From: ***Learning Mathematics for a New Century, 2000 Yearbook*.** (2000) Burke, Maurice J. and Curcio, R. Frances (Eds.). National Council of Teachers of Mathematics, Reston, VA**. ***

12

Supporting Students' Ways of Reasoning about Data

Kay McClain

Paul Cobb

Koeno Gravemeijer

OUR purpose in this article is to describe how one group of students came to reason about data while developing statistical understandings related to exploratory data analysis. In doing so, we will present episodes taken from a seventh-grade classroom in which we conducted a twelve-week teaching experiment. (The first two authors shared teaching responsibilities during the teaching experiment.) One of the goals of the teaching experiment was to investigate ways to support middle school students' development of statistical reasoning proactively. Our interest was piqued by current debates about the role of statistics in school curricula (National Council of Teachers of Mathematics 1989, 1991; Shaughnessy 1992). The image that emerged for us as we read this literature was that of students engaging in instructional activities in which they both developed and critiqued data-based arguments.

As we worked to develop an instructional sequence, we viewed the use of computer tools as an integral aspect of statistical reasoning rather than as technological add-ons. As such, the two computer tools we designed were intended to support students' emerging mathematical notions while simultaneously providing them with tools for data analysis. In the twenty-first century, access to information will continue to be enhanced by new technologies. This fact highlights the importance of providing students with

---

The research reported was supported by the National Science Foundation under Grant No. DMS-9057141 and RED-9353587 and by the Office of Educational Research and Improvement under Grant No. R305A60007. The opinions expressed are solely those of the authors.

opportunities to develop and critique data-based arguments in situations that are facilitated by technologies. It is therefore imperative that learning opportunities of this type in which students can develop deep understandings of important statistical ideas become central aspects of school curricula.

In the following sections of this paper, we first outline the intent of the instructional sequence we used in the seventh-grade classroom and describe the role of the computer tools in this sequence. Against this background we then describe the classroom and present episodes intended to highlight students' development of sophisticated ways to reason about data.

INSTRUCTIONAL SEQUENCE

As we began to design the instructional sequence to be used in the seventh-grade classroom, we attempted to identify the big ideas in statistics. Our plan was to develop a single, coherent sequence and thus tie together the separate, loosely related topics that typically characterize middle school statistics curricula. In doing so, we came to focus on the notion of distribution. This enabled us to treat notions such as mean, mode, median, and frequency as well as others, such as "skewness" and "spread-outness," as characteristics of distributions. It also allowed us to view various conventional graphs such as histograms and box-and-whiskers plots as different ways of structuring distributions. Our instructional goal was therefore to support students' gradual development of a single, multifaceted notion–that of distribution–rather than a collection of topics to be taught as separate components of a curriculum unit. In formulating hypotheses about how the students might reason about distributions, one of our primary goals was that students would think about data sets as entities that have properties in their own right rather than as collections of points (Hancock, Kaput, and Goldsmith 1992; Konold et al. in press; Mokros and Russell 1995). We conjectured that if students did begin to think about data in this way, they could then investigate ways of structuring data sets that would help them identify trends and patterns.

As we began mapping out the instructional sequence, we were guided by the premise that the integration of computer tools was crucial in supporting our mathematical goals. Students would need efficient ways to organize, structure, describe, and compare large data sets. This could best be facilitated by the use of computer tools for data analysis. However, we tried to avoid creating tools for analysis that would offer either too much or too little support. This quandary is captured in the current debate about the role of technologies in supporting students' understandings of data and data analysis. This debate is often cast in terms of what has been defined as expressive and exploratory computer models (Doerr 1995). In one of these approaches–the expressive–students are expected to recreate conventional graphs with only an occasional nudging from the teacher. In the other approach–the exploratory–students work with computer software that presents a range of conventional graphs with the expectation that the students will develop mature mathematical understandings of their meanings as they use them. The approach that we took when designing computer-based tools for data analysis offers a middle ground between the two approaches. It introduces particular tools and ways of structuring data that are designed to fit with students' current ways of understanding while simultaneously building toward conventional graphs (Gravemeijer et al. in press).

