Research Report

Is Developing Scientific Thinking All About Learning to Control Variables?

Deanna Kuhn and David Dean, Jr.

Columbia University

ABSTRACT—*Academically low-performing urban sixth graders engaged in inquiry activity received a suggestion that they focus their investigation on the role of a single factor. This suggestion had significant effects on their use of a superficially dissimilar strategy—controlling the variation of other factors. This latter strategy has received the lion's share of attention in research on the development of scientific reasoning. These results have implications, we propose, for what undergoes development with respect to scientific thinking and how this development can best be facilitated.*

"Inquiry" now appears as a curriculum goal in the American national curriculum standards for science (National Research Council, 1996) for every grade from 2nd or 3rd up through 12th and appears in most state standards as well. The national goals of inquiry learning for Grades 5 through 8, for example, 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. (p. 145)

Under "Design and conduct a scientific investigation," subskills identified include "systematic observation, making accurate measurements, and identifying and controlling variables" (p. 145).

Ideally, educators' efforts to implement these standards should be based on a sound grounding of psychological knowledge about how these skills develop, but this has not been the case (Kuhn, 2005). Psychologists, following from the work of Inhelder and Piaget (1958), have focused their research attention almost exclusively on children's developing skill in controlling variables in their design of experiments (Zimmerman, 2000). Most recently, Klahr and Nigam (2004) have presented evidence that brief direct instruction effectively teaches this skill, and on this basis they advocate such instruction as the most efficient, desirable method of developing inquiry skills.

Children in the preadolescent age range, however, have also been shown to develop inquiry skills without direct instruction if they engage in repeated encounters with situations that require these skills. They are unlikely to show much progress in the brief single session that Klahr and Nigam (2004) employed as a comparison condition, but microgenetic studies (Kuhn, 1995; Siegler & Crowley, 1991) show characteristic patterns of advancement over time and sustained engagement (Kuhn, Black, Keselman, & Kaplan, 2000; Kuhn, García-Mila, Zohar, & Andersen, 1995; Kuhn, Schauble, & García-Mila, 1992; Schauble, 1990, 1996).

What might be the strengths of this labor-intensive, time-inefficient means of fostering the development of inquiry skills relative to efficient, brief, direct instruction? The most important strength, we argue, has to do with the metastrategic level of understanding that ideally accompanies strategic learning (Kuhn, 2001a, 2001b). When individuals intentionally select a strategy they have available to apply to a task, they do so with some understanding (correct or incorrect) that this strategy is one that will serve their objectives. The strategy is selected and executed voluntarily. This metastrategic understanding cannot be assumed to be present or of the same quality when they are

Address correspondence to Deanna Kuhn, Box 119, Teachers College, Columbia University, New York, NY 10027; e-mail: dk100@columbia.edu.

instructed to use a strategy. The implications are significant with respect to how individuals may act when the instructional context is withdrawn and they resume voluntary control of their own behavior. These longer-term outcomes are not known in the case of Klahr and Nigam's (2004) study as the authors did not report specifically on students' continued use of the control-of-variables strategy in situations demanding it.

Another potential disadvantage of direct instruction is that it may be unduly narrow, and hence also constrain the development of meta-level understanding. Inquiry is a complex, multifaceted activity. It is more than a single strategy. If educators are to achieve the much-desired objective of involving students in "authentic" science (Chinn & Malhotra, 2002), the integrity of the whole must be respected. In the present study, we addressed this issue by demonstrating in an experimental design that a minimal intervention with respect to an aspect of inquiry activity that has been largely ignored as peripheral yields significant effects on an aspect that has long been treated as constituting the activity's core.

Participants were from a population of sixth-grade students at a low-performing inner-city public middle school where the large majority of students are academically at risk. In contrast to their counterparts (and even younger students) from more advantaged populations, students from this population had shown the ability to make only limited progress in developing inquiry skills through engagement and exercise (Keselman, 2003), and we contemplated introducing some more structured intervention, which we ultimately did (Kuhn, 2005; Kuhn & Dean, 2005). We speculated, however, that offering the students a modest initial suggestion might make their inquiry activity more productive, and this is the effort we report on here.

