Direct Instruction vs. Discovery: The Long View

DAVID DEAN JR., DEANNA KUHN
Teachers College, Columbia University, New York, NY 10027, USA

Received 20 June 2006; revised 4 October 2006; accepted 12 October 2006

DOI 10.1002/sce.20194
Published online 14 December 2006 in Wiley InterScience (www.interscience.wiley.com).

ABSTRACT: D. Klahr and M. Nigam (2004) make a case for the superiority of direct instruction over discovery learning in students' mastery of the control-of-variables strategy central to the scientific method. In the present work, we examine acquisition of this strategy among students of the same age as those studied by Klahr and Nigam, as well as follow central features of their design in directly comparing the two methods. In contrast to their design, however, we follow progress over an extended time period and a range of equivalent tasks. Three groups of 15 fourth-grade students, of diverse socioeconomic background, were compared. One group engaged in 12 sessions over 10 weeks working on problems that required the control-of-variables strategy for effective solution. Another group engaged in the same activity, preceded by a session involving direct instruction on the control-of-variables strategy. A third group received only the initial direct instruction, without subsequent engagement and practice. In this longer term framework, direct instruction appears to be neither a necessary nor sufficient condition for robust acquisition or for maintenance over time. The patterns of attainment observed here point instead to a gradual and extended process of acquisition and consolidation. *Sci Ed* **91:**384–397, 2007

INTRODUCTION

A recent study by Klahr and Nigam (2004) presents evidence to support a claim of the superiority of direct instruction over discovery learning, a claim that a number of researchers

Correspondence to: Deanna Kuhn; e-mail: dk100@columbia.edu
Present address of David Dean Jr.: Center for AIDS and STD, School of Social Work, University of Washington, Seattle, WA 98104, USA.

(e.g., Kirschner, Sweller, & Clark, 2006; Mayer, 2004; Rittle-Johnson, 2006), as well as policymakers, now endorse. Klahr and Nigam administered brief direct instruction on the control-of-variables procedure to a large group of third- and fourth-grade students. They indicate that in order to facilitate comparison with their other condition, they used "an extreme type of direct instruction in which the goals, the materials, the examples, the explanations, and the pace of instruction are all teacher controlled" (p. 662). In their comparison condition, the same problem was presented but students were allowed to design their own procedures and received no instruction or feedback. The problem was one in which students were provided with appropriate materials and asked to determine how different variables (ball material and surface, length, and steepness of ramp) affected the distance that balls traveled after rolling down an incline. Students in the direct instruction condition, Klahr and Nigam report, outperformed those in the comparison condition on both direct and transfer assessments.

In the work presented here, we study acquisition of the same strategy that Klahr and Nigam (2004) studied (control of variables), among students of the same age, and incorporate the central features of their design. A major difference is time frame. We examine acquisition and maintenance over a longer time period—almost 6 months—in contrast to the single acquisition session, with transfer assessment 1 week later, in Klahr and Nigam's study. Furthermore, we follow students' mastery of this specific strategy across varied content over the entire period, whereas Klahr and Nigam's follow-up assessment shifted to a less well-defined variety of concepts associated with the scientific method. Under our more extended assessment, we hypothesized, the relative strengths of direct instruction and discovery might appear somewhat different. Given the significant policy implications of this debate, the "long view" adopted in the present work seems worthy of investigation.

The control-of-variables strategy studied by Klahr and Nigam (2004) is a key component of inquiry skills, which now appear in the American national curriculum standards for science (National Research Council, 1996) at every grade beginning with second or third through twelfth and appear in most state standards as well. In the national science standards, the goals of inquiry skill development for grades 5–8 are the following (National Research Council, 1996):

- identify questions that can be answered through scientific investigations;
- design and conduct a scientific investigation;
- use appropriate tools and techniques to gather, analyze, and interpret data;
- develop descriptions, explanations, predictions, and models using evidence;
- think critically and logically to make the relationships between evidence and explanations.

Under "design and conduct a scientific investigation," subskills identified include ". . . systematic observation, making accurate measurements, and identifying and controlling variables."

Klahr and Nigam's (2004) intention is to demonstrate that direct instruction is a more effective means of acquiring the control-of-variables strategy than is "discovery learning," which they define as the student discovering or constructing this skill for himself or herself. There exists a certain irony, if not conceptual incoherence, in their intention, in that the control-of-variables strategy is a component of inquiry skill and inquiry skill is broadly understood to mean skill in discovering or constructing knowledge for oneself. Nonetheless, it is conceivable that inquiry methods are not the best ones for teaching inquiry skills and Klahr and Nigam's proposition thus deserves empirical investigation. The counterintuitive nature of the proposition, however—in suggesting the superiority of a method other than

involving students in activities that demand inquiry as a means of fostering inquiry skills—means that claims for such superiority should be especially well documented, particularly in demonstrating the scope and stability of what students have acquired. Hence, the long view taken here.