The instructional sequence developed in the course of the seventh-grade teaching experiment involved two computer minitools. In the initial phase of the sequence, which lasted for almost six weeks, the students used the first minitool to explore sets of data. This minitool was explicitly designed for this instructional phase and provided a means for students to manipulate, order, partition, and otherwise organize small sets of data in a relatively routine way. Part of our rationale in designing this tool was to support students' ability to analyze data as opposed to simply "doing something with numbers" (McGatha, Cobb, and McClain 1998). When data were entered into the tool, each individual data value was shown as a bar, the length of which signified the numerical value of the data point (see fig. 12.1).

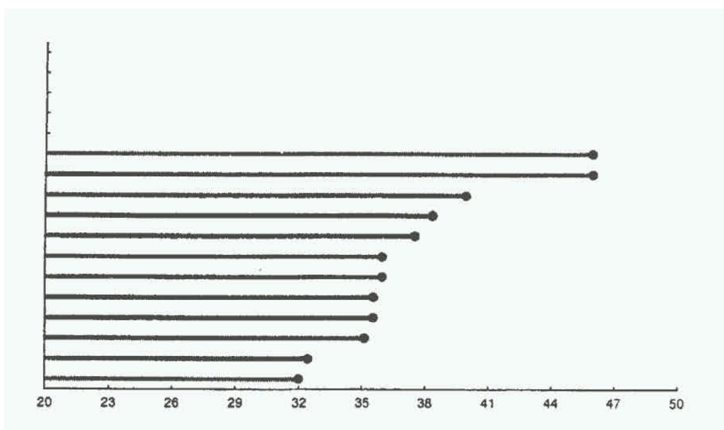

Fig. 12.1. Data displayed on the first minitool

A data set was therefore shown as a set of parallel bars of varying lengths that were aligned with an axis. Its use in the classroom made it possible for students to act on data in a relatively direct way. This would not have been possible had we used commercially available software packages for data analysis whose options typically include only a selection of conventional graphs. The first computer minitool also contained a value bar that could be dragged along the axis to partition data sets or to estimate the mean or to mark the median. In addition, there was a tool that could be used to determine the number of data points within a fixed range.

The second computer minitool can be viewed as an immediate successor of the first. As such, the endpoints of the bars that each signified a single data point in the first minitool were, in effect, collapsed down onto the axis so that a data set was now shown as collection of dots located on an axis (i.e., an axis plot as shown in fig. 12.2).

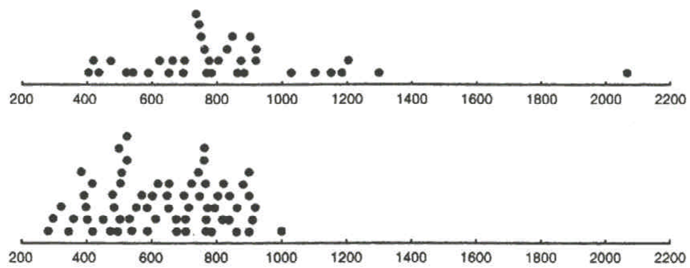

Fig. 12.2. Data displayed on the second minitool

The tool offered a range of ways to structure data. The tool's palate of options did not correspond to a variety of conventional graphs, unlike the palates of typical software packages that are available commercially. Instead, we designed the various options after identifying in the literature the various ways in which students structure data when they are given the opportunity to develop their own approaches while conducting genuine analysis (Hancock et al. 1992). Two of the options can be viewed as precursors to standard ways of structuring and inscribing data. These are organizing the data into four equal groups so that each group contains one-fourth of the data (precursor to the box-and-whiskers plot) and organizing data into groups of a fixed interval width so that each interval spans the same range on the axis (precursor to the histogram). However, the three other options available to students do not correspond to graphs typically taught in school. These involve structuring the data by (1) making your own groups, (2) partitioning the data into groups of a fixed size, and (3) partitioning the data into two groups of equal size. The first and least sophisticated of these options simply involved dragging one or more bars to chosen locations on the axis in order to partition the data set into groups of points. The number of points in each group was shown on the screen and adjusted automatically as the bars were dragged along the axis. The key point to note is that this tool was designed to fit with students' ways of reasoning while simultaneously taking important statistical ideas seriously.

