Diverse Descriptions of Experimental Practice as Supports for Learning

Michael J. Ford
5M34 W. W. Posvar Hall, University of Pittsburgh, Pittsburgh, PA 15260
Tel: 412-624-6021 Fax: 412-648-7081
Email: mjford@pitt.edu

Abstract: Descriptions of practice are selective, therefore instructional designs based on alternative descriptions of practice likely support learning in different ways. This empirical exploration of the relationship between descriptions of experimental practice and the learning they support utilizes two concurrent sixth grade design experiments. Each design experiment was based on a practice theory of instruction stemming from one of two divergent descriptions of experimentation. One view of experimentation emphasizes the logic of controlling variables, whereas an alternative description locates instructional leverage in precisely those aspects of experimentation necessarily backgrounded by a focus on logic, that is, in the mundane actions of constructing experimental apparatus. Learning outcomes documented through multiple assessments identify ways in which divergent descriptions of practice support learning in different ways. The results suggest that logic may not play the role often assumed in learning to experiment and point to apparatus construction as an activity from which students glean some flexible capabilities that may be educationally fundamental to experimentation.

Introduction

There has been a recent effort in educational research to describe the practices through which professional communities accomplish their work (e.g., Burton, 1999; Hall, 1999; Hall & Stevens, 1995; Roth & McGinn, 1997; Saxe, 1991; Stevens, 2000). Underlying this effort is a rationale that effective instructional designs can be informed by an understanding of how conceptual, symbolic, material and social resources support a professional community's practices. All descriptions of practice highlight some aspects while backgrounding others, and thereby reflect choices regarding aspects of the practice theorized to be key.

What is key to scientific practice is a subject of significant debate among philosophers, especially recently as sociologists have offered alternative descriptions of science that challenge more classic views. These debates, sometimes heated (thereby earning the label "science wars"), have resulted in a cascade of divergent descriptions of scientific practice. Each description purports to represent what is special about science, that is, what aspects of its practice are responsible for explaining the success of science and distinguishing scientific knowledge from other knowledge forms.

This poses a challenge for the science educator who wishes to inform instructional design with descriptions of practice, particularly now that learning "to do" science has been identified as a goal in the National Science Standards (1996). Descriptions of practice function differently for the educator than for the philosopher, however. Whereas the philosopher highlights aspects of practice that are fundamental in that they distinguish science from other endeavors, the educator requires aspects of practice that are fundamental in a different sense. *Educationally fundamental* aspects of scientific practice are those that provide support for a continuing process of learning about the scientific endeavor and how to take part in it. Given the importance of preparing students for lives that are increasingly influenced by science, it is essential that research explore how the aspects of scientific practice highlighted in a description and purported to be fundamental in a philosophical sense also support student learning and function fundamentally in an educational sense. The present study represents a modest step toward addressing this issue.

This study focuses on experimentation. Although experimentation is not the only frame of scientific work, it is popularly considered representative of what distinguishes science from other endeavors. One description of experimental practice highlights the logical structure of experiments, locating the distinguishing features of this practice in the logical method by which experiments are carried out. The logic of controlling variables, or control of variables strategy (CVS) has become emblematic not only of experimentation but more broadly as a structural characterization of cognitive maturity as formal operational thought (Inhelder & Piaget, 1958).

An alternative, divergent description of experimentation put forward by Pickering (1995) locates the source of experimental leverage not in logical method but rather in the process of constructing an apparatus. Pickering argues that to characterize the creation of experiments in terms of any universal method is to fundamentally misrepresent the activity. Moreover, Pickering characterizes apparatus construction not merely in terms of the specialized machinery typically present in laboratories, but more broadly as the objectification of nature in variables, measures and representations that afford a particular focus on situational features for systematic study. The presence of these two divergent views of what matters most in experimentation leaves the science educator in a quandary. Which description should guide the design of science instruction on experimentation? If experimentation is fundamentally about logical method, then this aspect of practice should be highlighted and other aspects of practice, like the mundane actions instantiating a sound design in an apparatus, should be backgrounded. However, if the activity of apparatus construction is what experiments are really about, then the mundane activities of apparatus construction should be highlighted for students, and abstract logic should be backgrounded.

