MATHEMATICAL THINKING AND LEARNING, *2*(1&2), 51–74

Inventing Data Structures for Representational Purposes: Elementary Grade Students' Classification Models

Richard Lehrer and Leona Schauble
Department of Educational Psychology
University of Wisconsin–Madison

This study concerns the development of children's understanding of data and classification. Three intact classrooms of students and their teachers (18 at 1st and 2nd grade, 25 at 4th grade, and 22 at 5th grade) worked over several sessions to categorize drawings made by other children and to progressively "mathematize" their categorization rules. The goal was to develop a model that would make explicit their initial guesses about the grade level of the artists. In small groups, students developed, applied, and made iterations of revisions to their data models. The details of instruction and duration varied somewhat in accord with the capabilities and needs of students of different ages. The youngest children evolved systems of attributes that described their categories in a post hoc fashion but failed to come to regard those rules as a model to guide classification. In contrast, 4th and 5th graders considered their category systems as models that logically constrained the members admitted to categories. These students came to appreciate dimensional attribute–value structure, although many continued to include redundant or extraneous information. They incorporated and discussed a variety of kinds of decision rules, including ways of combining information, such as differentially weighing diagnostic attributes. By engaging with data characterized by prototypic rather than crisp membership values, students had the opportunity to see the intellectual work performed by practices of data modeling.

Classification of objects and events is a ubiquitous component of cognition, evident in early infancy yet changing in nature throughout childhood (e.g., Inhelder &

Requests for reprints should be sent to Richard Lehrer, Department of Educational Psychology, University of Wisconsin, 1025 West Johnson Street, Madison, WI 53706. E-mail: rlehrer@facstaff.wisc.edu

Piaget, 1964; L. B. Smith, 1989). Consequently, the study of categories and concepts has long occupied center stage, although there remains considerable controversy about the essential qualities of categories and the potential mechanisms governing their formation (e.g., Anderson, 1990; Fodor, 1995; McClelland & Rumelhart, 1986; E. E. Smith & Medin, 1981). In addition to its centrality as a psychological phenomenon, classification is essential to reasoning in many disciplines from evolutionary biology to robotics. Therefore, there is a correspondingly long tradition of developing mathematical models to render classification transparent to practitioners in these disciplines (e.g., Panchen, 1992; Sneath & Sokal, 1973).

Our study intersects these two trends, presenting our investigation of how children's perceptions of resemblance can serve as the basis for developing a mathematics of classification. As we explain, we attempted to examine children's progressive "mathematization" of their perceptions of similarities and differences among sets of drawings. Children's propensities to categorize served as a foundation for reasoning about classification. We asked children to develop a mathematical model that would serve to make explicit and to justify their initial guesses about the likely grade levels of the artists of those drawings.

Because the problem posed to students dealt with fuzzy rather than crisp sets, common school practices that equate data classification with tables were insufficient for problem solution. In this way, the problem posed went well beyond attribute blocks and other typical school activities related to pattern and classification.

Unlike most studies of the development of reasoning, ours was situated within classrooms. Although classrooms typically do not afford high degrees of experimental control and related aids to strong inference, they do provide opportunities to investigate the development of mathematical reasoning when this reasoning is bootstrapped by teachers and peers. Because it is unlikely that children will spontaneously develop a mathematics of classification, we collaborated with teachers to develop tasks and situations that would foster the growth of mathematical reasoning about data and classification. Previous research suggested that reasoning about data is fostered when students are asked to invent and revise models (Hancock, Kaput, & Goldsmith, 1992; Lehrer & Romberg, 1996; Lehrer & Schauble, in press). Data modeling is a multicomponential process of posing questions; developing attributes of phenomena; measuring and structuring these attributes; and then composing, refining, and displaying models of their relations. In this study, we do not break these components apart to investigate them separately. Instead, we examine their orchestration in intact classrooms.

We worked with teachers at three different grade levels (first and second, fourth, and fifth) to develop tasks and related forms of instruction that afforded students opportunities to develop, critique, and revise data classification models. The tasks were designed so that students could readily classify stimuli perceptually ("with your eyes"). The challenge, then, was to develop a mathematical model to account for what they so readily perceived. Instructional means were devised in collaboration with teachers so that students would have need to both invent and revise their models, although the implementation varied with teachers' preferences and preexisting classroom cultures. Our purpose was both to study the change over time that occurred within age groups and to compare the amount and rate of learning accomplished by children across different grades. This work, then, is in the form of a teaching experiment. Because the teaching is adapted to the needs and capabilities of students in each of the participating classes, the results are necessarily understood as outcomes of procedures that differ somewhat from group to group.

METHOD

Participants

Participants in the research were one intact class at each of three grade levels: a combined first- and second-grade class, a fourth-grade class, and a fifth-grade class. The average age of the 18 first and second graders was 7 years 1 month (range = 6 years 2 months–8 years 2 months). The 25 fourth graders averaged 9 years 8 months (range = 8 years 7 months–10 years 2 months). The average age of the 22 fifth graders was 10 years 7 months (range = 9 years 7 months–11 years 8 months). Across the three grade levels, the sample included 30 boys and 35 girls.

These students and their teachers were participants in a district-wide research reform project targeted toward improving elementary mathematics and science through an emphasis on modeling (Lehrer & Schauble, in press). This suburban/rural district is approximately 15 miles (24 km) from the state capital in a midwestern state. Adults in the district represent a mix of professional, blue-collar, and farming occupations. Socioeconomic status is generally middle class; approximately 5% of the students are eligible for free or reduced lunch.

Materials

In preparation for the unit on data classification, teachers asked all students in the school to contribute drawings that would be used as the stimuli. In their home classrooms, students from all grades drew two pictures on different occasions. One was a self-portrait; the other was the student's attempt to portray "two houses, one near and one far" (see Braine, Schauble, Kugelmass, & Winter, 1993). From this set of approximately 400 drawings of each kind (portraits and houses), the three participating teachers selected sets of drawings that they considered "representative" of the drawings of kindergartners, first graders, third graders, and fifth graders. Twelve portraits and 12 drawings of houses were assembled as "most representative" of each of four grades (kindergarten, first, third, and fifth). Each set was then sorted into comparable subsets with four drawings at each grade level to produce a Set A, Set B, and Set C for portraits and a Set A, Set B, and Set C for houses.

Multiple copies of the sets were made for students to examine. On these copies, the age of the artist and other identifying information were omitted. Each drawing was assigned a fabricated name (e.g., "Wanda") so that students could easily refer to the drawings; students were informed that these names were fabrications.