As we worked to outline the sequence, we reasoned that students would need to encounter situations in which they had to develop arguments based on the reasons for which the data were generated. In this way, they would need to develop ways to analyze and describe the data in order to substantiate their recommendations. We anticipated that this would best be achieved by developing a sequence of instructional tasks that involved either describing a data set or analyzing two or more data sets in order to make a decision or a judgment. The students typically engaged in these types of tasks in order to make a recommendation to someone about a practical course of action that should be followed. An important aspect of the instructional sequence involved talking through the data creation process with the students. In situations where students did not actually collect the data themselves, we found it very important for them to think about the types of decisions that are made when collecting data in order to answer a question. The students typically made conjectures and offered suggestions about the information that would be needed in order to make a reasoned decision. Against this background, they discussed the steps that they might take to collect the data. These discussions proved critical in grounding the students' data analysis activity in the context of a recommendation that had real consequences.

CLASSROOM EPISODES

During the teaching experiment, the students explained and justified their reasoning in whole-class discussions. This was facilitated by the use of a computer projection system that allowed students to display the data sets and show the ways that they had structured the data. In addition, the students produced written arguments that we often used as a basis for classroom discussions. As the sequence progressed, we also asked the students to draw inscriptions in order to support their recommendations. (These inscriptions included nonconventional graphs and symbolic summaries of the data such as an axis marked with a three- or five-point summary.) In doing so they had to develop ways to show trends and patterns that supported their argument without reproducing data sets in their entirety.

One of the initial task situations that was posed to the students involved comparing two separate brands of batteries, Always Ready and Tough Cell. In discussing the task situation, students noted that a better battery would be one that lasted a long time. They framed the discussion around their experiences with using batteries in several types of electronic devices and reflected on how often they had to purchase batteries. They then offered numerous suggestions for how batteries might be tested. Against the background of this discussion, they were asked to analyze data from a sample of ten batteries of each of two brands that had been tested to determine how long they would last (see fig. 12.3).

The students then worked in pairs on the computers and used the minitool to organize and structure the data in order to help them make a decision. Afterwards, they discussed the results of their analysis in a wholeclass setting.

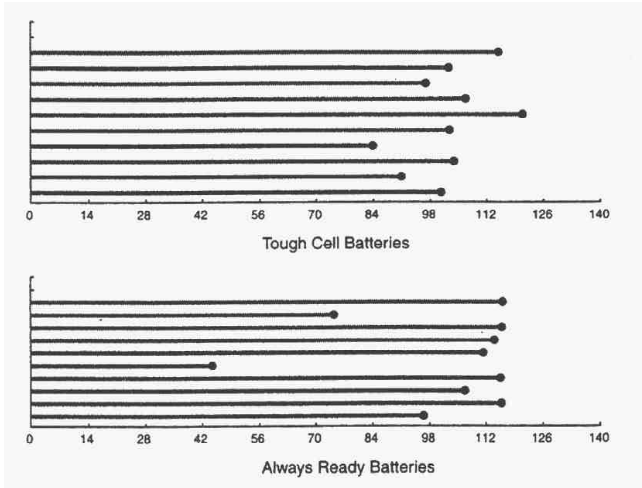

Fig. 12.3. Data on two brands of batteries

Celia was the first student to share her argument. She began by explaining that she used the range tool to identify the top ten batteries out of twenty that were tested. In doing so, she found that seven of the ten longest lasting were Always Ready batteries. During the discussion of Celia's explanation, Bradley raised his hand to say that he compared the two brands of batteries a different way.

Bradley: Can you put the representative value on 80? Now, see there's still [Always Ready batteries] behind 80, but all the Tough Cell is above 80 and I'd rather have a consistent battery that is going to give me above 80 hours instead of one. I just have to guess.

Teacher 1: Questions for Bradley? Janine?

Janine: Why wouldn't the Always Ready battery be consistent?

Bradley: All your Tough Cells is above 80 but you still have two behind 80 in the Always Ready.

Janine: Yeah, but that's only two out of ten.

Bradley: Yeah, but they only did ten batteries and the two or three will add up. It will add up to more bad batteries and all that.

Janine: Only wouldn't that happen with the Tough Cell batteries?

Bradley: The Tough Cell batteries show on the chart that they are all over 80 so it seems to me they would all be better.

Janine: *(nods okay).*

Bradley based his argument on the observation that all the Tough Cell batteries lasted at least 80 hours. He used the value bar to partition the data and determined that Tough Cell was a more consistent brand.

In comparing Celia's and Bradley's arguments, it is significant to note that while the students understood what Celia had done (compared the number of batteries of each brand in the top half of the data), her choice of the "top ten" was open to question. For instance, one student pointed out that if she had chosen the top fourteen batteries instead of the top ten, there would be seven of each brand. Celia's choice of the top ten was arbitrary in the sense that it was not grounded in the context of the investigation. Bradley, however, gave a rationale for choosing 80 hours that appeared to make sense to the students. He wanted batteries that he could be assured would last a minimum of 80 hours. As a consequence of this rationale for partitioning, his argument appeared to be accepted as valid.