Determining “what experiments are about” in a philosophical sense is not the purpose of this paper. The guiding concern here is to collect information regarding the educational utility of these divergent descriptions. In what follows, I describe two concurrent design experiments, each aligning with one of these visions of experimental practice. The features of these design experiments support several aims. One aim is to identify how these different instructional foci influence what students “take away,” thereby filling a practical need for the science educator. Toward this aim, students from both instructional conditions were assessed in several ways, both on paper and pencil tasks and on an open experimental performance task. By assessing broadly how these divergent instructional experiences position students for subsequent experimental situations, this study also aims to generate hypotheses about learning mechanisms and to suggest what longer-term trajectories of engagement in experimental activities might be most fruitful for learning about experimentation and more broadly about science.

Method

Participants

This study was conducted in two sixth-grade classes in a school that serves grades six through eight. This school is in a suburb of a medium-sized Midwestern city, and at the time of this study about 70 percent of the district’s inhabitants considered themselves "white" whereas about 30 percent considered themselves members of minority groups.

This study was part of a multi-year collaboration between several teachers in this school and researchers from a nearby university. The participating teacher had 22 years of middle school teaching experience and managed science instruction for two classes of students at the time of this study. The school's teaching staff was organized into "houses," or groups of teachers, who shared instructional responsibilities for groups of students. The house of focus in this study was comprised of two teachers who divided among themselves the subjects taught to a group of students. This house of students consisted of 24 boys and 18 girls. This group contained six students who were in an ESL (English as a second language) program, four students who were in an "at risk" program, and five students who received remedial help in reading. This study took place at the beginning of the school year, which allowed for these 42 students to be randomly assigned to their two house classes. In order to encourage the formation of equable groups, the students who participated in the ESL, "at risk," and remedial reading programs were randomly but equally divided between the two classes. All students in this house experienced the instruction described below, but only those students who agreed to data collection participated in the study. In the control of variables (CVS) condition, 21 out of 23 students agreed to participate, and in the APPARATUS condition, 18 out of 21 students agreed to participate.

Procedure

Both instructional conditions were co-taught by the author and the participating teacher. The units were planned to be of approximately the same duration. The CVS condition lasted eight 45-minute periods and the APPARATUS condition lasted ten periods. All students were assessed with paper-and-pencil instruments before instruction and both paper-and-pencil instruments and a performance task after instruction. Descriptions of these assessments follow descriptions of the two instructional conditions below.

Control of Variables (CVS) Condition

For this study, I chose a particularly elegant instructional design to represent the view of experimental practice as logical method. This design, developed and extensively documented by Klahr and others (Chen & Klahr, 1999; Toth, Klahr, and Chen, 2000; Klahr, Chen, and Toth, 2001), has been extensively demonstrated to target the control of variables strategy. Klahr, et al. (2001) identified the instrumental characteristic of CVS instruction in terms of how it effectively isolates the abstract logic of CVS from potentially distracting elements of context.

This instructional design isolated the general, logical form of the control of variables strategy both explicitly through direct instruction and implicitly through structured practice setting up experiments on several experimental contraptions, the designs of which were instrumental in effectively isolating CVS. Elements of context theorized to distract from acquisition of abstract logic were backgrounded by the straightforward operationalization of variables and outcomes in these contraptions. To further minimize potential distractions, instruction identified for students the tactile actions necessary to manipulate all variable settings. Students worked in groups to set up experiments on these contraptions in four domains: ramps, sinking objects, spring extension, and pendulums. Table 1 summarizes the variables and outcomes in these CVS instructional domains.