In particular, weaknesses in the inquiry process arise long before students get to the phase of designing and interpreting experiments. A first, critical phase is formulating a question to be asked. Unless students understand the purpose of the activity as seeking information that will bear on a question whose answer they do not already know, inquiry often degenerates into an empty activity of students' securing observations for the purpose of illustrating what they already take to be true (Kuhn, 2002, 2005). But it is not sufficient simply for the student to understand the need to formulate a question; the student must also be able to formulate effective questions. In a context of multiple variables potentially affecting an outcome, students who have developed an understanding of the need to access an available database as a source of information may nonetheless still initially pose ineffective questions, in particular because they aim to discover the effects of all variables at once. It may be this ineffective intention that leads them to simultaneously manipulate multiple variables (in effect, overattending to them, rather than underattending by failing to control them, as is typically assumed). In the present study, therefore, our intervention was simply to suggest to students that they try to find out about only one thing at a time.

METHOD

Participants

Participants were 30 sixth-grade science students in an urban public school serving a lower-income population. Students at this school are identified as academically at risk, and most perform below grade level. The students were almost all of African American and Hispanic ethnicity and 11 or 12 years of age. Of the 30, 17 were female and 13 male. Twelve students served in an experimental condition, and the remaining 18 in a control condition. An additional 12 students participated in an alternative control condition.

Procedure

Because previous work had repeatedly established poor pretest performance in this population, a posttest-only design was employed. Students in the experimental group participated in twelve 45-min sessions over 8 weeks, working in pairs with the Earthquake Forecaster inquiry software during their science period (as did students in the alternative control condition, who worked with Earthquake Forecaster during the same period without the additional intervention the experimental group received). Students in the main control group remained in their regular science classes. At the end of this period, students in all groups underwent an individual assessment with this same program.

Experimental Group

Initial Session. Earthquake Forecaster asks students to investigate five binary variables—water pollution (high or low), water temperature (cold or hot), soil depth (deep or shallow), soil type (igneous or sedimentary), and elevation (high or low)—and ascertain their role in earthquake risk. In each of four cycles of investigation during a session, students choose which of the variables they want to find out about (with the option of choosing one or more) and then are able to select a site that reflects a combination of variable levels of their choice (e.g., they choose whether they wish to see a site having high or low water pollution, high or low elevation, and so forth for the other three variables). An outcome appears, in the form of a gauge displaying the risk level (one of four alternatives, from low to extreme). Students are then asked to draw an inference regarding each of the five variables, indicating whether it does make a difference in earthquake risk, it does not make a difference, or they are unsure whether it makes a difference. Students are then prompted to make any notes they wish to in an electronic notebook.

At this point, a second investigation cycle begins, with the variable levels and outcome from the preceding cycle remaining displayed in a corner of the screen. The process is repeated for the third and fourth cycles. At the end of the fourth cycle, students are thanked for participating, and the program shuts down.

Second and Subsequent Sessions. The researcher introduced the second and subsequent sessions with this additional suggestion:

> Today let's try to find out about just one feature to start. A lot of you disagree about the soil type—whether it makes a difference if it's igneous or sedimentary. Today let's all try to find out for sure about the soil type to figure out if it has anything to do with the earthquake risk.

At each session, a different variable was suggested as the focus of investigation.

Immediate Assessment. At a final session with this program, students worked alone on Earthquake Forecaster. The suggestion to focus on a particular variable was not included.

Transfer Assessment. At the next session, students worked alone on a parallel inquiry program, Ocean Voyage. This program is structurally identical to Earthquake Forecaster and varies only in content. In Ocean Voyage, students investigate whether variables such as crew size and sail type affect a ship's progress toward its destination.

Delayed Assessment. Three months following the transfer assessment, students again spent an individual session working with Earthquake Forecaster.

Control Groups

Students in the main (assessment-only) control group engaged in only a single individual session working with Earthquake Forecaster. This was their only assessment.

In the alternative (practice) control group, participants were involved in the same weekly engagement with Earthquake Forecaster as participants in the experimental group, but did not receive the suggestion to focus their investigations on a single variable. The purpose of including this alternative control group was to establish that any superiority of the experimental group was attributable to the manipulation itself, rather than to the practice provided by engagement with the program. This group received the immediate assessment.