The study was planned collaboratively by project researchers and participating teachers (standard practice under the reform effort under way in the district). We agreed that, under the guidance of their teachers, the participating students would assume the role of "junior researchers"—investigating a set of drawings carefully, trying to decide the grade of the artist for each, and eventually developing and revising a classification model that would accurately identify the correct grade of the artists of novel sets of pictures.

Self-portraits and near–far houses were assigned as the topics for the pictures because younger and older children draw these items in characteristically different ways. Some of these differences are obvious on casual comparison, but some are not likely to be noticed without careful study (e.g., Braine et al., 1993; Gardner, 1980). Hence, we considered it likely that, as children continued to inspect the drawings, they would identify additional diagnostic features and relations among features. We conjectured that this growing body of classroom lore about the drawings would, in turn, provoke model revision.

There are reasonably robust age differences in the way that children draw these items (e.g., as Braine et al., 1993, established, younger children tend to portray two houses near and far by placing two same-sized houses on a horizontal line, whereas older children use a diagonal placement and differing sizes to indicate proximity). However, there is also considerable variability within any grade level. Therefore, we conjectured that students would need to find ways to combine information in their data classification models and to make decisions about how to cope with mismatch between the expected outcomes predicted by the model and the observed values (i.e., the actual grade levels of the artists). The form of these data also afforded opportunities for students to consider how to handle sets of attributes that were not disjoint (e.g., an attribute could be present at more than one grade level) and to develop and evaluate appropriate data structures (e.g., lists and tables).

The participating teachers conducted this investigation in their classrooms in October of the school year. There was some variation in starting times and in duration: The fourth-grade teacher began about 2 weeks after the first- and second-grade and fifth-grade teachers; the first and second grade worked on this unit for 1 week, whereas the fourth and fifth grades worked for 3 weeks. Differences in duration were due to each teacher exercising judgment about available classroom time and how far the students could profitably proceed with these ideas.

RESULTS

The procedure varied somewhat from grade to grade, as teachers monitored and assessed their students' capabilities with this task. Therefore, we report results for each grade level in turn, integrating the findings with a description of what occurred in each of the classrooms.

First and Second Grade: From Global Evaluation to Attributes

The first- and second-grade teacher, Eric James, worked with the data classification task for approximately 1 week, or until he judged that his young students had accomplished about what could be expected from the exercise. These children worked only with the self-portraits and did not progress to work with the near–far houses. As we explain next, over this period of instruction most of these young children never really grasped the idea of a model—especially the separation of a model from its referent. Instead, many of them appeared to consider the activities of assigning portraits to appropriate grade levels and developing clear descriptors of portraits within grades as two unrelated tasks. We observed significant improvements in the specificity and clarity of children's attribute descriptors over the course of the week, but for the most part, children assigned portraits to grades by exercising holistic perceptual judgments, not by systematically applying attributes.

Initial sorting. The teacher framed the task by instructing each group of 2 or 3 students to "take a set of 12 pictures, spread them all apart, and then gather them together into four piles in terms of which grade you think did these pictures." Because many of the children could not yet read or write, Eric asked a collaborating fifth grader to work with each group of students and to serve as group recorder.

The children began sorting the pictures as if the task were self-evident. As one of the fifth-grade recorders noted, "They're really confident that they're right." Moreover, initially, students failed to work together on the task. Instead, for the most part, each child in the group picked up a picture, decided on the grade level of its artist, and placed it into a category. There were few or no attempts to reconcile any differences in opinion. When pressed by an adult or a fifth grader to justify a choice, the students tended to do so with globally evaluative statements: "This definitely would be from kindergarten. I know, because it's dumb." "This one goes to third grade, because it looks better than most of them but not as good as the fifth grade."

However, as the children were encouraged with consistent prompting from the teacher, researchers, and fifth-grade partners, their discussion began to shift from evaluative statements to mention of features: "Tim is kindergarten, because he's bare naked and sort of simple. He has no hair, no hands, no shoes." "Wanda has no hands, no shoes, no stomach. Round feet." However, even as children's conversation began to shift from evaluative generalities to descriptors, they focused for the most part only on single features of individual pictures. On the 1st day, they made no progress toward defining dimensions that crossed grades or portraits. Yet, over the course of this day, the children began to examine features at a finer level of detail. For example, at one point, Lane criticized a portrait on the grounds that the hands were not of equal size; one was "smaller and fatter" than the other.

Students appeared to be more concerned with "being right" about the grade-level classifications than with finding clear and distinguishing descriptors. Angeline's concerns were typical: "I wonder if these are right. I think the kindergarten group is definitely right. I'm definitely sure about the fifth grade." Moreover, having sorted the pictures into piles by grade level, students were reluctant to reconsider or revise their first groupings upon reflection or the introduction of new cases.

The most general thesis advanced and agreed to by the students was that older children make portraits that include "more detail" than those of the younger children. Unfortunately, there was no specification of what "detail" meant. Those students who did make progress in generating a category system relied heavily on adults for assistance. The teacher and the two participating researchers offered assistance by proposing counterexamples, asking about the consequences of a rule, and encouraging students to compare pictures across grade levels. In contrast to the students in the older grades, who are described later, the first and second graders did not initiate these activities.

At the close of the 1st day, each fifth-grade helper prepared a list summarizing the current model of his or her group. As Figure 1 demonstrates, for the most part these models were paragraphs of text arranged in whatever way the fifth graders found convenient.

Discussion of first models. On the 2nd day of instruction, Eric explained that he hoped to "give some feedback about which of the ideas seem really clear." To that end, he orchestrated a whole-class discussion in which students concentrated on each of the models in turn. This discussion seemed critical in turning the children's attention primarily from whether they were "right" about grade membership of the pictures toward whether the features in their models were clear and explicit. For example, in their model, Rachel and Jordan included "potato body" as a descriptor for the kindergarten pictures. In the class discussion, Eric asked them what "potato body" means, and Rachel explained, "It means there's no feet and no body, and his arms are in his head."

During this discussion, the children questioned for the first time whether a proposed feature was accurate and precise. For example, some students questioned

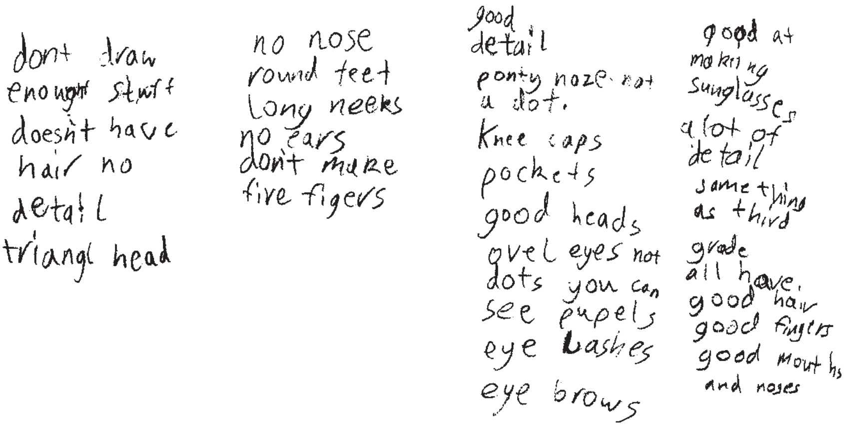

FIGURE 1 First model developed by one first-grade group.

whether it was really true that the kindergarten pictures had "no feet" and noted that to decide, one would have to determine what counts as feet. One child proposed that the authors of the model might have "meant no shoes." To this suggestion, Eric responded, "Would that make their thinking clearer?"