Table 1. CVS Instructional Domain Outcomes and Variables

	Domain			
	Springs	Ramps	Sinking	Pendulum
Outcome	Extension	Rolling speed	Dropping speed	Period
Binary Variables	Weight size	Ramp length	Object shape	String length
	Coil diameter	Ramp height	Object size	Object weight
	Coil length	Ramp surface	Object material	Object color
	Wire diameter	Ball selection	Drop height	Release height

The students' work was organized in phases. First, students worked in groups to explore the contraptions, and were prompted to set up a valid comparison to target a particular variable's effect. After this initial exploration, the teacher managed a whole-class discussion, during which CVS was explained in general form and in terms of its essential function distinguishing between good and bad (i.e., confounded) tests. The teacher also illustrated this principle through several comparisons on the ramp contraption. Following this instructional phase, student groups returned to their work designing experiments on the contraptions. They worked in a "round robin" manner, completing four experiments on each contraption each day. Students organized their work individually on experimental record sheets. These sheets were provided and represented all the variables in each of the contraptions in a table format. Students were required to circle the settings for each variable instantiated on the contraption and to note the outcome, that is, whether or not the targeted variable made a difference. A more thorough description of the experimental contraptions can be found in Chen & Klahr (1999), and a more thorough description of the experimental record sheets can be found in Toth, Klahr & Chen (2000). After students had completed all of their experiments, the teacher managed a closing discussion about CVS in general, and asked each group to demonstrate two good experiments on each of the contraptions in front of the class.

APPARATUS Condition

Experimental practice in actual contexts includes aspects other than the logic of controlling variables. These aspects include the mundane actions of instantiating a logically sound experimental design by arranging material resources, conducting measurements, and creating symbolic representations of conditions and results. These aspects of apparatus construction each involve distinctive challenges. Whereas the CVS design effectively “black-boxed” these challenges in order to isolate abstract logic as instructional target, the APPARATUS condition was designed so that students would encounter and address these challenges as they emerged. Instruction supported this process by soliciting potential solutions from students, representing scientific criteria for evaluating alternative solutions, and supporting effective student choices.

The unit invited students to conduct ramp experiments to address a somewhat ambiguous question, “How does steepness affect speed?” The salience of the question was established during a race in which one student ran across a hill and another ran the same distance down it. Students objected that the race wasn’t fair, and the class moved to the classroom to address this issue in more depth. Instruction did not provide pre-designed materials to focus on this issue, but rather provided unassembled material (racquetballs, wooden planks of varying lengths and boxes) and invited students to design ramp experiments collaboratively. On the first day, students were prompted to generate a ramp setup that could address the question, “How does steepness affect speed?” and a sequence of actions by which information could be collected. Students individually recorded candidate ramp experiments in journals and then discussed this issue as a class. During this discussion, students generated several ways of operationalizing both “steepness” and “speed.” Students noted that steepness could be varied systematically in two ways, either by holding the plank length constant and changing the number of height support boxes, or by holding the number of boxes constant and varying the plank length. Two options for operationalizing speed were generated as well. “Speed” was conceived in terms of the time it takes for the ball to roll down the ramp, and alternatively in terms of how far it rolls after leaving the ramp. The teacher noted that considering the speed *on* the ramp was analogous to a runner’s speed *on* a hill, so she argued that speed should be measured as the time of the ball roll from top to bottom. However, the teacher allowed the students to pursue both ways of operationalizing “steepness.” She suggested that the class run two experiments, one varying plank length while holding the ramp height constant, and the other by varying ramp height while holding the plank length constant.

Students conducted the former experiment first. According to the initial plan, students were assigned to one of four groups, the task of each was to assign a “roll time” to one of four defined ramp conditions. Results from the four ramp conditions were to be pooled and represented before the class, and subsequent discussion was to relate the data to the question, “How does steepness affect speed?” Dividing ramp conditions among groups was instrumental in highlighting the importance of standardizing material arrangements and measurement protocols. Although the initial apparatus plan prescribed that each group use four boxes and one of four planks (2 ft., 3 ft., 4 ft., 6 ft.), several features of material arrangement remained ambiguous (e.g., positioning of the plank on the box) as did the measurement protocol. Although students initially considered the measurement task unproblematic, when they made their ramps and began collecting information, they quickly realized that these other details required addressing. Most groups noted that other groups used a different procedure for measuring ball rolls, and all groups found it difficult to assign a “roll time” to their ramp because successive trials did not yield identical results. These issues were brought before the whole class as unanticipated problems that needed to be solved before the experiment could go forward. The teacher organized a discussion to generate a standard ramp setup and a standard measurement protocol, thereby specifying what counts as a “ball roll” and the means by which numbers would be assigned to “ball roll” features. Significant discussion took place about the benefits and drawbacks of different arrangements for releasing the ball, starting the timer, and coordinating the stopping of the timer with the ball’s departure from the ramp. Students also debated various ways to determine the representative ball roll time in the presence of multiple divergent measurements. They ultimately decided that all group members would take four careful measurements and each group’s data should be summarized with a mean.