RESULTS

Our initial question was whether the manipulation was effective. Did students accept and act on the suggestion to focus their investigation on the effect of a single variable? Although observation during the intervention confirmed that they did, of more interest was whether they would do so when the suggestion was omitted and there was no direct influence on them in this respect. The results were clear-cut. At the immediate assessment, each student participated individually in four cycles of Earthquake Forecaster, and in each cycle, every participant indicated only a single variable as the one he or she intended to find out about. By contrast, students in the main control group intended to find out about a single variable only 11% of the time, and 83% of the time they intended to find out about three or more variables in a single comparison. The mean number of variables a student in the control group intended to find out about was 3.1. (Results were comparable in the alternative control group.)

Of ultimate interest, however, was the effect of the manipulation on students' ability to investigate effectively and draw valid inferences as a result of their investigations. We took valid inferences to be the ultimate indicator of successful investigation, defining a valid inference as a determinate inference (i.e., the variable makes a difference or the variable does not make a difference) that the evidence generated supports adequately. Given four instances to compare across four investigative cycles in a session, a student was able to make a maximum of three valid inferences at each assessment. As shown in Table 1, 75% of students in the experimental group made mostly or exclusively valid inferences at the immediate assessment on Earthquake Forecaster. In contrast, no students in the main control group did so. In the alternative control group, inference performance was also poor, confirming that the superiority of the experimental group was not attributable to engagement with the activity itself.

TABLE 1

Valid Inferences as a Function of Group and Assessment

	Percentage of students			
Group and assessment	Two or three valid inferences	One valid inference	No valid inferences	Mean number of valid inferences per student
Experimental group				
Immediate assessment	75	8	17	1.70
Transfer assessment	33	33	33	0.90
Delayed assessment	25	42	33	1.00
Main (assessment-only) control group: immediate assessment	0	44	56	0.40
Alternative (practice) control group: immediate assessment	8	25	67	0.40

The difference between the experimental and main control groups in mean number of valid inferences at the immediate assessment was significant, $t(28) = 4.79$, $p < .01$. The experimental group's performance dropped off at the transfer assessment with the Ocean Voyage program, as it did 3 months later with the original Earthquake Forecaster program. A comparison of the experimental group at the transfer and delayed assessments with the main control group (immediate assessment) showed a significant difference at the .05 level for the delayed assessment, $t(28) = 2.24$, $p = .033$, and a nearly significant difference for the transfer assessment, $t(28) = 1.99$, $p = .057$.

DISCUSSION

The implications of these simple data are straightforward, we believe. Historically, the literature on the development of scientific thinking has overemphasized control of variables, almost to the exclusion of any other aspects of the process of scientific inquiry. The present results suggest that this focus has been misplaced, or, at the very least, has constrained investigation of the development of scientific thinking. In particular, the data point to the importance of the initial question-formulating phase of scientific inquiry (whether conducted by novice or expert scientists), arguably because it organizes and gives meaning to the activity that follows (Lehrer, Schauble, & Petrosino, 2001).

In this report, we have examined the question-formulating phase only in the rather narrow form of a decision as to which one of a number of potential variables will be the object of investigation. Clearly, question formulation is a broader endeavor. Yet the present results demonstrate the significance of novice scientists' gaining command of this key phase of question asking. Developing an awareness and appreciation of the task goal of assessing the effect of one variable carries forward to the later (design and inference) phases of inquiry. The novice investigator may eventually cease to vary other variables across two-instance comparisons because of an increasing sense that those variables are not relevant to the comparison being made. Once the other variables are left alone, and thereby "neutralized" as Inhelder and Piaget (1958) described it, the way is prepared for increased usage and increasing metastrategic understanding of the power of controlled comparison.