During this discussion, students were also observed mentioning *multiple features* of individual pictures. For example, as one child remarked about a picture, "It has no fingers. It has no feet. You can barely see the mouth. And his hair looks like a hat." Later in the discussion, Whitney's comment shows that the children were beginning to compare a common feature across portraits within a grade: "On all three of the kindergartners, it goes body, body, body. No body, then that tiny girl. So they go like, kind of in a row." In a couple of instances, children even compared a feature across grades. As Katie said, "This one was done by a first grader because she has dot eyes and the other ones [those from the other grades] are oval." Yet, even with these signs of progress, holistic judgments remained more compelling to many of the children, who found discussion of descriptors less useful than claims such as "I think this one looks like a kindergartner or a first grader" or "I know what grades they were in, all."

On the basis of this discussion, Eric suggested that the children might want to revise their groupings. Although most of the students agreed that this would be a good idea, they did not get the opportunity to do so until the following day. As we discuss next, when this opportunity arose, most of them failed to take advantage of it.

Making a group model. On the 3rd day, Eric compiled the lists of the fifth-grade recorders into what he referred to as a "super-smart list." As he wrote

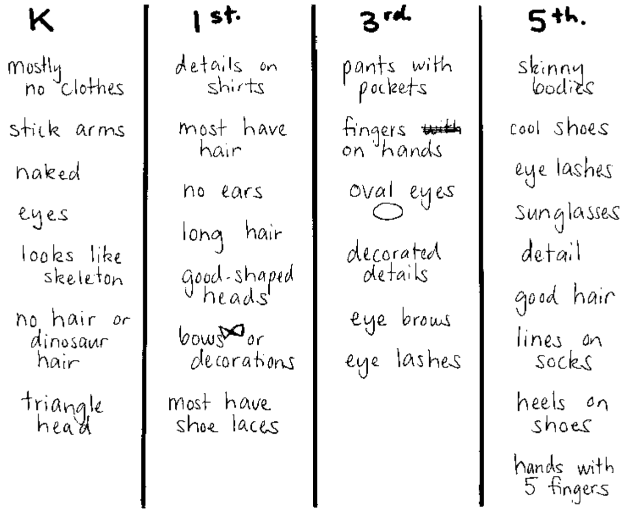

FIGURE 2 Composite model developed by the first-grade class.

this list on the board (see Figure 2), he frequently paused to ask about one descriptor or another, “Is this really clear?” Children readily responded with elaborations. For example, Rachel explained, “Good hair means they have, like, ponytails.” Eric also asked students to visualize what a descriptor might mean: “Looks like a skeleton. Do you know what skeletons look like? That’s a good description. Close your eyes. See if you can picture stick arms, a triangle head, looks like a skeleton.”

When the group list was compiled, Eric suggested that children “might want to change your piles a little bit now that the whole class is thinking together.” One child remarked that one of the pictures designated as “kindergarten” in her group might actually be a third-grade picture. However, when asked to justify the change, she simply responded, “Because it looks like third.” An adult tried to focus her on the descriptors written on the board, but she appeared satisfied with her “looks like” criterion.

Applying the model. On the 4th day, Eric announced that the class would see some brand new drawings, but first he asked students to review the list on the board and decide whether any of the descriptors “go together in your thinking.” The children decided that “no clothes” and “naked” in the kindergarten category go together. “Looks like a skeleton” and “stick arms” were also deemed related.

To focus children on descriptors, Eric attached a blank, rose-colored piece of paper to the board and said:

> See this picture of Rosie? I'm going to tell you some things about Rosie. See if you can figure out which group she's in. Let's pretend I said to you, "This is Rosie, and she has heels on her shoes." Raise your hand to tell me where Rosie would go.

As he continued along this vein (with "Peachie" and "Greenie"), Eric also encouraged students to "keep their eyes open" and report "can't tell" if he used a clue that was not diagnostic. Eventually, the class began to yell, "Tell me more! Tell me more!" in those cases when Eric mentioned a descriptor that appeared under more than a single grade in the list.

As he brought this class discussion to a close, Eric asked children to reform their small groups and use the class "rules" to sort a brand new set of pictures. To focus attention on the "rules," he announced, "When Mr. James comes around, he's gonna say, 'What rules are you using?' I'm going to mark it on your paper with a dot if I see you using it." In general, although the groups began by considering the rules, they abandoned them when there was no clear match between features of a picture and the descriptors on the board. They also showed no concern when a rule seemed to apply across more than one grade level; instead, they were likely to assign a picture to a grade by "eyeballing" it and then searching for descriptors to justify their choices.

Checking the model. On the 5th and final day of instruction, Eric tried to encourage thinking about dimensions by suggesting the cross-grade application of a rule. "Who can tell us a 'teeth' rule? Tell me how it fits one of the grades; that's what we mean by a rule." However, in response to this request, the students suggested "teeth" descriptors for fifth grade and kindergarten but failed to sustain the discussion long enough to carry it across all grades. The focus on teeth was aborted when Whitney proposed a new rule about eyes. At this point, Eric abandoned the attempt to focus on dimension and suggested that the small groups check their sets from the previous day by matching them against the rules.

The children's activity was quite diagnostic of these first and second graders' limitations in this task. In one group, the children conscientiously "checked" the rules by making a check mark next to each attribute on the list at a time, noting whether it was present or absent for each picture considered. However, surprisingly, they appeared not to even be considering whether a given picture's placement violated one of the rules. It was as if reading and checking were totally divorced in their minds from the problem of grade classification. Even though they found a number of mismatches with the rules, they remained quite confident that, for instance, "This is definitely the first grade." In a postclass conversation, Eric remarked that he had observed one group of students "checking" their piles with no recourse to the rules at all. When an adult challenged whether these piles were consistent with the rules, the rule sheet could not initially be found; eventually, it was discovered that one of the children was standing on it.