Following these refinements to the apparatus, students returned to their groups and conducted measurements on their ramps. Each group recorded their results as a list of “ball roll” times, and results were organized by the teacher into a single table. This table was a focus of discussion, during which students decided to average the “ball roll” times and debated what the data implied for their question. Students followed the same procedure for the second experiment, except that systematic variation of “steepness” was achieved not through different ramp lengths, but through a different number of height supports with the same plank lengths. In the second experiment, students scrupulously adhered to the protocols they had established for ball rolls, and the experiment produced data that reflected a different answer to the question, stemming from the different way of operationalizing “steepness” in the apparatus. Students generally seemed confused about these divergent results, but due to time constraints the unit ended without resolution of this issue.

Assessments

All students completed three assessments: a CVS pre-test, a CVS post-test, and an experimental performance assessment. The CVS pre-test was given the day before both instructional conditions began, and the CVS post-test was given the day after each instructional condition ended. Immediately following the CVS post-test, the performance assessment was given each day to two pairs of students from each instructional condition until all participating students had completed it. CVS pre-tests and post-tests were paper and pencil instruments that presented a sequence of experimental comparisons. Each comparison contained three variables with two possible settings, and each condition was presented both in pictures and in words. Students were instructed to evaluate the comparison presented and to determine whether it was a "good test" (unconfounded) or a "bad test" (either confounded or target feature not varied) by circling those words on the page. The CVS pre-test was borrowed from Chen & Klahr (1999), and contained twelve items, each in the domain of airplanes. Each item represented three features (tail size, body shape, and wing length) and their settings on two airplanes. The CVS post-test was borrowed from Toth, Klahr, and Chen (2000), adhered to the same format, and contained fifteen items, three items in each of five domains. These domains involved plant growth, cookie baking, airplanes, drink sales, and running speed.

The performance assessment was an open task in which student pairs were to design and execute an experiment to answer the question, "If you drop a ball, how does the drop affect the bounce?" The author accompanied the students to a conference room, which contained a white board and dry-erase markers, a table and chairs. On the table were a can of two racquetballs and a can of three tennis balls. Students were told that they could use any materials present in the room, and if they required some additional materials, they could return to their classroom to retrieve them. The question was written on the board, and students were prompted to both plan and do an experiment to answer the question. Students worked together typically for about 40 minutes. When students determined an answer, they announced this to the author and reported their answer, thereby concluding the task.

Data Forms and Analysis

Each item on the CVS pre-tests and post-tests was scored as "correct" or "incorrect," and each student score indicated the number of correct responses. Performance on the CVS pre-test provided a covariate for testing instructional effects on post-test performance. The experimental performance assessments were video recorded. Student performances were coded for aspects of experimental practice that transferred from instructional conditions.

Results

CVS Assessments

Using the CVS pre-test as a covariate, the CVS students outperformed the APPARATUS students on the CVS post-test (ANCOVA $F(1,37) = 5.617$, $p=0.023$). Whereas the CVS group's mean performance increased from 69% on the pre-test to 84% on the post-test, mean performance of the APPARATUS group remained about the same, decreasing from 71% on the pre-test to 70% on the post-test. These results are in line with results reported in other studies on the effects of CVS instruction, therefore this represents a successful replication. In what follows, this study builds on this replication by exploring how experimental performance was influenced more broadly by what students "took away" from CVS instruction.