METHOD

Participants

Participants were 44 fourth-grade students in a university-affiliated urban independent elementary school. The school is unique in that 50% of its enrollment consists of children of the university faculty and senior administrators, while the other 50% of spaces in the school are reserved for children of families in the surrounding, largely lower income neighborhood, chosen by lottery and provided sufficient financial aid to enable them to attend. The student body is thus more diverse than is typical in most schools with respect to race, ethnicity, parent education level, and student ability. Once students enroll in the school, they are not distinguished as to mode of entry. It was therefore not feasible, nor did we consider it appropriate, to identify them or examine their performance in the present study as a function of family (university or community) status.

Design

Previous assessments with fourth and sixth graders at this school (described below) had established that students in fourth grade showed no initial competence with respect to the strategy under investigation, a conclusion consistent with much previous research (see Kuhn, 2001, for review). Individual pretest assessment was therefore deemed unnecessary, and the design consists of postintervention comparisons of three groups who underwent different forms of intervention.

The three fourth-grade classes at the school each contained 15 students and had been composed so as to be equivalent with respect to children's gender, ability and achievement, and social characteristics. Although we did not have access to these data, school administrators informed us that median and range of scores on standardized academic achievement tests were equivalent across the three groups. These classes were randomly assigned to one of three conditions: (a) extended engagement with problems requiring the control-of-variables strategy for effective solution (henceforth referred to as the *practice* condition), hypothesized to give students opportunity to develop the strategy of interest, (b) single-session *direct instruction*, designed to teach the strategy, and (c) a third condition consisting of the combination of the first two (*direct instruction plus practice*). The groups are henceforth referred to as the PR, DI, and DI/PR groups, respectively.

The DI condition was modeled after Klahr and Nigam's (2004) direct-instruction procedures as closely as possible (see description below). Except for its extension over a much longer period of time, the practice condition resembled Klahr and Nigam's discovery condition in engaging students in a problem that required the control of variables strategy for effective solution and hence providing them opportunity to construct the strategy.

In addition to the overall analysis implied by the design, we addressed two specific research questions: (a) How does practice, with or without DI, compare to DI alone? (b) Do effects differ at different time intervals?

Tasks

Three parallel computer-based inquiry tasks were employed during practice and assessment: ocean voyage (OV), earthquake forecaster (EF), and avalanche hunter (AH). The

three are structurally identical and differ only in content. Each introduces five potential variables that may affect an outcome and asks the student to investigate and determine which do and which do not affect outcome. In each version, two of the five variables have no effect and the other three have additive (noninteractive) effects on outcome. Figures 1–3 show the main screens for earthquake forecaster. As each screen is displayed, a voiceover presents the identical text orally, thus eliminating any challenge that reading the text may have posed for any of the participants while at the same time accommodating those who prefer the visual mode.

Introductory screens describe the importance of identifying earthquake risk and explain the student's task as junior earthquake forecaster. The first interactive screen (Figure 1) asks students to identify the goal of their first investigation. The next (Figure 2) asks them to select a particular instance for examination. The next (Figure 3) presents the outcome associated with this instance and asks the student to draw conclusions (makes a difference, doesn't make a difference, or don't know, for each variable). A final screen (not shown here) offers the opportunity to make notes in an electronic notebook. The cycle then repeats four times, giving the student the opportunity to examine four instances and draw conclusions regarding each variable after each one. In the second through fourth iterations of the cycle, the results for the preceding cycle remain displayed (as shown in Figure 3).

To execute the task effectively, a student must access and compare two instances that differ with respect to the levels of only a single variable, in order to assess the effect of that variable, and then must draw the appropriate conclusion (based on whether the manipulation is associated with a change in the outcome variable). This procedure must be repeated to identify the causal or noncausal status of each of the variables.