The students continued to engage in similar investigations using the first computer minitool for several weeks. As they did so, many of them frequently used the value bar to partition the data. They would place the bar at a particular value along the axis and then reason about the number of data points that were above or below that value. It is important to note that the value at which they partitioned the data were typically not arbitrary. For instance, in a task about health care, many placed the bar at 65 years, arguing that this enabled them to focus on senior citizens in comparison to the rest of the population. We should clarify that we did not anticipate that the students would use the value bar in this way. Our expectation when we designed the tool was that they might use the bar to estimate the mean of a data set. Instead, the students adapted this feature of the minitool to their current ways of thinking about data.

As we worked with the students, we developed and modified tasks by analyzing their reasoning in each classroom session. We introduced the second minitool once we judged that the students had developed a variety of ways to organize the data and an understanding of what is involved in a data-based argument. Because the second tool displayed data as an axis plot, we were able to increase the number of data points in the sets.

One of the initial investigations with the second computer tool involved analyzing data on driving speeds on a very busy road in the city. As drivers are known to speed on this particular stretch of highway, the police department had decided to set up a speed trap to try to slow the traffic. Students discussed what information would be necessary to determine if the speed trap was effective. After much discussion on both issues of safety related to speed and on the specifics of this particular problem, students were asked to compare data on the speeds of sixty cars before the speed trap was set up with the speeds of sixty cars a month later to decide if the speed trap was effective (see fig. 12.4).

After students had analyzed the data using the computer minitool, we asked them to develop a written argument that could be submitted to the Chief of Police. In the subsequent whole-class discussion, students read their reports as part of their explanations. The first argument was presented by Janine.

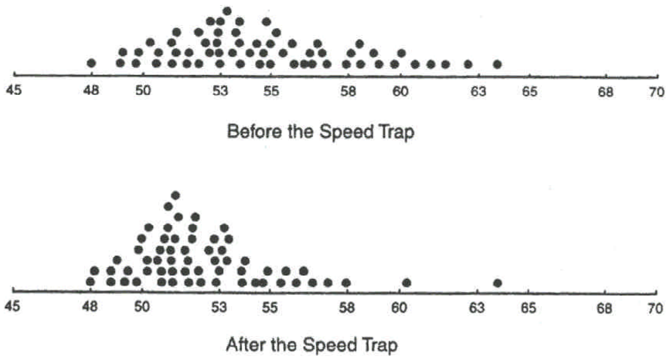

Fig. 12.4. Data on speeds before and after the speed trap

Janine: If you look at the graphs and look at them like "hills," then for the before group it is more spread out and more are over 55. If you look at the after graph, then more people are bunched up closer to the speed limit, which means that the majority of the people slowed down.

After Janine finished, the teacher used the projection system to display the data on the white board.

Teacher 1: Okay, Janine said if you look at this like hills ... now think about this as a hill *(draws a hill over the data in the first data set as shown in* fig. 12.5) and think about this as being a hill *(draws a hill over the second data set as shown in fig. 12.5)*. See what Janine was talking about? Before the speed trap the hill was spread out, but after the speed trap the hill got more bunched up and less people were speeding.

Students seemed to understand Janine's argument and saw the relevance of her idea of "hills," as indicated by Kent's comment below.

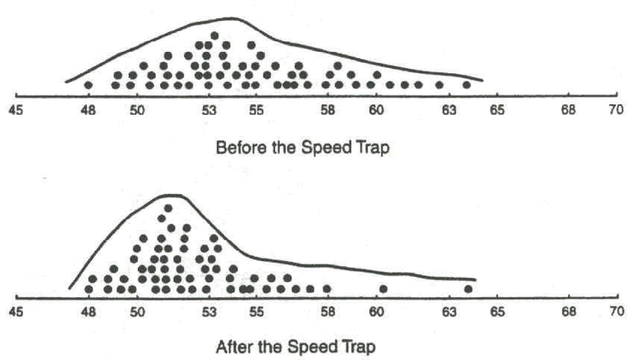

Fig. 12.5. Speed data with "hills" shown

Kent: They were slowing down. I want to compliment Janine on the way that she did that. I couldn't find out some way to compare, and I think that was a good way.