Experimental Performance Assessment

A coding scheme was developed to identify specific dimensions of difference between CVS and APPARATUS student performances in the "ball drop" tasks. This coding scheme is organized according to five variables. Three of these variables are binary (yes or no) and represent the presence or absence of a particular feature in student performance. To be coded as present, it was sufficient for a feature to exhibit a single presence during the entire task. The other three variables take the form of ordinal categories, ranking performances by levels of sophistication. Table 2 summarizes the coding results on both the binary and the ordinal variables. A second coder not involved with the collection of data or the conduct of the study coded 50% of the performances randomly selected from each condition. Inter-rater reliability was found to be 100%.

These five features of experimental performance afford the generation and evaluation of experimental results in precise, systematic ways, and thus align with concerns that are central to effective experimentation.

The binary variables address whether students quantified an outcome even once during their performance and whether students established a standard release height for balls across compared trials. Whereas only 20% (2 of 10) CVS student pairs quantified an outcome, 87.5% (7 of 8) APPARATUS student pairs did so. APPARATUS students also were more likely to use a standard release height across trials for comparison, although the contrast between groups was not as strong on this variable. Three other dimensions of student experimental performance were coded as scores on ordinal category scales. These variables characterize performances in terms of the sophistication with which students used multiple trials, represented the results of trials, and summarized their results.

Performances were analyzed in terms of how students used multiple trials in their data collection and were coded in a "multiple trials" score. All student pairs utilized multiple ball drops, and performance was assigned a score of 1 if students never explicitly mentioned that the aim of these multiple drops was to verify a result. A score of 2 was assigned if students mentioned verification of a result, but did not hold the ball's release height constant across subsequent trials, which is a necessary condition for effective verification. Performances were assigned a score of 3 if students ran at least one verification trial while keeping the release height constant across trials being compared. A score of 4 was assigned to performances in which students made an explicit *plan* to run multiple trials and followed through with this plan. Whereas 9 of 10 CVS performances ran multiple trials without any explicit aim to verify a result, 6 of 8 APPARATUS performances included a plan to run multiple trials within conditions.

Table 2. Percentages of "ball drop" performances coded within each category

		CVS(n=10)				APP(n=8)			
		No	Yes			No	Yes		p-value
Quantify Outcome		80	20			12.5	87.5		< 0.01*
Std. Release Height		60	40			12.5	87.5		0.05*
	1	2	3	4	1	2	3	4	
Multiple Trials	90	10	0	0	0	0	25	75	<0.0001**
Represent Results	80	10	10	-	0	37.5	63	-	<0.001**
Summarize Results	90	0	10	0	25	25	25	25	<0.01**

* Fisher Exact Test
** Two-Sample Wilcoxon

Performances also were coded for a "representing results" score, which ranked performances in terms of whether and how students visually represented trial results. A score of 1 was assigned to performances in which students made no visual representations of trial results and a score of 2 was assigned to performances in which students represented a trial result with a gesture (for example, a hand motion indicating the bounce height). Performances in which students created visual marks (e.g., a line segment on the board to represent bounce height or records of quantitative measurements) to represent trial results received a score of 3. Whereas 8 of 10 CVS students did not represent information gleaned from trials in any explicit way, all APPARATUS students did so, with 5 of 8 utilizing more enduring marks rather than temporal gestures.

Finally, performances were assigned a "summarizing results" score, which rated performances in terms of how trial results were summarized. Performances in which a conclusion was reached without referring to or summarizing the trial results were assigned a score of 1. An example of this would be if students observed several ball bounce trials and concluded with a statement like, "The racquet ball bounces more." A score of 2 was assigned to performances in which students summarized results with a decreased level of precision. An example of this category would be if students measured multiple trials of a particular experimental condition, then declared that those results were "more" or "less" than the other experimental condition, without describing "how much" more or less. A score of 3 was assigned to performances in which results were summarized at the same level of precision as the trial results. An example of this category would be if students measured 14 inches and 15 inches as trial results for one condition and summarized these with a statement like, “Let’s just say 14.” Performances in which students employed a systematic, repeatable summary procedure for their results received a score of 4. Whereas CVS students overwhelmingly generated conclusions without explicitly summarizing their experimental results, APPARATUS performances were evenly distributed across these levels of sophistication.