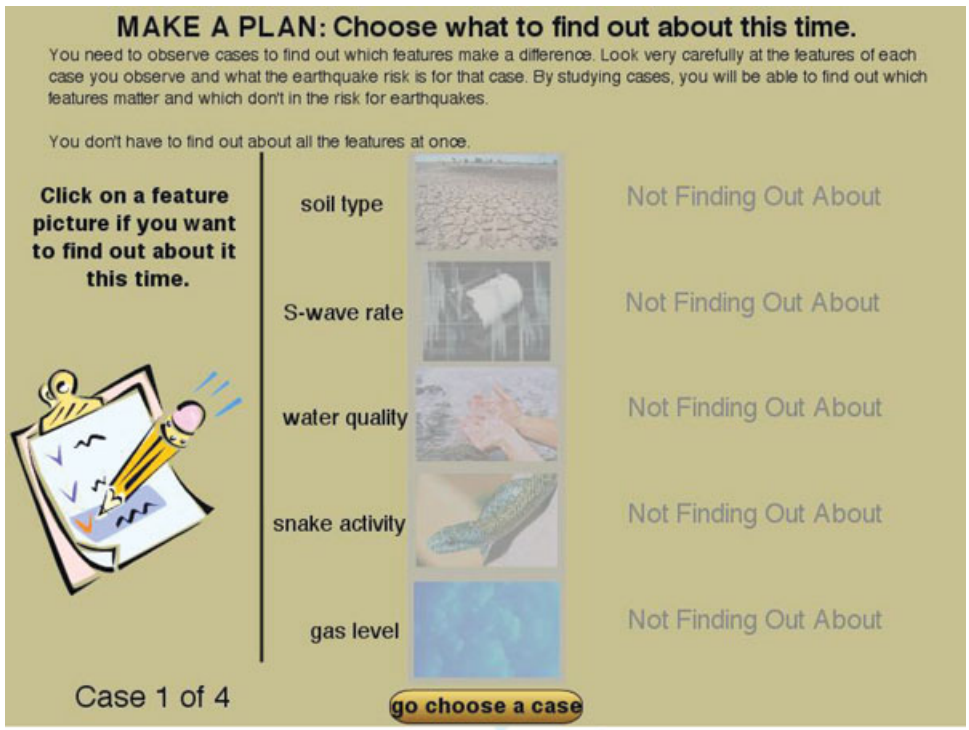

Figure 1. Find out screen. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

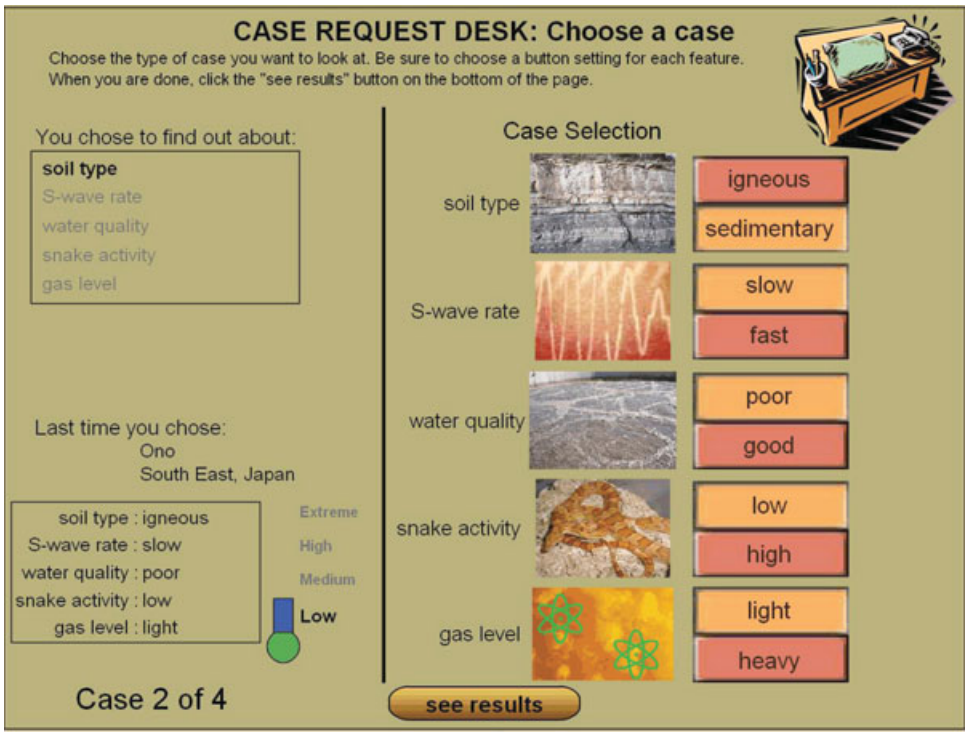

Figure 2. Case request screen. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