In this particular task, students attempted to find a way to organize and describe the data so that they could made a recommendation. The key point for us is that the notion of a data set as a distribution of data points emerged as the students discussed Janine's "hills" interpretation. She was concerned with how the data were distributed and focused on qualitative proportions of the data (e.g., the majority). In this episode, the students began to reason about global trends in entire data sets. Previously, they had focused on the number of data points in parts of data sets.

A further shift in the students' reasoning can be seen in an episode that occurred six days later. In developing the task, we had reasoned that we needed to find a situation in which the number of data points in the two data sets were very different. We hoped that this would lead to opportunities to question students' arguments that involved simple partitions and additive reasoning rather than proportional reasoning. The task we developed asked the students to analyze the T-cell counts of two groups of AIDS patients who had enrolled in different treatment protocols. A lengthy discussion revealed that the students were quite knowledgeable about AIDS and understood the importance of finding an effective treatment. Further, they clarified the relation between T-cell counts and a patients' overall health: increased T-cell counts are desirable. In the task, students were given data on the T-cell counts of 46 patients in a new, experimental treatment and the T-cell counts of 186 patients in a standard protocol (see fig. 12.6). The students were asked not only to make a recommendation about which protocol was more effective, but also to develop inscriptions that could be used to support their arguments.

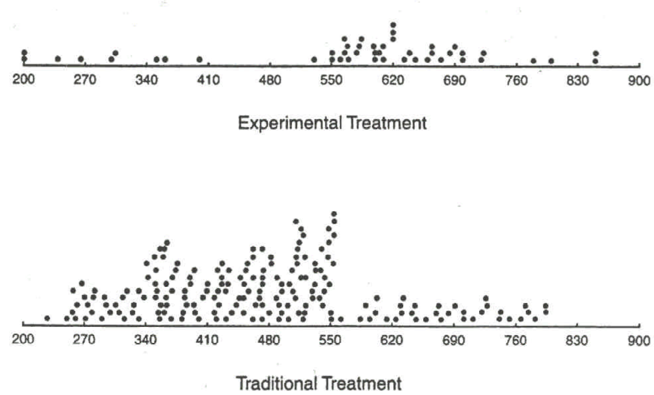

Fig. 12.6. Data on T-cell counts of AIDS patients

As the students worked both at their computers and in groups, we monitored their activity in order to select students whose arguments might provide opportunities for shifts in mathematical thinking to occur. In one of the first reports that was discussed, the students had partitioned the data at a T-cell count of 550; they found that most of the data in the standard protocol was below a 550 T-cell count and most of the data in the experimental protocol was above a 550 T-cell count. During the discussion, the teacher clarified that these students had chosen the T-cell count of 550 because the "hill" of one data set was mostly below this value and the "hill" of the other was mostly above (see fig. 12.7).

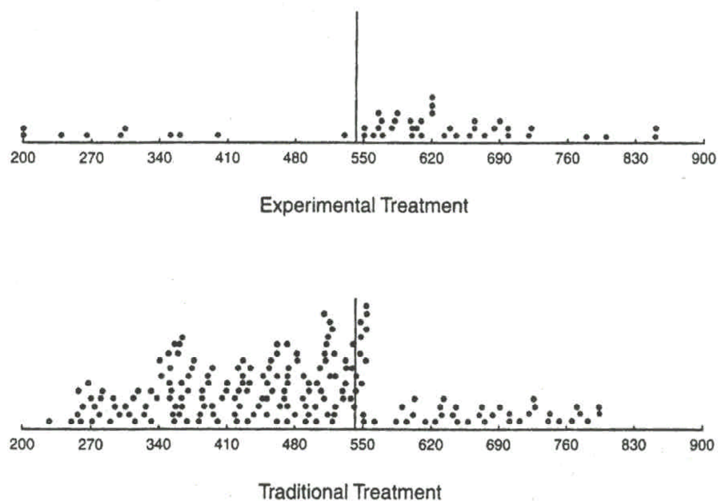

Fig. 12.7. AIDS data with partition between the "hills"

Thus, the students had partitioned at a particular value in order to develop a quantitative description of a perceived qualitative difference between the two data sets. Toward the end of the discussion, Janine made the following comment.

Janine: I think it would be helpful to know how many of the possible [patients] were in that range.