Discussion

The results indicate that CVS instruction generated a significant effect on the CVS post-test. Nevertheless, CVS students did not approach the open “ball drop” task with the sophistication that marked the performances of APPARATUS students. On the face of it, this seems to suggest that the role of CVS logic in learning to experiment may not be instrumental as is widely assumed. However, it is far from clear whether CVS students actually acquired the targeted logical principle, because the isomorphism between the CVS instructional contraptions and the CVS assessment items suggests that increased performance on the post-test may be due to a training effect. CVS students demonstrated that when particular features are provided, that is, defined variables with settings easily determined as either “the same” or “different,” they can evaluate those settings in light of a valued pattern, that is, the target variable settings as “different” and all other variables as “the same.” The “ball drop” task, which did not include these cueing features, did not provide CVS students an opportunity to demonstrate what they had learned.

The APPARATUS students’ relatively sophisticated approach to the “ball drop” task, however, represents a contrast to this dependence on contextual cues. Even though the task was open and ill-defined, APPARATUS students targeted quantitative information to address the “ball drop” problem, and their work generally resulted in careful scrutiny and quantification of targeted ball drop features. The experimental structure that yielded a solution to this problem was neither embedded within the task nor imposed from an acquired abstract logical template, but rather seemed to emerge in student work. The distinguishing features of APPARATUS performances yielded by this initial analysis (namely, quantifying an outcome, standardizing release height, running multiple trials, recording results for reflection, and summarizing results systematically) represented solutions to problems encountered during APPARATUS instruction. These solutions shaped both the end products of those experiments and, also likely shaped students’ conceptions of what it means to conduct an experiment. This conception of experiment likely guided APPARATUS students to grapple with issues similar to those addressed in the APPARATUS instruction. This similarity of issues is paralleled by a similarity in the way the issues emerged and were addressed. Both APPARATUS instruction and “ball drop” performances seemed to reflect an *iterative* process of instantiating an apparatus feature, encountering material resistances, and revising the apparatus to overcome those resistances. For example, in the “ball drop” task, individual ball drops supplanted simultaneous drops as a solution to logistical problems encountered measuring bounces simultaneously. In turn, release heights were standardized to allow comparison of individual ball drops. As during APPARATUS instruction, these performance features addressed material resistances, which emerged in response to students’ efforts to scrupulously characterize features of the events under scrutiny.

This interplay between goal-directed instantiations and material resistances is perhaps most significant in APPARATUS students’ process of defining variables. In general, variables play a key function of bridging the real world with the represented world. Meanings of words like “bounce height” in an everyday sense are determined by everyday semantics, but in experiments, the meanings of such terms become limited exclusively in terms of the actions through which a quantitative descriptor is assigned to a feature. For example, whereas in an everyday sense, the word “height” means roughly some aspect of elevation, “height” came to mean something more specific for APPARATUS students in this task, such as, for example, the number on a meterstick observed to most closely align with a momentary apex of the ball’s path. Of course, such an observation must be framed by a particular material orchestration to make it meaningful. A meterstick must be held vertically, with one end placed on the bounce surface and the numbers increasing in value as an increase in distance from that surface. Even the “bounce event” itself must be scrupulously structured. The ball must be released in a particular way, from a particular distance above the bounce surface, and the drop must be near enough the meter stick to allow for observation. Such material orchestrations and structured observations were not established at the outset of student work, but rather emerged in an iterative manner as initial candidates for “bounce height” were first instantiated and then evaluated. The ways in which these initial candidates for “bounce height” were subsequently found lacking (e.g., two measurements of the same bounce were found to vary widely) became, for APPARATUS students, indicators of confounds, and subsequent revision of the apparatus targeted the removal of these confounds. Hence, iterative action and revision was key to the definition of variables, through a *creation* of confounds, as material orchestrations and structured observations met with resistance, and the *removal* of these confounds through scrupulous revision of these variable definitions.