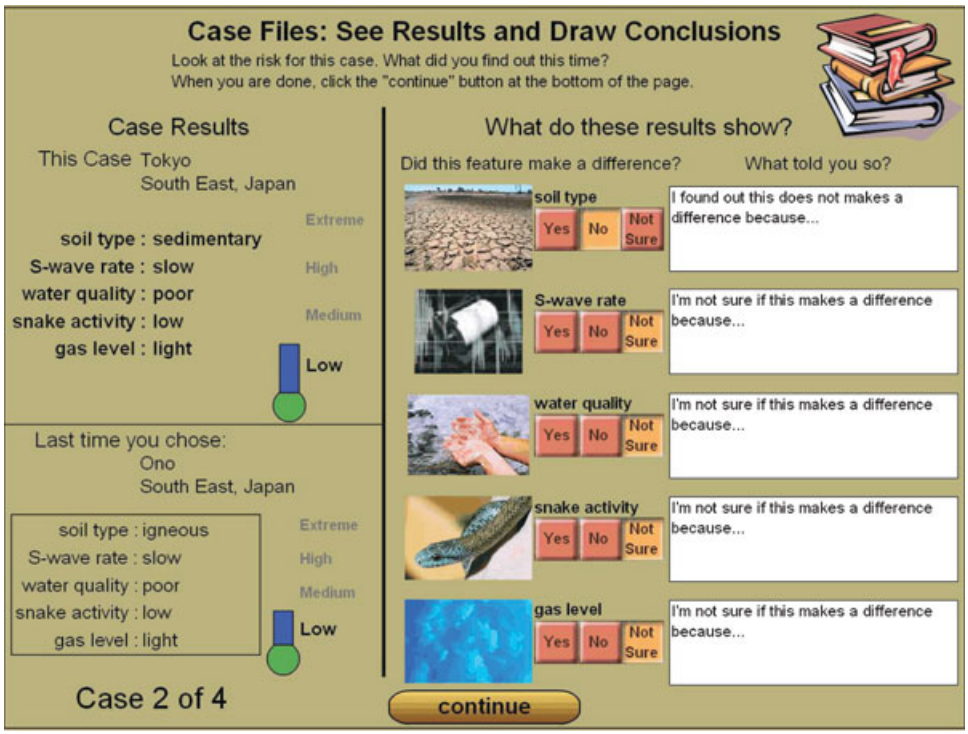

Figure 3. Results and conclusions screen. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

Pilot Assessment

Pilot assessment was undertaken during the spring of the year preceding the study to establish the general skill level of students at this school with respect to the skills of interest. For this purpose, the EF task was administered individually to two classes of that year's fourth graders (30 students) and one class of sixth graders (19 students). None of these students participated in the present study, which was conducted with fourth graders the following school year.

In their work on the EF task, none of these pilot fourth graders ever displayed the control-of-variables strategy, and two sixth graders displayed it. Neither of the sixth graders, however, followed his controlled comparisons of two instances with an appropriate inference. On this basis, it was concluded that fourth graders from this population show no competence in this strategy, and individual pretest assessment of the study participants was deemed unnecessary.

Procedure

Introductory and Direct-Instruction Session. Students in all conditions participated in an introductory session. Its purpose was to introduce the activity to all groups and in addition to present the direct instruction to those groups receiving it (DI and DI/PR groups). The instruction was modeled after that used by Klahr and Nigam (2004), except for the content which was mail-order music catalogs, chosen to be familiar and of interest to students. All groups underwent the introductory and postinstruction phases of this introductory session, to ensure group equivalence except for the direct instruction. Only the DI and DI/PR groups underwent the instruction phase. The session lasted approximately 45 minutes for each of these two groups and 35 minutes for the PR group since the direct instruction phase was omitted in their case.

Introduction (All Groups). The researcher introduced mock-ups of types of catalogs it was explained a CD company was contemplating using to advertise. The catalogs varied on four dimensions: format (booklet or foldout), color (multi or single), illustration (artist photos or CD covers), and number of CDs displayed (few or many). The student worked with a set of the 16 possible catalogs (actual physical mock-ups of the different kinds of catalogs). The student was asked to identify possible comparisons that would help to find out whether features made a difference or made no difference to catalog sales. It was indicated that sales records for each catalog were available, but no records were actually shown. Students were first asked to identify one comparison (pair of catalogs), and then a second comparison, to find out if one of the features (format) made a difference. The procedure was then repeated for a second feature (color), yielding a total of four comparisons constructed by each student.

Direct Instruction (DI and DI/PR Groups Only). Again following Klahr and Nigam (2004), the researcher then introduced a series of four comparisons and commented on them as follows:

> *Comparison 1* (confounded—format and color both varied). Is this a good comparison? No. Let me tell you why. This is a bad comparison because Pat changed both features. If you change both features in a comparison you can't tell which one makes a difference.

> *Comparison 2* (unconfounded—only format varied). Is this a good comparison? Yes. Let me tell you why. This is a good comparison because Pat only changed one feature in the comparison so Pat can be sure that it's that feature that made the difference.
> *Comparison 3* (unconfounded—only color varied). Same instruction as comparison 2.
> *Comparison 4* (confounded—format and color both varied). Same instruction as comparison 1.

Postinstruction Assessment (All Groups). The procedure was identical to that of the introductory phase, except that students were first asked to investigate a new feature (number of CDs) and then one of the features (color) that they had previously investigated.

Practice Sessions. These sessions, extending over multiple weeks, allowed students in the PR and DI/PR conditions opportunity to construct the control-of-variables strategy over time (or, in the case of DI/PR students, if they were able to do so, to apply and consolidate the direct instruction they had received at the initial session as they worked on these problems).

Practice-Only (PR) Condition. The initial practice session took place the following week and began with an introduction to the first computer program (OV) students would use. It was explained that if they investigated very carefully, they would be able to find out which features make a difference to outcome and which do not. It was also emphasized that what they find out may be different from what they think now might make a difference. They were also told:

> Even if you think that you have it all figured out, we're going to ask you to keep working on the problems and checking your conclusions a little longer to be absolutely sure, since sometimes peoples' conclusions change as they keep investigating. You want to be able to show others how you know what you're claiming is correct. It's going to take you more than one class period to figure everything out, so you'll have more chances after today to continue your investigation.

Finally, students were told that they would work in pairs and that they needed to discuss with their partner and make sure both partners were in agreement before they made any choices or decisions. Before pairs began working independently, the teacher used an audio/video projector to take the group as a whole through one cycle of the program, making sure they understood what to do at each point in the cycle.