Summary: First and second grade. By the time they completed the week of instruction, children were able to do a reasonable job of applying one rule at a time and deciding whether a given picture was consistent with it. They continued to have difficulties when they encountered mismatches. Even though Eric had discussed the idea of a “rule-breaker” (e.g., a picture that does not obey all the rules for that grade level), this discussion seemed beyond the comprehension of most of the students who, for the most part, were using the features as mere descriptions and not as rules at all. That is, rather than using rules to categorize the pictures, children seemed to be using them to describe the pictures. Although a few children made progress in understanding the category system, as indicated by the discussions reported earlier, these advances did not appear to “take hold” for the long term, and most of the first and second graders did not achieve them. For most of these students, progress over the course of instruction consisted of changes in the quality of the descriptors themselves—a shift from global, evaluative comments to feature-based descriptions that became more clearly stated, more fine-grained, and more accurate as the activity progressed.

Fourth Grade: From Mutually Exclusive Attributes to Dimensions

Approximately 5 of the students in the fourth grade had completed the body portraits data model task with their teacher, Erin diPerna, in their third-grade year. Therefore, the fourth-grade teacher, Sally Hanner, worked with the near–far houses pictures as stimuli. This work occupied approximately 15 days of class time that extended over a month. Like the first and second graders, the fourth graders also struggled to express their “rules” in language clear enough so that an uninformed participant could apply them. However, they made more progress than the first graders in achieving this objective, partly because they had no difficulty with the idea of applying the features as rules that a given picture must meet to be included in a category, rather than as after-the-fact descriptions of pictures already assigned to a category by other means. Moreover, the fourth graders went on to consider whether it might be worthwhile to apply a common set of descriptors across categories—a dimensional structure that included attributes of different levels.

Initial sorting. The initial procedure that the students agreed on for assigning pictures to grade-level categories was to vote within their teams of 3 students. They further agreed that "majority wins," although there was considerable discussion about how to handle "three-way splits." Some students advocated for selecting the middle value (e.g., assigning a picture to third grade if there were votes for first, third, and fifth), but one student objected that, in this case, pictures would never get assigned to the kindergarten or fifth-grade categories. Sally emphasized that the point was not to form the categories but to "leave a thinking trail" by writing descriptors:

> Not for an individual picture but for each grade level. … Come up with some words and phrases that describe what they're like. What makes the pictures be in that same group? The most important thing in this task is not to identify the right grade level of the drawings but to have clear thinking and communicate your thinking.

Alex, a student, objected to the validity of the enterprise, pointing out, "Well, you never know if a fifth grader draws like a third grader." Another student concurred, suggesting that maybe someone with a learning disability might make a drawing like a kindergartner. However, Tyler pointed out that such a model might be reasonable to aim for even if it could not account for all cases: "Most of the time, but not all of the time, fifth graders have better drawings than the younger kids."

In contrast to the first and second graders, who primarily worked in parallel to categorize the pictures without considering each others' judgments, these older children reviewed each of the pictures together. Agreement was reached quickly about each picture, and unlike the younger children, students spontaneously justified their assignments by mentioning features of pictures ("This one is just like two blobs. They're both kind of sloppy. They're both kind of small."). Some groups were quick to spot conventions used to represent near and far. For example, Justin noted, "Brit was kind of smart because she made the 'far' house kind of smaller [than the 'near' house]; it wasn't really the same size." In fact, by the close of these discussions, some of the groups were already applying criteria across grade categories, although the criteria themselves were ambiguous (e.g., more and less "messy" or "organized").

Initial models. On the 2nd day of instruction, Sally handed out a sheet on which students could record their initial descriptors in a common structure. The sheet was divided into quadrants labeled *K, 1, 3,* and *5,* respectively. As we discuss, this structure may have influenced the kinds of initial models that students produced.

Knowing that students would have difficulty understanding the need for clarity and precision, Sally asked what an outsider to the classroom might notice in all the pictures categorized within a grade. In reply, students' comments ranged from general features meant to apply to all the pictures within a category ("They're like, sloppy") to features that described only one picture within a category ("This one looks like a purse").

When children convened in their small groups to record their descriptors on the standard format sheet, children were observed revising the placement of some pictures as they considered whether they matched or mismatched the descriptors for that category (recall that the first and second graders very rarely recategorized a picture).

As in the first and second grade, the fourth graders also believed that "detail" characterized pictures categorized in the older grades. However, unlike the younger students, they considered it important to say what counted as detail. Children in one group concluded that detail included bushes or trees, windows, doors, roofs, siding, bricks, paths or roads, doorknobs, and people. During this discussion, children were observed arguing about whether a descriptor should be included on the list if it applied to one picture deemed to be in the category but not to another.

On the 3rd day of instruction, Sally introduced the importance of clarity of description: "We're ready to see if another team can take a set of drawings and descriptors and use those descriptors to sort the drawings into the same set of piles." Richard Lehrer, one of the researchers, explained:

> A lot of times we try to do this in factories. There are all kinds of nuts and bolts to be sorted. If you hire people to do this, it's very expensive and it's boring. Industries try to program robots to do this. But the robot is very, very stupid. … You should be able to take these descriptions and think not about everything *you* know but what they really say. Almost as if you were writing a computer program for a robot. Remember, you're much smarter than any robot.

As Lehrer played "robot" and evaluated students' descriptors, the children began to ask questions such as "What do we mean by 'no details'?" and "How do we *know* this drawing took a little longer to do than that one?" Children began to discuss whether one could fairly apply criteria such as "really good drawing." As Leslie summarized the problem:

> We're sort of, like, trying to figure out how we can make our descriptions for our pictures better so a robot, a stupid robot, could use them. We pretty much know this [drawing] is fifth [grade], but we need to find out a way so that *he* [the robot] would know that this is a fifth grader.

However, the usefulness of this stance clearly eluded some fourth graders. Lehrer applied one group's model to a drawing and concluded that according to the model, the drawing should be classified as a kindergarten drawing. Brittany objected, "But I can look at it, and I know it isn't a kindergarten drawing." It remained difficult for students to understand the need for a model to accomplish something that they felt they could do perfectly well by everyday common sense. Moreover, although many students began to understand the difficulties with including descriptors like "detailed" and "advanced" in their models, it was not immediately apparent to them what they should substitute.

