained directly from the computer. For example, a step-by-step tutorial of CVS, following the same script used by the human instructor, could be presented to the children within the virtual materials. Implementing this tutorial requires making several design decisions about the specific features of the instruction, including whether to include prompts as audio or written text, and how to have children respond to the open-ended prompts. If media influences learning, each of these design choices could influence children's learning outcomes. Manipulating whether a computer or a human provides instruction is a further test of Clark's (1983, 1994) claim because it evaluates the general assumption that children learn more from working with a teacher than a computer. Without data addressing this question, it would be premature to generalize the results of this study to the claim that media will never have an influence on learning.

Even if children learn equally from different media when instructional method is controlled, clearly some methods are unique to a particular medium (Kozma, 1991, 1994). For example, some research has found improved learning when a graphical representation is automatically linked to a symbolic representation (Moore, 1993; Thompson, 1992), but this feature is difficult to implement using media other than computer presentation. It is important that research explore all kinds of instructional methods to develop the most effective instruction, whether this is limited to one medium or can be implemented with several different media.

Should teachers abandon the hands-on science materials they have been using with their classes over the years? Not necessarily, but, as this study indicates, neither should they be wedded to them on the basis of folk psychology or vague theory. Much remains to be learned about the impact of physical and virtual materials in different domains and instructional contexts, and that investigation can be carried out by using the same approach to science that we taught to the children in this study.

ACKNOWLEDGMENTS

This work was supported in part by grants from NICHHD (HD25211) and the James S. McDonnell Foundation (96-37) to David Klahr.

We would like to thank the children, parents, principals, and teachers at Immaculate Conception School and Sacred Heart Elementary School for their participation and cooperation. We also thank Jennifer Schnakenberg for her assistance in collecting data and Jolene Watson for assistance in coding, both of whom had insightful suggestions during several stages of the project. We appreciate the constructive comments on earlier drafts from Ryan Baker, Sharon Carver, Norma Chang, Kenneth Koedinger, Amy Masnick, Brad Morris, Anne Siegel, Robert Siegler, Eva Erdosne Toth, and Corinne Zimmerman.

Portions of these results were presented at the meeting of the American Educational Research Association, April 2001, Seattle, WA.

REFERENCES

Agresti, A. (1996). *An introduction to categorical data analysis.* New York: Wiley.

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

Armstrong, A., & Casement, C. (1998). *The child and the machine: Why computers may put our children's education at risk.* Toronto, Ontario: Key Porter Books.

Baxter, G. P. (1995). Using computer simulations to assess hands-on science learning. *Journal of Science Education and Technology, 4,* 21–27.

Berk, E. G. (1999). Hands-on science: Using manipulatives in the classroom. *Principal, 78,* 52–55, 57.

Bredderman, T. (1983). Effects of activity-based elementary science on student outcomes: A quantitative synthesis. *Review of Educational Research, 53,* 499–518.

Chao, S., Stigler, J. W., & Woodward, J. A. (2000). The effects of physical materials on kindergartners' learning of number concepts. *Cognition and Instruction, 18,* 285–316.

Char, C. (1991). *Computer graphic feltboards: New software approaches to children's mathematical exploration* (Report No. 91-1). Newton, MA: Education Development Center, Inc. (ERIC Document Reproduction Service No. ED347080)

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

Clark, R. E. (1983). Reconsidering research on learning from media. *Review of Educational Research, 53,* 445–459.

Clark, R. E. (1994). Media will never influence learning. *Educational Technology Research and Development, 42,* 21–29.

Clements, D. (1999). "Concrete" manipulatives, concrete ideas. *Contemporary Issues in Early Childhood.* [Online]. Retrieved from http://www.triangle.co.uk/ciec/

Diem, K. G. (2001). Turn any science topic into hands-on fun. *Science Scope, Apr-01,* 46–49.

Education Week. (2002). Technology Counts '02: E-Defining Education. *Education Week on the Web.* [Online]. Retrieved from http://www.edweek.org/sreports/

Ferguson, E. L., & Hegarty, M. (1995). Learning with real machines or diagrams: Application of knowledge to real-world problems. *Cognition and Instruction, 13,* 129–160.

Fuller, E. W., May, D. H., & Butts, D. P. (1979). The science achievement of third graders using visual, symbolic, and manipulative instructional treatments. *Journal of Research in Science Teaching, 16,* 129–136.

Gabel, D., & Sherwood, R. (1980). The effect of student manipulation of molecular models on chemistry achievement according to Piaget level. *Journal of Research in Science Teaching, 17,* 75–81.

Glasson, G. E. (1989). The effects of hands-on and teacher demonstration laboratory methods on science achievement in relation to reasoning ability and prior knowledge. *Journal of Research in Science Teaching, 26,* 121–131.

Healy, J. M. (1999). *Failure to connect: How computers affect our children's minds—and what we can do about it.* New York: Simon & Schuster.

Koedinger, K. R., Anderson, J. A., Hadley, W. H., & Mark, M. A. (1997). Intelligent tutoring goes to the big city. *International Journal of Artificial Intelligence in Education, 8,* 30–43.

Kozma, R. B. (1991). Learning with media. *Review of Educational Research, 61,* 179–211.

Kozma, R. B. (1994). Will media influence learning? Reframing the debate. *Educational Technology Research and Development, 42,* 7–19.

Kuhn, D., Garcia-Mila, M., Zohar, A., & Andersen, C. (1995). Strategies of knowledge acquisition. *Monographs of the Society for Research in Child Development, 60*(4), 1–128.

Moore, J. L. (1993). Comparisons of a physical model and computer representations in reasoning and learning about linear functions. *Dissertation Abstracts International, 54*(02), 464A. (UMI No. 9317796)

National Research Council. (1996). *National science education standards*: *Observe, interact, change, learn.* Washington DC: National Academy Press.

Piaget, J. (1960). *The child's conception of physical causality.* Oxford, England: Littlefield, Adams.

Resnick, M. (1998). Technologies for lifelong kindergarten. *Educational Technology Research and Development, 46,* 43–55.

Richardson, A. E. , Montello, D. R., & Hegarty, M. (1999). Spatial knowledge acquisition from maps and from navigation in real and virtual environments. *Memory and Cognition, 27,* 741–750.

Riley, J. P. (1979). The influence of hands-on science process training on preservice teachers' acquisition of process skills and attitude toward science and science teaching. *Journal of Research in Science Teaching, 16,* 373–384.

Schauble, L. (1996). The development of scientific reasoning in knowledge-rich contexts. *Developmental Psychology, 32,* 102–119.

Shymansky, J. A, Kyle, W. C., & Alport, J. M. (1983). The effects of new science curricula on student performance. *Journal of Research in Science Teaching, 20,* 387–404.

Siegel, S., & Castellan, N. J. (1988). *Nonparametric statistics for the behavioral sciences* (2nd ed.). New York: McGraw-Hill.

Sowell, E. J. (1989). Effects of manipulative materials in mathematics instruction. *Journal for Research in Mathematics Education, 20,* 498–505.

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.

Technology and Learning. (2002). *Software Product Guide.* [Online]. Retrieved from http://www.techlearning.com/content/resources/product_guide/

Thompson, P. W. (1992). Notations, conventions, and constraints: Contributions to effective uses of concrete materials in elementary mathematics. *Journal for Research in Mathematics Education, 23,* 123–147.

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.

White, B. Y., & Frederiksen, J. R. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and Instruction, 16,* 3–118.