Students worked on the OV task for a total of 12 sessions, at a frequency of once or twice per week (depending on the school schedule), over a total period of 10 weeks. In subsequent sessions, the teacher reminded students of the preceding information as necessary but the need for instruction became minimal and was phased out over the next few sessions. As additional practice, at the fifth session, and again at the ninth and eleventh sessions, a "claim sheet" was introduced, asking the pair of students to make a claim of causality or noncausality ("makes a difference" or "doesn't make a difference") about a feature and to indicate what evidence they had to show their claim was correct.

DI and DI/PR Conditions. The practice sessions for the DI/PR group were identical to those just described for the PR group. The DI group received their regular classroom science instruction in lieu of any practice sessions.

Assessment and Maintenance Sessions

Following the 10-week period just described, the procedure from this point was identical for all three groups, except that the DI group did not participate in the maintenance phase described below.

First Posttest Assessment (Familiar Content). The first posttest assessment occurred the week after the PR and DI/PR groups had completed 12 OV practice sessions. All students were assessed individually. The student was asked to demonstrate using the OV program "how to find out whether a feature makes a difference or doesn't make a difference." If a student indicated she or he already knew the effect of all features, the student was asked to imagine that another student disagreed and to demonstrate how it could be demonstrated to that student that the feature does or does not make a difference.

Transfer Assessment (Unfamiliar Content). During this same week, students in all groups were assessed individually using a new task identical in structure but differing in content (EF). Procedure was identical to that for the preceding assessment.

Maintenance Sessions (PR and DI/PR Groups Only). To consolidate and help ensure that skills attained during the practice sessions would be maintained, beginning the next week PR and DI/PR groups engaged in additional practice, once per week for 5 weeks, working with the EF task. The procedure was identical to that used in the earlier practice sessions.

Delayed Posttest Assessment (Familiar Content). During the next week, all students underwent individual assessment with the EF task. The procedure was identical to that used in the first two posttest assessments.

Delayed Transfer Assessment (Unfamiliar Content). During this same week, all students underwent individual assessment with a new task (AH). The procedure was identical to that used in previous assessments.

RESULTS

Immediate Effects of Direct Instruction

Before conducting the main analyses, we wished to confirm that the direct instruction had been successful in producing the anticipated learning. Accordingly, we conducted an analysis of performance at the initial instruction session itself. Consistent with the findings of microgenetic research (Kuhn, 1995; Siegler, 2006), students did not perform consistently across the four comparisons that constituted the postinstruction assessment at this initial session; most showed a mixture of correct (unconfounded) and incorrect (confounded) comparisons. The two DI groups were combined for analysis, since their experience had not differed at this point. For the 29 students receiving DI, the mean number of correct (unconfounded) comparisons constructed was 1.76 (SD = 1.13), of a possible 4. In contrast, the mean for those in the PR condition was 1.00 (SD = 0.66).[1] This difference was significant, $t(42) = 2.25$, $p = .030$, indicating the instruction had an effect.

[1]Because no justifications were required, some of these comparisons could have been unconfounded by chance and do not necessarily represent true competence.

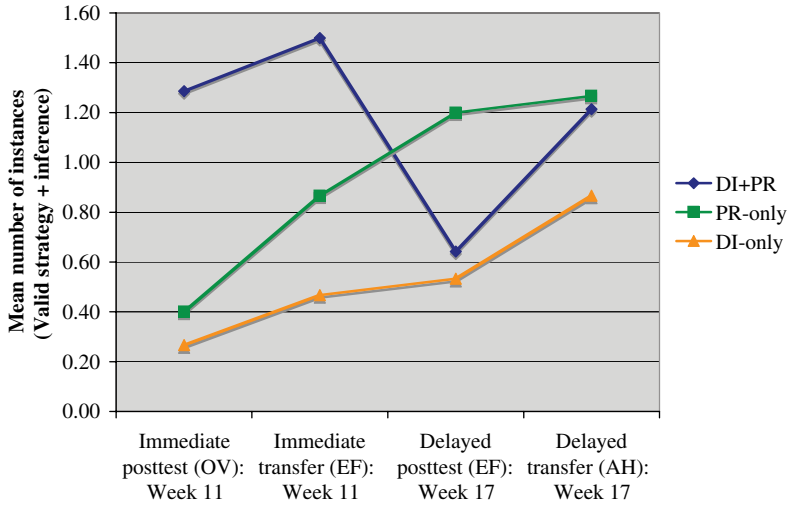

Figure 4. Means for valid strategy + inference by group. Maximum score = 3. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

Performance of Groups Over Time

Strategy and Inference. In analyses of the four main assessments indicated above, and again following Klahr and Nigam (2004) as closely as possible, we considered students to have demonstrated mastery of the control of variables strategy if they compared two instances that differed with respect to a single variable and then drew an appropriate inference with respect to the varied variable (thereby confirming that selection of the two unconfounded instances had not been by chance). At each of the four assessments described above, the number of such correct comparison–inference sequences (of a possible 3) executed by a student was identified. (No comparisons were possible following the first of the four instances examined because only one instance was available.) Means across students for each of the four assessments are shown in Figure 4.