At a student's suggestion, the teacher then drew a diagram that recorded the number of patients in each treatment with T-cell counts above and below 550 (see fig. 12.8).

Teacher 2: Hey, I've got a question for everybody. Couldn't you just argue rather convincingly that the old treatment was better 'cause there were 56 people over 550? Fifty-six patients had T-cell counts greater than 550, and here there are only 37, so the old has just got to be better. I mean, there are 19 more in there, so that's the better one, surely.

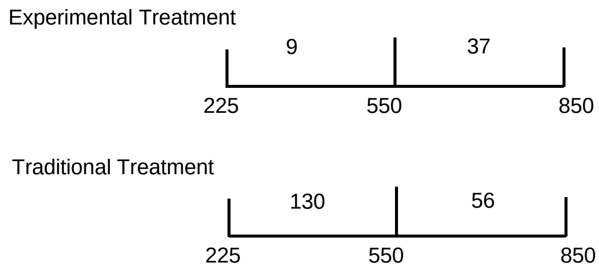

Fig. 12.8. Inscription showing partitions of AIDS data

Bradley: But there is more in the old.

Jake: Thirty-seven is more than half of 9 and 37, but 56 is not more than half of 130 and 56.

Kent: I've got a suggestion. I don't know how to do it *(inaudible).* Is there a way to make 130 and 156 compare to 9 and 37? I don't know how...

Kent's suggestion indicates that he wanted to find a way of comparing the information in the diagram. It was as if he was stating design specifications for the ideas of relative frequencies. The comments of the other students reveal that they also questioned the teacher's additive argument. The teacher capitalized on these contributions by introducing percentages as a solution to Kent's proposal. The students quickly calculated the percentages of patients in each treatment above and below a T-cell count of 550 and used the results to substantiate their initial arguments. After they had finished, Bradley made an observation.

Bradley: See, coming up with the percent of data cancels all the different numbers of data.

In the next report that was discussed, the students had used the computer minitool to organize the data into four equal groups. The inscription they developed consisted of axes marked only with each of the resulting intervals, similar to a box-and-whiskers plot (see fig. 12.9).

Bradley: On the four equal groups you can tell where the differences is in the groups.

Teacher 1: Can you do that by looking at this [inscription]? So, what do you see when you look at this, Bradley?

Bradley: That the new treatment was better than the old treatment.

Teacher 1: And what are you basing that comment on?

Bradley: Because the three lines for the equal groups [for the new treat ment] were all above 525 compared to only one on the old one.

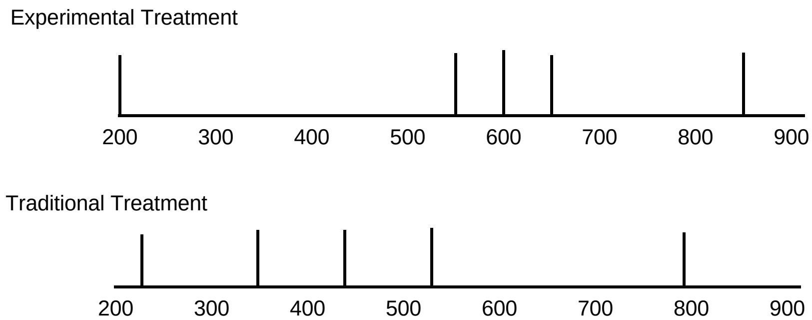

Fig. 12.9. Inscription of AIDS data partitioned into four equal groups

In this exchange, Bradley clarified the conclusions that could be drawn from the inscription. He could see that 75 percent ("three lines") of the patients in the new treatment had T-cell counts greater than 525, whereas only 25 percent ("only one") of the patients in the old treatment had comparable T-cell counts. For him, this justified recommending the new treatment.

In the subsequent discussion, students focused on not only the validity of the argument but also the adequacy of the inscription. In doing so, they began to develop ways of using the four-equal-groups inscription to make arguments about the data. As they worked to understand the inscription, the validity of the argument was strengthened.

Martin: I think it would help to have the numbers to know how many were in each group.

Bradley: It doesn't really matter where all the data is, 'cause you know which group is better, 'cause you know where the data is. 'Cause you already see where the data is in the groups.