Over subsequent days, students continued to work in groups to refine their descriptor lists and finally to compile them as a model on large sheets of poster paper. As they worked on their descriptors, most of the groups simply listed their descriptors under the appropriate "box" on the recording sheet, a strategy that encouraged the consideration of one grade level at a time for purposes of identifying descriptors. As Figure 3 illustrates, most of students' initial models (six of the

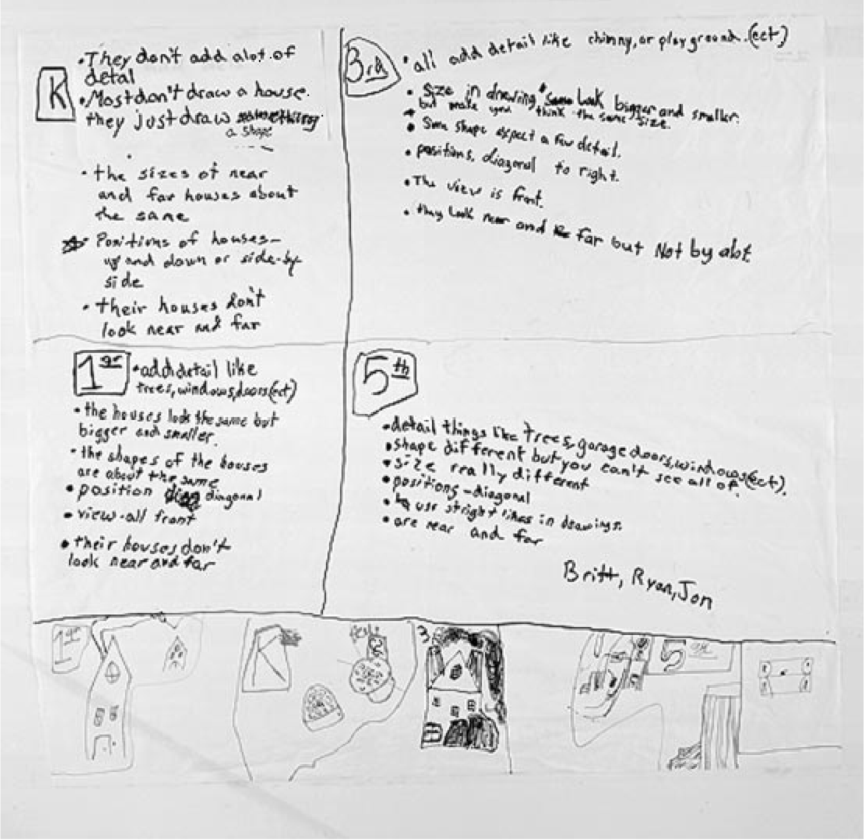

FIGURE 3 "Box model" developed in the fourth grade.

seven groups) maintained the box structure that was originally suggested on the recording sheet.

At this point, many of these students believed that rules should serve to identify exclusive sets. For example, if a descriptor applied to more than one grade level (e.g., “Near house is drawn larger than far house”), these children would reject it altogether. That may be because the box structure of their model discouraged thinking about combining information across features. A related “bug” in the fourth graders’ reasoning about these models concerned their conviction that, to work, a descriptor needed to apply to every case in the set. Hence, finding an exception to a descriptor in the category (e.g., if one third-grade drawing did *not* draw the near house bigger) was considered grounds for rejecting the descriptor. Some children held both these beliefs; some held one or the other. Taken together, these conceptions resulted in models that failed to apply features systematically across grades—the resulting models had a “chopped up,” ad hoc quality.

In contrast, Lehrer asked one group of students to consider how a single attribute could be considered at every grade level. This tabular form of inscription was familiar to children, but they had never considered its potential for classification. Perhaps children’s previous history with distinctive features (e.g., squares vs. circles) and nonoverlapping sets biased them toward the disjoint sets implied by the worksheet. After seeing how one attribute could distinguish among some of the grades, students extended the column to a dimensional model in which grade level was recorded down the side of the paper and common features were identified as general attributes across the top. Levels for each attribute were specified in the intersecting cells (see Figure 4).

Evaluating models. The advantages of this model were not immediately clear to the class. When the children began to share their models with the entire class, Nikki displayed her group’s attribute model (see Figure 4), explaining, “We used position, size, and dimensions.” Ryan objected that Nikki’s model didn’t have nearly as many clues as the other models: “Compared to ours, we have sections and they have sections, but we wrote more than two words [in each section].” Many of the children were apparently operating under the assumption that the more ingredients a model had, the better. Brittany noticed that the format of Nikki’s model was also different from the others: “We wrote in like, sections, and they put it in columns and rows. A row of position, a row of sizes, and a row of dimensions.” When asked, “What’s the advantage to that table? What’s it good for?” students concluded that Nikki’s model was different from theirs primarily in its graph format, but they seemed to consider the format to be a trivial matter. As one child suggested gently, “It doesn’t matter how you write it; it matters what’s in it.”

Shortly thereafter, Sally returned to a discussion of the models. Children compared similarities and differences in their models, crossed out rules that didn’t

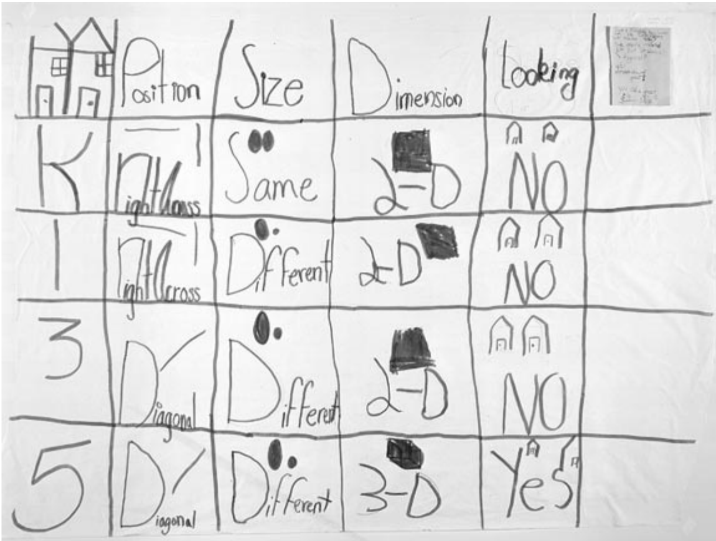

FIGURE 4 Dimensional model developed in the fourth grade.

work very well, and preserved those that did. Then, to help focus the children on clarity, she invited Mark Rohlfing, a fifth-grade teacher, to visit the classroom and apply the children's models to a novel set of pictures. There was some discussion about whether it was fair to evaluate a model by applying it to pictures that were not used to construct it originally. Nathan suggested that the model should work with any set of pictures, regardless of whether those pictures were used to make the model in the first place.

Mark, the visiting teacher, applied each model in turn, thinking aloud as he did so. He remarked:

> I can definitely tell this one [Nikki's attribute model, Figure 4] was quick. This one [Brittany's box model, Figure 3] is hard because I have to read every single thing on there. … On these, I had to read through every word until I found one that matched. Then, I had a hard time remembering how many matches.