Because we had multiple objectives in examining these data, we followed the recommendation of a statistical specialist and conducted an overall ANOVA to establish that experimental condition had a significant effect, as well as two specific comparisons to identify the effects of the variables of theoretical interest, DI and time.

Repeated measures ANOVA of the data shown in Figure 4 yielded a main effect for group, $F(2, 41) = 4.74$, $p = .014$, a main effect for assessment occasion, $F(3, 123) = 2.85$, $p = .040$, and also a significant interaction between them, $F(6, 123) = 2.72$, $p = .016$. A comparison of the DI group ($M = 0.53$), against the two practice groups ($M = 1.16$ for the DI/PR group and 0.93 for the PR group), was also significant, $F(1, 41) = 11.34$, $p < .05$ (although the two PR groups did not differ significantly from one another). Also significant was a contrast between the two initial and two delayed assessments, with correct explanations more frequent at the delayed assessments as reflected in Figure 4, $F(1, 41) = 6.09$, $p < .05$.

Explanation. Each of the three assessment tasks required students to explain the basis for each inference made. Regarded as adequate explanations for indeterminacy inferences (those not asserting a definite conclusion) were those that referred to the evidence that had been generated and indicated it was insufficient to allow an inference (e.g., "I haven't found

out yet"). Regarded as adequate explanations for determinant inferences (the feature makes a difference or does not make a difference) were explanations that referred to evidence that had been generated and that was sufficient to support the inference. Such explanations were counted as correct only if they accompanied a valid (nonconfounded) comparison and appropriate determinant inference.

The number of correct explanations (of a possible 20, the maximum number of inferences the student had an opportunity to make given four instances, each involving five variables) was identified for each participant at each assessment. Means across participants for each of the four assessments are shown in Figure 5. Repeated measures ANOVA yielded no significant effects for group or for assessment occasion. However, the interaction of group and assessment type was significant, $F(6, 123) = 2.40$, $p = .031$, The contrast between the two immediate and two delayed assessments was significant, with correct explanations more frequent at the delayed assessments as reflected in Figure 5, $F(1, 41) = 5.54$, $p < .05$. The contrast between the DI group (M = 5.20) and the two practice groups (means = 7.38 for the DI/PR group and 8.82 for the PR group), however, did not reach significance.

Patterns of Performance Across Time and Tasks. The patterns in Figures 4 and 5 do not tell the complete story of performance over time, as they do not indicate how many individuals contributed to successful group performance. Given the means were far from ceiling, they could reflect the very high achievement of a few or the more modest achievement of a larger number. For this reason, we also examined results in the form seen in Figure 6, which shows the proportion of students in each group who showed mastery a majority of the time (two of three possible instances of valid strategy followed by appropriate inference). We also identified the proportion of students in each group who ever showed the correct strategy–inference sequence (not shown); these proportions were of course higher (mostly in the range of 60%–75%), but the patterns across groups were similar and hence this figure is omitted. We also identified the proportions by group who showed the correct strategy–inference–explanation sequence a majority of the time. With the added criterion of correct explanation, these proportions are lower (mostly in the range of 10%–30%) but again are not shown as the pattern was similar.

It is Figure 4, in conjunction with Figure 6, that tells the clearest story. The DI/PR group, we see, shows an initial advantage, one that is not matched by the group who received

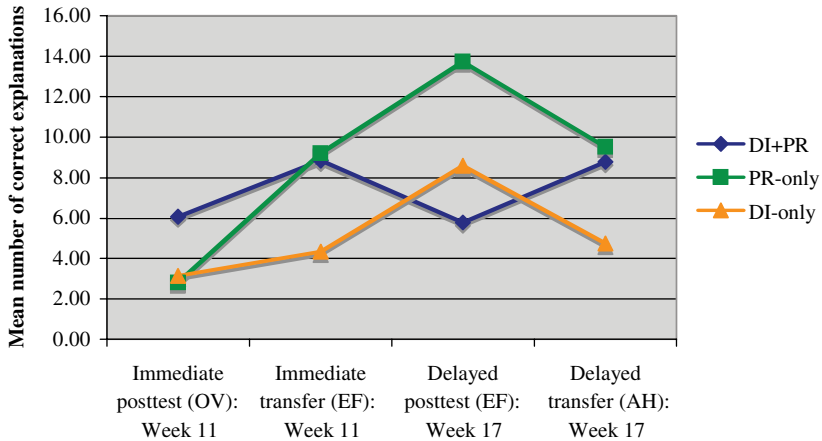

Figure 5. Means for explanation by group. Maximum score = 20. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