For Pickering (1995), the resistances that scientists encounter as they construct their apparatus plays a central role in the development of an experiment. He refers to such resistances as "material agency," because it represents a constraining factor on the apparatus development, unpredictable at the outset and irreducible to the human realm, that provides scientists with the information necessary to ultimately make the apparatus work. In this paper, I have briefly described how variable definitions made material agency visible (at least implicitly) to students and guided the iterative redefinition of variables. The contrast of this account of learning experimentation with the popular account as imposition of an *a priori* logical template is clear. Further analyses of this data set will develop this account in more detail by tracing the role of material agency to include the circumstances under which material agency was encountered, whether those circumstances were shaped by the transfer of particular actions from APPARATUS instruction, and whether such acts reflect not explicit reactions to, but implicit anticipations of, material agency in the "ball drop" task.

Another focus of learning leverage suggested by this study is the students' own sense of agency they gained from the APPARATUS instruction. For example, in order to target apparatus construction as an emergent, iterative process, APPARATUS instruction necessarily focused on not only procedural and representational, but also social, tactile, and affective dimensions of experience. Students were led to encounter difficulties and provided at least some authority to make decisions about solutions. APPARATUS students were responsible for evaluating the options available at each stage of their apparatus development, and they experienced outcomes directly as they acted upon their choices to discover ultimately how they would affect their apparatus' success. Given the way APPARATUS student performances reflected ability and confidence to fruitfully engage an ill-defined task, this affective aspect of APPARATUS instruction seems an important subject for future study.

Finally, these results may indicate an instrumentality of students' understanding something about the "whole" of a practice rather than merely its isolated parts. Taken as a whole, the aspects of APPARATUS performances suggest some broader understanding of experimentation as part of a modeling endeavor (Lehrer, Schauble, & Petrosino, 2000). To what extent can an understanding of these aspects of experimental practice be separated from knowledge of the quantitative modeling enterprise as a whole? Such a broad understanding may provide a crucial coherence through which students come to understand how agency, tools, argumentation patterns, and norms of science combine in the modeling process. Indeed, the ways in which understanding particular practices is dependent upon their place in the larger whole may remain unexplored with an exclusively decompositional approach employed by classic experimental methods. In this way, "practice" holds promise as a fruitful unit of analysis not only for philosophical speculation but also for future research on learning and instructional theory.

References

Burton, L. (1999). The practices of mathematicians: What do they tell us about coming to know mathematics? *Educational Studies in Mathematics, 37*, 121-143.

Chen, Z., & Klahr, D. (1999). All other things being equal: Acquisition and transfer of the control of variables strategy. *Child Development, 70*(5), 1098-1120.

Hall, R. (1999). Following mathematical practices in design-oriented work. In C. Hoyles, C. Morgan & G. Woodhouse (Eds.), *Rethinking the mathematics curriculum* (pp. 27-49). Philadelphia, PA: Falmer Press.

Hall, R., & Stevens, R. (1995). Making space: A comparison of mathematical work in school and professional design practices. In S. L. Star (Ed.), *The cultures of computing* (pp. 118-145). London: Basil Blackwell.

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

Klahr, D., Chen, Z., & Toth, E. E. (2001). Cognitve development and science education: Ships that pass in the night or beacons of mutual illumination? In S. Carver & D. Klahr (Eds.), *Cognition and instruction: Twenty-five years of progress* (pp. 75-119). Mahwah, NJ: Erlbaum.

Lehrer, R., Schauble, L., & Petrosino, A. (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*. Mahwah, N.J.: Lawrence Erlbaum Associates.

Pickering, A. (1995). *The mangle of practice: Time, agency, and science*. Chicago: University of Chicago Press.

Roth, W. M., & McGinn, M. K. (1997). Science in schools and everywhere else: What science educators should know about science and technology studies. *Studies in Science Education, 29*, 1-44.

Saxe, G. B. (1991). *Culture and cognitive development: Studies in mathematical understanding*. Hillsdale, NJ: Erlbaum.

Stevens, R. (2000). Divisions of labor in school and in the workplace: Comparing computer and paper-supported activities across settings. *The Journal of the Learning Sciences, 9*(4), 373-401.

Toth, E. E., Klahr, D., & Chen, Z. (2000). Bridging research and practice: A research-based classroom intervention for teaching experimentation skills to elementary school children. *Cognition and Instruction, 18*(4), 423-459.