He also remarked on the clarity of descriptors: "I'm not sure what 'except a few details' means. What does 'diagonal to the right' mean? I'm not really sure how much 'a lot' is." Mark also modeled how to categorize a picture when the attribute values are not exclusive: "My guess is it [this picture] fits three out of three [attributes] in one grade but two out of three in K, so my guess is it's a first grade. … So, the most matches wins?"

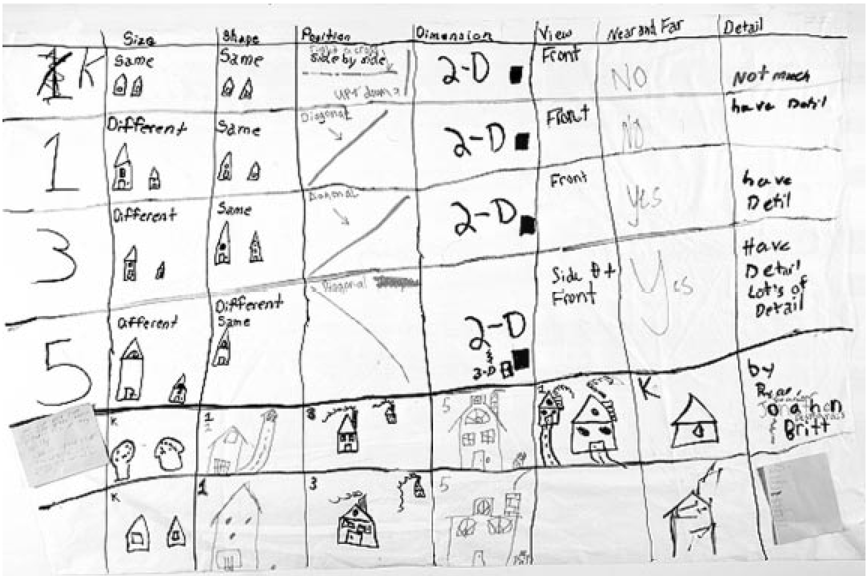

FIGURE 5 Final model of one fourth-grade group.

After each of the models was evaluated, the children concluded that Nikki's attribute model was the fastest to apply, and as Nathan noted, it did well on accuracy, too: "Practically every one you did for this one you got right. I think you only got one wrong."

The children's final models, which they developed after this discussion, clearly showed the influence of the model-evaluation session and the success of Nikki's model. Only one of the final models was a box model; the rest were dimensional models. An interesting intermediate case is the final model developed by Brittany's group. (Recall that their first model was Figure 3, developed on the principle, "the more detail, the better.") Figure 5, their final model, has the structure of an attribute model, but it still illustrates the "more stuff is better" heuristic. Unlike other children, who sought to develop the most powerful or most efficient model (the one that attains the best accuracy with the fewest rules), Brittany's model retains redundant and ambiguous information—it is a hybrid of the group's first model and Nikki's stripped-down attribute model.

Summary: Fourth grade. The challenges encountered by the first and second graders also emerged in Grade 4, but here they were more quickly surmounted. The fourth graders, unlike the younger children, clearly considered the model as a model, that is, as logically constraining the members that could be admitted to categories. Yet, these children struggled with understanding why one would need a formal model (when their eyes alone could categorize pictures so effectively) and with ways of bracketing their knowledge to develop descriptors that were clear and unambiguous enough to apply without special knowledge of the stimuli. Particularly problematic were children's initial expectations that attribute values should apply exclusively within grade categories and that each member of a category needed to exemplify every rule for that category. By the close of instruction, some children seemed to appreciate that some rules are "more powerful" than others, but for many, the value of exhaustiveness appeared to outweigh the perceived value of parsimony.

Fifth Grade: Comparing the Structure of Models

In contrast to the fourth graders, the fifth graders focused less on fine-tuning descriptors and more on the overall structure of the models themselves. To illustrate this point, we focus here primarily on the cycles of models that the fifth graders produced and explain how questions of structure governed the model revisions at each cycle.

Initial models. The fifth graders began by working in small groups to sort the portraits "by eye" into four grade groups. They then attempted to account explicitly for the results generated by their "mind's eye." Their prototypical initial model consisted of a list of attributes. As in the fourth grade, most of these attributes were unique to each grade, although the attribute of detail often was qualified across grades (e.g., "First graders have bodies, clothes, and more detail than kindergartners"). The paragraph structure of the models was apparently related to the worksheet that Mark Rohlfing, the fifth-grade teacher, borrowed from Sally.

When these initial-list models were completed, Mark encouraged the groups to swap models so that each group could try to apply the model of a different group. The emphasis on having students try to use the models invented by others was unique among the classes we studied; moreover, it played an important role in inspiring model revision. When the children attempted to apply the descriptions provided by another group, problems immediately became apparent: "Can you tell me what detail means?" and "How can we say that Grade 3 drawings have eyes with pupils if not every single one has them at the third grade?"

The difficulties students initially encountered were similar to those experienced by the younger children—how to choose and describe meaningful attributes and how to deal with fuzzy sets, with their corresponding emphasis on family resemblance instead of necessary and sufficient features. Another failing became evident when students discovered that the models classified the majority of portraits into at least two grades. None of the initial models had the dimensional, attribute–value structure noted previously. Mark, the classroom teacher, relied on students' difficulties with using each others' models to prompt revision. He also introduced the word *attribute* and encouraged students to think about how an attribute of the portraits might take on different values across grades.

Intermediate models. To prompt revision, we introduced easel paper and asked groups to construct their models on it so they would be visible to the rest of the class. Once that was accomplished, we focused the children on comparing similarities and differences among the models they displayed. We also "seeded" the idea of dimensional structure by asking some of the groups to describe how one attribute, like the nature of the legs drawn in the portrait, varied by grade.

After working to revise their initial models, students compared and discussed problems they were continuing to experience when they attempted to apply the models developed by other groups. Some recurrent difficulties included ambiguous qualifiers, like *some* (Lehrer: "What did they mean by 'some'?" Student: "That's what we were trying to figure out!"). Many of the descriptions were also ambiguous: "In K it said, 'basic head shape.' We didn't know what basic head shape was." However, most of the discussion centered on the structure of the models. In this discussion, the contrast between initial models and variants of dimensional structure occupied center stage. For example, when considering a model, one group noted:

Student 1: It's [the model discussed] a mixture of Group 1's [a dimensional structure] and a table [here meaning the paragraph structure of the worksheet].

Student 2: It's the same subject [meaning attribute] but different answers, so that's like a table.

Lehrer: So, it's sort of like a table with attributes and many values within each grade. Was that easy to use, then?