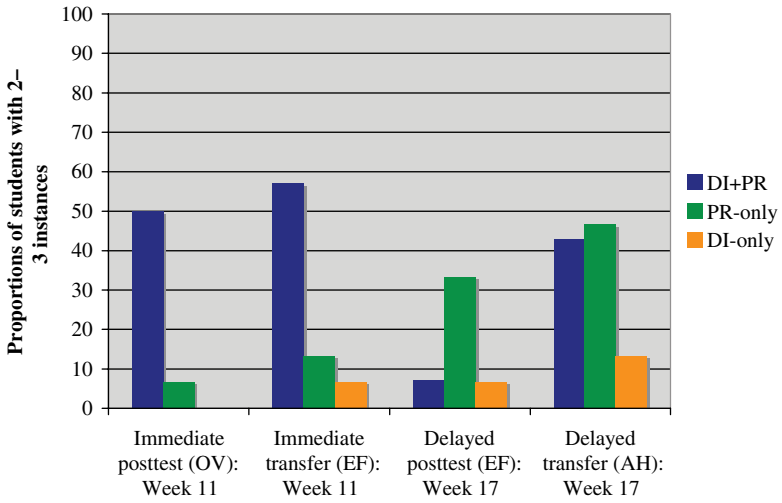

Figure 6. Proportions of successful students by group and time. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

only DI without the additional benefit of PR. This high level of performance, however, is not maintained with additional time, dropping sharply at the third assessment although recovering to a good extent when the final new task is introduced.

The DI group, in contrast, shows negligible effect of the instruction by the time of the initial assessment (at 11 weeks). This group improves steadily, although modestly, over time and assessments, an effect that may be attributable to the practice effect of the assessments themselves.

The PR group also improves steadily. At the first assessment, hardly any students in this group have consolidated use of the new strategy (Figure 6), but the additional practice on a different new task between second and third assessments enables nearly half of them to do so.

Performance on a new task at the end of an extended time interval (fourth assessment) might be regarded as the most significant measure of students' ultimate gains. Here the results are clear. The two practice groups, with or without direct instruction, do better than the DI-only group.

DISCUSSION

The present results replicate Klahr and Nigam's (2004) finding that among children of this age, brief direct instruction is capable of producing a significant level of correct performance with respect to the control of variables strategy, immediately following instruction. Examined over a longer time frame, however, our results indicate that direct instruction is neither necessary nor sufficient to accomplish this goal. Klahr and Nigam's study, note, does not speak to the issue of maintenance over time, nor does it directly address transfer, since the target strategy is not specifically assessed in their transfer assessment.[2] In the present study, in contrast, we examine transfer of the specific strategy under consideration across multiple parallel versions of the task over an extended time period.

[2]Earlier studies by Klahr and associates, e.g., Chen and Klahr (1999), similarly have limitations with respect to their demonstration of transfer and/or maintenance. For further discussion, see Kuhn and Dean (2005), Kuhn and Franklin (2006), and Zimmerman (2006).

The trends reflected in Figure 4 are informative in several respects. At the initial assessment, some 11 weeks after the direct instruction session, the instruction still shows an effect, but only when it is combined with regular practice (DI/PR group). DI without this practice is not sufficient to sustain, after 12 weeks or longer, a result any better than what can be achieved without it, merely through engagement, or practice. This practice, with or without DI, appears sufficient to produce at least as strong, if not stronger, performance. DI is not a necessary component.

Such practice, however, does not achieve its maximum effect after 12 sessions working with a single problem. The new content introduced at the second assessment shows some suggestion of reinvigorating the engagement of the PR group, some of whom we suspected may have developed habitual modes of response to the first task toward the end of the 12 sessions they worked on it. Further advance is seen, however, after they had the opportunity for sustained engagement with a different problem (third assessment). At this point, the initial advantage shown by the DI/PR group drops off sharply and the PR group in fact shows a higher level of performance, although by the fourth assessment, with new content, performance of the DI/PR and PR groups is identical with respect to both strategy–inference (Figure 4) and explanation (Figure 5). DI, then, has yielded no long-term advantage. The modest improvement over time shown by the DI group suggests that, rather than DI, it is the engagement with the assessment tasks that functions as a form of practice that contributes to improvement of performance, possibly along with any natural mental development that might occur over this 6-month period.

The present study, note, does not purport to demonstrate the merits of engagement/practice methods, compared to direct instruction, with regard to efficiency of instruction. Our interest is not in establishing how fast the strategic understanding examined here can be acquired, but rather how well it can be acquired. Students in the two practice conditions spent much greater "time on task" than those in the direct instruction condition. Given this practice led to significant and lasting gains in strategic understanding for the majority of students, do we then need to ask whether these gains could not be accomplished more quickly, with less student time devoted to their accomplishment? If so, does the time saved warrant the introduction of direct instruction that would not be necessary if more time were taken? Given the centrality of scientific investigation to the epistemology of science (not to mention to the U.S. science education standards), we do not see the student time devoted to attaining and consolidating this strategic and metastrategic understanding as needing to be minimized. Indeed, Klahr and Nigam's adoption of the experimental psychologist's focus on efficiency of instruction departs significantly from the perspectives of those in the field of science education concerned with the teaching of science process skills (Duschl & Grandy, 2005; Metz, 2004; Reiser, 2004; Sandoval, 2005: White & Frederiksen, 2005).