The "displacement" of instruction—from one aspect of scientific investigation to another, superficially dissimilar one—that we observed supports the claim that there is more that undergoes development than isolated procedures. Metastrategic competence—the ability to reflect on and manage strategic knowledge and to relate it to task goals (Kuhn, 2001a, 2001b; Kuhn & Franklin, in press; Kuhn & Pease, in press; Shrager & Siegler, 1998; Siegler, in press)—is a critical aspect of what develops, one that figures centrally in the matter of how individuals will choose to employ a newly learned procedure. The fact that almost half of the participants in the main control condition displayed the valid inference strategy occasionally (see Table 1) is consistent with the now sizable body of data based on microgenetic research methodology (Siegler, in press). Yet none of these students employed the valid strategy as their dominant strategy. The challenge, then, is not to teach students how to execute the strategy, but rather to help them understand why to use it—knowledge that is metastrategic in nature. The fact that frequency of usage recedes when the problem environment is altered or the dense problem environment is suspended altogether (the transfer and delay assessments, respectively) is consistent both with microgenetic research (Kuhn, 1995; Siegler, in press) and with a sociocultural perspective on learning (Anderson, Greeno, Reder, & Simon, 2000).

In a more applied vein, previous work with this population (in contrast to more advantaged populations), as we noted earlier, showed that they make only minimal progress by means of extended engagement with the task environment alone. It does not follow that direct instruction is the only, much less the best, approach to employ with such populations. There exists an unfortunate history of such an assumption being made. In the present case, we believe, our findings suggest that this approach may offer such populations less than they need and can profit from.

Acknowledgments—This research was supported by National Science Foundation Grant 0206043 to the first author.

REFERENCES

Anderson, J.R., Greeno, J.G., Reder, L.M., & Simon, H.A. (2000). Perspectives on learning, thinking, and activity. *Educational Researcher*, *29*(4), 11–13.

Chinn, C.A., & Malhotra, B.A. (2002). Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks. *Science Education*, *86*, 175–218.

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. New York: Basic Books.

Keselman, A. (2003). Supporting inquiry learning by promoting normative understanding of multivariable causality. *Journal of Research in Science Teaching*, *40*, 898–921.

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. (2001a). How do people know? *Psychological Science*, *12*, 1–8.

Kuhn, D. (2001b). Why development does (and doesn't) occur: Evidence from the domain of inductive reasoning. In R.S. Siegler & J. McClelland (Eds.), *Mechanism of cognitive development: Neural and behavioral perspectives* (pp. 221–249). Mahwah, NJ: Erlbaum.

Kuhn, D. (2002). What is scientific thinking and how does it develop? In U. Goswami (Ed.), *Blackwell handbook of childhood cognitive development* (pp. 371–393). Malden, MA: Blackwell.

Kuhn, D. (2005). *Education for thinking*. Cambridge, MA: Harvard University Press.

Kuhn, D., Black, J., Keselman, A., & Kaplan, D. (2000). The development of cognitive skills to support inquiry learning. *Cognition and Instruction*, *18*, 495–523.

Kuhn, D., & Dean, D., Jr. (2005). *Scaffolding the development of inquiry skills in academically at-risk populations*. Unpublished manuscript, Teachers College, Columbia University, New York.

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

Kuhn, D., García-Mila, M., Zohar, A., & Andersen, C. (1995). Strategies of knowledge acquisition. *Monographs of the Society for Research in Child Development*, *60*(4, Serial No. 245).

Kuhn, D., & Pease, M. (in press). Do children and adults learn differently? *Journal of Cognition and Development*.

Kuhn, D., Schauble, L., & García-Mila, M. (1992). Cross-domain development of scientific reasoning. *Cognition and Instruction*, *9*, 285–327.

Lehrer, R., Schauble, L., & Petrosino, A.J. (2001). Reconsidering the role of experiment in science education. In K. Crowley, C. Schunn, & T. Okada (Eds.), *Designing for science: Implications from everyday, classroom, and professional settings* (pp. 251–277). Mahwah, NJ: Erlbaum.

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

Schauble, L. (1990). Belief revision in children: The role of prior knowledge and strategies for generating evidence. *Journal of Experimental Child Psychology*, *49*, 31–57.

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

Shrager, J., & Siegler, R.S. (1998). SCADS: A model of children's strategy choices and strategy discoveries. *Psychological Science*, *9*, 405–410.

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

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

Zimmerman, C. (2000). The development of scientific reasoning skills. *Developmental Review*, *20*, 99–149.

(Received 9/27/04; Revision accepted 11/15/04; Final materials received 11/16/04)