Here, the conversation focused on parts of the model that were dimensional, as indicated by inclusion of attributes, but that also included multiple values for an attribute at a grade. Figure 6 displays an intermediate model developed by one group that displays an attribute–value structure but still retains the multiple values of the initial "paragraph" models.

Students also noted variation among the models with respect to the diagnostic significance of the attribute–value pairings. For example, one student said:

> We found that something they did really well is explained the K and first-grade attributes. We found the right people for each of those grades. Their attributes for K and for first were good. Something they need to improve on is third and fifth because they need to be more descriptive of what

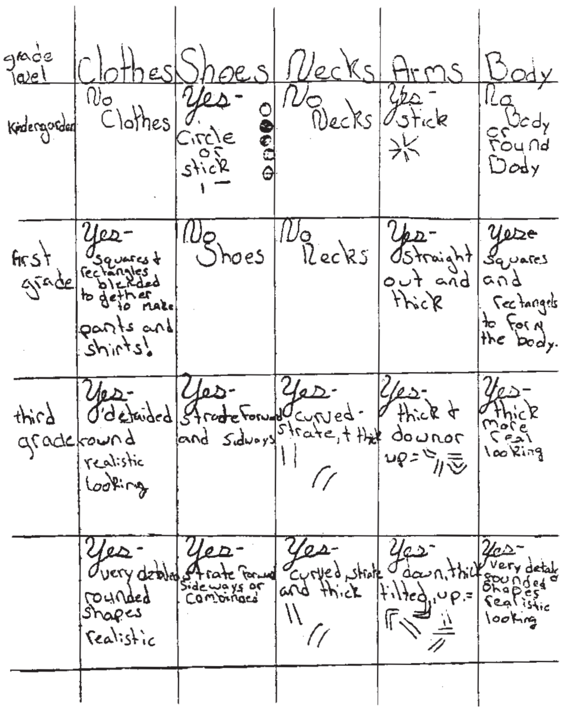

FIGURE 6 First model by a fifth-grade group.

> the third and fifth graders had and didn't have. We found a lot of kids that could be in third grade and only two that could be in fifth from their attributes.

Students also began to deal with the nature of fuzzy sets:

> Well, not every single one [portrait] has to fit every single one [attribute]. Say, if you had a K picture that fit every single one [of the attributes] but didn't have stick legs [a K attribute in the model], it could still be [drawn by] a kindergartner. It's more like a general relation.

Attention to the idea of partial rather than complete matching of attributes led directly to considerations about how to represent the degree of match:

Lehrer: So, you count up the number of matches to make your decision?
Student: That's exactly what we did. In their group, they said we couldn't find a lot of matches. More like 4 out of 5 or 3 out of 5.

In summary, during the second cycle of modeling, we seeded the dimensional data structure of rows defining grades and columns of attribute–values by explicitly suggesting to some groups that they consider how the values of a single attribute might change over grades. This form of assistance prompted the “spread” of dimensional data structures and instigated first considerations of the possibility of partial matching and the related problem of how to represent explicitly the degree of match (i.e., how to combine information). The initial difficulties students had with developing unambiguous descriptions of values for attributes like eyes persisted, as did students’ propensity to attempt to find the necessary and sufficient set of distinguishing attributes at every grade level. Nevertheless, at the end of this second cycle of modeling, the modal form of student model, exemplified in Figure 6, incorporated attribute–value notation even as it preserved elements of the initial paragraph models.

Third modeling cycle. To prompt further efforts to revise models, we provided students with a novel set of portraits and asked them to use their models to attempt to classify the grade level of each artist. Students noticed that many of their models were too “fine-tuned.” By this they meant that the models captured the idiosyncrasies of the first set of portraits but at the expense of poorer fit to the second set of portraits and other portraits in general. This recognition prompted further revisions, which we exemplified with a class composite model displayed in Figure 7, attributing the composite to one of the participating researchers (Vera).

This model illustrated how single-valued attributes could be employed in a tabular structure to enable quick and fairly accurate classification, a proposition tested with a third, larger set of portraits. Students noted that Vera’s model (the class composite) was “a lot more efficient,” “It doesn’t use words like ‘sometimes,’ ” and “I think the pictures helped.” Students went on to discuss whether the sheer number of attributes in Vera’s model was what was most important, but they concluded that it was their quality. They also noted that the distinction between kindergarten and first-grade drawings was much easier to capture than that between third and fifth grades. During this last cycle of modeling, students also noted that no model was able to classify all the pictures accurately, a result they had not even considered during the first cycle of modeling. Like the fourth graders, one group of fifth graders used iconic examples to supplement the text to represent attribute–value relations in a dimensional structure, as is apparent in Figure 8. Only one of the five groups used this solution in a final model, however. Most preferred textual descriptors of attributes and values.

Summary: Fifth grade. Like the younger students, the fifth graders also encountered problems that arose from vague descriptors, qualifiers, and their own

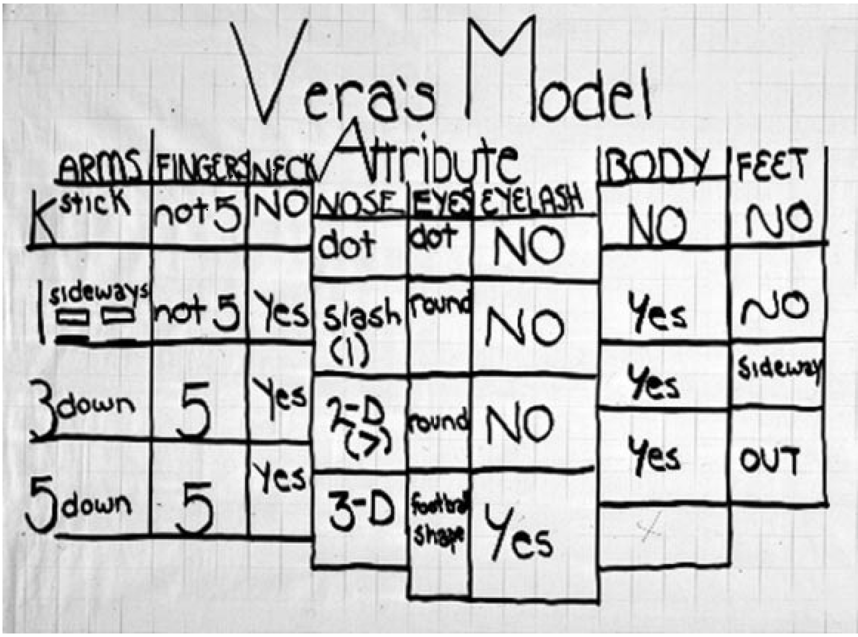

FIGURE 7 "Vera's model," proposed by the researchers.