Other comments that Bradley made indicated that he (and several other students) had come to understand that 25 percent of the data were in each interval. He could therefore "see where the data is in the groups," whereas Martin needed to know exactly how many data points were in each interval. Differences such as these indicate the range in the students' interpretations. As a consequence of this diversity, the teacher encouraged the students to explain their reasoning by using the computer minitool and the inscriptions they had created. This enabled students such as Martin to contribute to the discussion of the four-equal-groups inscription in personally meaningful ways.

By the end of the classroom teaching experiment, over half of the students routinely described a part of a distribution as a proportion or percentage of the whole. In doing so, they reasoned about what Konold and colleagues (in press) have called *group propensities* (i.e., the rate of occurrence of some data value within a group that varies across a range of data values). Konold and colleagues argue that propensity is at the heart of what they call a *statistical perspective.* As we have seen, many of the students also structured data sets using what the statistician David Moore calls the *five-point summary* (i.e., extreme values, median, and quartiles) in order to characterize differences between distributions for a specific reason or purpose. In addition, their arguments now involved justifying the statistics they used to compare data sets. For example, students justified partitioning the data at a T-cell count of 550 based on the location of the hills as opposed to an arbitrary value such as the midpoint of the range. The whole-class discussions during the final sessions also indicate that almost all the students were now able to make and understand arguments that focused on patterns in how the data were distributed. It is significant to note that the students often changed their initial judgments in the course of whole-class discussions. This reveals that their ways of reasoning were constantly being challenged and modified by other's arguments.

CONCLUSION

We would stress that the purpose of the instructional sequence was not that the students might come to create specified graphs in particular situations or calculate measures of central tendency correctly. Most could already do the latter, although with little understanding. Instead, it was that they might develop relatively deep understandings of important statistical ideas as they used the computer minitools to structure data and make data-based arguments. It was for this reason that whole-class discussions throughout the classroom teaching experiment focused on the ways in which students organized data in order to develop arguments. In addition, students seemed to reconceptualize their understanding of what it means to know and do statistics as they compared and contrasted solutions. The crucial norm that became established was that of explaining and justifying solutions in the context of the problem being explored. This is a radically different approach to statistics than is typically introduced in middle schools. It highlights the importance of middle school curricula that allow students to engage in genuine problem solving that supports the development of central mathematical concepts.

There is much talk of preparing students for the information age but without fully acknowledging that the information in this new era will be largely statistical in nature. Cast in these terms, statistical literacy for the twentyfirst century will involve reasoning with data in relatively sophisticated ways. New computer tools will provide opportunities for students to deal with information that is readily available, but often unmanageable with only paper and pencil. Classroom experiences in which students use computer-based tools to help them think and reason about problem situations serve to prepare them for the twenty-first century.

REFERENCES

Doerr, Helen. "An Integrated Approach to Mathematical Modeling: A Classroom Study." Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, April 1995.

Gravemeijer, Koeno, Paul Cobb, Janet Bowers, and Joy Whitenack. "Symbolizing, Modeling, and Instructional Design." In *Symbolizing and Communicating in Mathematics Classrooms: Perspectives on Discourse, Tools, and Instructional Design*, edited by Paul Cobb, Erna Yackel, and Kay McClain. Mahwah, N.J.: Lawrence Erlbaum Associates, in press.

Hancock, Chris, James Kaput, and Lynn Goldsmith. "Authentic Inquiry with Data: Critical Barriers to Classroom Implementation." *Educational Psychologist 27* (1992): 337-64.

Konold, Cliff, Alexander Pollatsek, Arnold Well, and Allen Gagnon. "Students Analyzing Data: Research of Critical Barriers:" *Journal for Research in Mathematics Education* (in press).

McGatha, Maggie B., Paul Cobb, and Kay McClain. "An Analysis of Students' Statistical Understandings:' Paper presented at the Annual Meeting of the American Educational Research Association, San Diego, April 1998.

Mokros, Jan, and Susan Russell. "Children's Concepts of Average and Representativeness." *Journal for Research in Mathematics Education 26* (January 1995): 20-39.

National Council of Teachers of Mathematics. *Curriculum and Evaluation Standards for School Mathematics.* Reston, Va.: National Council of Teachers of Mathematics, 1989.

_______. *Professional Standards for Teaching Mathematics.* Reston, Va.: National Council of Teachers of Mathematics, 1991.

Shaughnessy, Michael. "Research on Probability and Statistics: Reflections and Directions:" In *Handbook of Research on the Teaching and Learning of Mathematics*, edited by Douglas Grouws, pp. 465-94. New York: Macmillan, 1992.