All of the conclusions drawn here of course hold only for the specific age group investigated. Older or younger students could exhibit different effects of the various forms of intervention we have examined, and replication of the kinds of comparisons undertaken here is necessary to determine whether this is the case. Yet, the specific competency examined here is highlighted as a core component of the scientific inquiry skills that are now mandated in virtually all state and national science education standards by the fourth grade (National Research Council, 1996). How mastery of these skills is best achieved at this particular age level is thus a matter having enormous implications for current educational policy and practice.

In this context of educational implications, a final point to be noted is that our results suggest that such competency remains fragile at this age level. Our own and others' extensive work with this age group shows negligible spontaneously emerging competence at this age, even among academically able children, a finding corroborated in the pilot assessment

reported here. Despite the extensive practice fourth graders in the present study underwent, some of them still showed minimal to no indication of proficiency. Moreover, even among those who did show indications of having attained proficiency, the norm remains a mixture of usage of correct and incorrect strategies. Thus, both interindividual and intraindividual variability are the norm—a result consistent with the microgenetic literature on repeated engagement with the same or similar tasks (Kuhn, 1995; Kuhn & Franklin, 2006; Siegler, 2006; Siegler & Crowley, 1991). This pattern further supports our claim that a single instruction session is insufficient to produce the desired mastery. Although we have focused on only one key component of scientific inquiry here, development of the desired competence has a variety of interconnected components, as we have argued elsewhere (Kuhn & Dean, 2005), and does not occur overnight. In terms of both scope and time frame, it would appear to be a gradual and extended acquisition process that researchers hoping to contribute to educational practice should seek to understand.

The research reported here is drawn from a dissertation presented by the first author to the second author in partial fulfillment of the requirements for a Ph.D. degree at Teachers College, Columbia University. The authors would like to thank the students and teachers who participated in this work: Erica Chutuape, Jenny Lander, Mari McGrath, Nani Pease, and Nava Silton for assisting in data collection; Daniel Rubin for programming the data extraction program; Jane Monroe for statistical advice; and Jared Katz for programming the intervention task and assisting with the analysis. The authors would also like to thank the anonymous reviewers for helpful comments on earlier versions of the manuscript.

REFERENCES

Chen, Z., & Klahr, D. (1999). All other things being equal: Children's acquisition of the control of variables strategy. Child Development, 70(5), 1098–1120.

Duschl, R., & Grandy, R. (2005). Reconsidering the character and role of inquiry in school science: Framing the debates. Unpublished manuscript, Rice University, Houston, TX.

Kirschner, P., Sweller, J., & Clark, R. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist discovery, problem-based, experiential, and inquiry-based teaching. Educational Psychologist, 41, 75–86.

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

Kuhn, D. (1995). Microgenetic study of change: What has it told us? Psychological Science, 6, 133–139.

Kuhn, D. (2001). Why development does (and doesn't) occur: Evidence from the domain of inductive reasoning. In R. Siegler & J. McClelland (Eds.), Mechanisms of cognitive development: Neural and behavioral perspectives. Mahwah, NJ: Erlbaum.

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

Kuhn, D., & Franklin, S. (2006). The second decade: What develops (and how)? In W. Damon & R. Lerner (Series Eds.), D. Kuhn & R. Siegler (Vol. Eds.), Handbook of child psychology: Vol. 2. Cognition, perception, and language (6th ed.). Hoboken, NJ: Wiley.

Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? American Psychologist, 59, 14–19.

Metz, K. (2004). Children's understanding of scientific inquiry: Their conceptualization of uncertainty in investigations of their own design. Cognition and Instruction, 22, 219–290.

National Research Council (1996). The National Science Education Standards. Washington DC: National Academy Press.

Reiser, B. (2004). Scaffolding complex learning: The mechanisms of structuring and problematizing student work. Journal of the Learning Sciences, 13, 273–304.

Rittle-Johnson, B. (2006). Promoting transfer: Effects of self-explanation and direct instruction. Child Development, 77, 1–15.

Sandoval, W. (2005). Understanding students' practical epistemologies and their influence. Science Education, 89, 634–656.

Siegler, R. (2006). Microgenetic studies of learning. In W. Damon & R. Lerner (Series Eds.), D. Kuhn & R. Siegler (Eds.), Handbook of child psychology: Vol. 2. Cognition, perception, and language (6th ed.). Hoboken, NJ: Wiley.

Siegler, R., & Crowley, K. (1991). The microgenetic method: A direct means for studying cognitive development. American Psychologist, 46(6), 606–620.

White, B., & Frederiksen, J. (2005). A theoretical framework and approach for fostering metacognitive development. Educational Psychologist, 40, 211–223.

Zimmerman, C. (2006). The development of scientific reasoning skills: What psychologists contribute to an understanding of elementary science learning. Final draft of a report to the National Research Council Committee on Science Learning Kindergarten through Eighth Grade. Washington, DC: National Research Council.