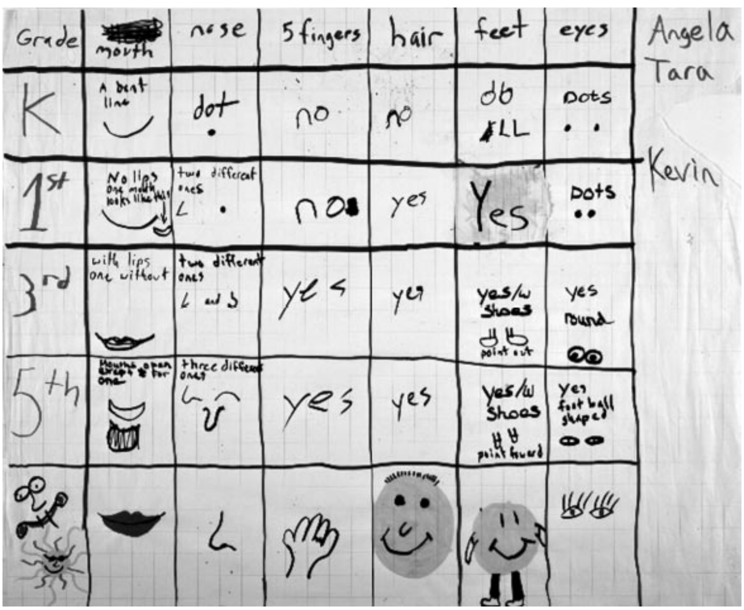

FIGURE 8 Use of iconic exemplars in an intermediate dimensional model proposed by one of the fifth-grade groups.

expectations that each rule should serve as both a necessary and sufficient condition for assignment to a category. However, most of the discussion and work in the fifth grade concerned the overall structure of models and the development of criteria for combining information across attributes. Fifth graders recognized most quickly, and with the least assistance, the utility of the attribute–value data structure and the need to develop decision rules for combining information to account for partial matching.

DISCUSSION

The very idea of a model was especially problematic for the youngest students, the first and second graders. Although we observed occasional glimmers of insight in individual children, for the most part, these students were highly reliant on adult assistance to make use of descriptors to classify cases, critique descriptors, and make even minor revisions. For the most part, their progress involved the development of more accurate and less ambiguous attributes within categories (grade levels). No lasting progress was made toward developing dimensions that would apply across grade levels.

In contrast, the fourth and fifth graders readily understood the “modeling game” (Hestenes, 1992), even though they struggled to implement it effectively. Like the younger children, they wrestled with problems of clarity and ambiguity. Occasional assistance and an audience for their work played an important role in developing models that could account for their perceptions of similarity. They made much greater progress than the younger students did toward developing and evaluating alternative data structures and agreeing on criteria for assigning cases to categories by combining attributes. Almost all the older students came to appreciate the value of a dimensional attribute–value structure for their model, although many children continued to prefer models that included redundant or additional information (such as illustrations of prototypes and additional descriptors) to leaner, more powerful models.

As a result of trying to fit their models to new data, some of the older students (i.e., the fifth graders) recognized the difficulties in generalizing a model that can be created by overfitting any particular set of data. This recognition played an important role in helping them appreciate the potential of more parsimonious models, although further work would have been required to explore the consequences of robustness. The oldest students also considered more explicitly the consequences of different types of decision rules, and like good statisticians everywhere, some even proposed weighing diagnostic attributes more heavily.

In addition to providing information about student learning, the cross-grade nature of the instruction made an important contribution to the professional development of the participating teachers and other teachers in the reform community.

During the fall that this case was pursued, teachers met to plan the outlines of instruction, talked regularly with each other about "how it was going," visited each others' classrooms to observe and actively assist in each other's teaching, analyzed videotapes of their classrooms together, and participated in interviews with researchers about how the experience had changed their thinking about their teaching. These activities played an important role in encouraging the consideration of children's thinking across grade levels. Although there was a previously existing tradition in the school district that supported within-grade "teaming," before the inception of this research there had been little precedent or institutional support for ongoing cross-grade work. Yet, such opportunities are of obvious importance for the generation and maintenance of developmentally oriented instruction.

ACKNOWLEDGMENTS

The alphabetical order of the authors denotes fully collaborative work and equal contribution.

This research was supported by the James S. McDonnell Foundation and by a grant from the U.S. Department of Education, Office of Educational Research and Improvement, to the National Center for Improving Student Learning and Achievement in Mathematics and Science (Contract R305A600007). We thank the participating teachers, Sally Hanner, Eric James, and Mark Rohlfing, and their students, and we gratefully acknowledge the contributions of our research staff and students, especially Nancy D. Giles and Vera Kemeny. The opinions expressed herein do not necessarily reflect the position, policy, or endorsement of the supporting agencies.

REFERENCES

Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Braine, L., Schauble, L., Kugelmass, S., & Winter, A. (1993). The representation of depth by children: Spatial strategies and lateral biases. *Developmental Psychology, 19*, 466–479.

Fodor, J. (1995). Concepts: A potboiler. In J. Mehler & S. Franck (Eds.), *Cognition on cognition* (pp. 133–151). Cambridge, MA: MIT Press.

Gardner, H. (1980). *Artful scribbles: The significance of children's drawings*. New York: Basic Books.

Hancock, C., Kaput, J. J., & Goldsmith, L. T. (1992). Authentic inquiry with data: Critical barriers to classroom implementation. *Educational Psychologist, 27,* 337–364.

Hestenes, D. (1992). Modeling games in the Newtonian world. *American Journal of Physics, 60,* 440–454.

Inhelder, B., & Piaget, J. (1964). *The early growth of logic in the child*. New York: Norton.

Lehrer, R., & Romberg, T. (1996). Exploring children's data modeling. *Cognition and Instruction, 14,* 69–108.

Lehrer, R., & Schauble, L. (in press). Modeling in mathematics and science. In R. Glaser (Ed.), *Advances in instructional psychology: Vol. 5. Educational design and cognitive science.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

McClelland, J. L., & Rumelhart, D. E. (1986). A distributed model of human learning and memory. In J. McClelland & D. Rumelhart (Eds.), *Parallel distributed processing* (Vol. 2, pp. 170–215). Cambridge, MA: MIT Press.

Panchen, A. L. (1992). *Classification, evolution, and the nature of biology*. Cambridge, England: Cambridge University Press.

Smith, E. E., & Medin, D. L. (1981). *Categories and concepts*. Cambridge, MA: Harvard University Press.

Smith, L. B. (1989). A model of perceptual classification in children and adults. *Psychological Review*, *96*, 125–144.

Sneath, P. H. A., & Sokal, R. R. (1973). *Numerical taxonomy*. San Francisco: Freeman.