SUPPORTING TEACHERS' UNDERSATNDING OF STATISTICAL DATA ANALYSIS: LEARNING TRAJECTORIES AS TOOLS FOR CHANGE

Kay McClain
Vanderbilt University
USA

This paper provides an analysis of a Teacher Development Experiment (Simon, 2000) designed to support teachers' understandings of statistical data analysis. The experiment addresses the following research question: Can the results from research conducted in a middle-grades mathematics classroom be used to guide teachers' learning? In both cases, activities from an instructional sequence designed to support the development of ways to reason statistically about data were the basis of engagement. Analyses of the episodes in this paper document that the learning trajectory that emerged from the teachers' activity did, in many significant ways, parallel that of the students.

INTRODUCTION

The purpose of this paper is to provide an analysis of the development of one group of teachers' understandings of statistical data analysis. The analysis builds from the literature on students' understandings by taking prior research as a basis for conjectures about means of supporting teachers' development. In particular, the analysis in this paper will focus on a collaborative effort conducted between the author of this paper and a cohort of middle-school teachers. The collaboration occurred during the 1999 – 2000 academic year. The teachers participated in monthly work sessions designed to support their understandings of effective ways of teaching statistical data analysis in the middle grades. Fundamental to this effort was support of the development of the teachers' content knowledge. The instructional activities utilized were taken from a classroom teaching experiment conducted with a group of seventh-grade students during the fall semester of 1997 (for a detailed analysis of the classroom teaching experiment see Cobb, 1999; McClain, Cobb, & Gravemeijer, 2000; and McClain & Cobb, in press). The intent of the instructional sequence is to support middle-school students' development of sophisticated ways of reasoning statistically about univariate data. The overarching goal is that they come to reason about data in terms of distributions. Inherent in this understanding is a focus on multiplicative ways of structuring data.

The intent of the teacher collaboration was to build from the mathematical practices that emerged in the course of the classroom teaching experiment (Cobb, 1999). The hypothesis was that the same general learning trajectory (i.e., a parallel progression of conceptual development) could serve as a basis for guiding the mathematical development of the teachers. This trajectory served as a conjecture about the learning route of the teachers and the means of supporting their development. During the collaboration, the conjecture was continually being tested and refined in the course of interactions with the teachers. The trajectory therefore offered a conjectured route through the mathematical terrain. This conjecture included not only taking the classroom mathematical practices as a basis for the learning route of the teachers, but also taking the accompanying *means of support* as tools for supporting the emergence of the mathematical practices. These tools included the choice of tasks, the use of computer-based tools for analysis, the use of the teachers' inscriptions and solutions, and the norms for argumentation

In the following sections of this paper, I begin by outlining the instructional sequence. I then describe the methodology used in the analysis followed by a description of the setting. Against this background, I provide an analysis of episodes from the work sessions intended to document the teachers' developing understandings of statistical data analysis. I conclude by returning to the instructional sequence and its underlying learning trajectory as a means of supporting teachers' mathematical development.

INSTRUCTIONAL SEQUENCE

In developing the instructional sequence for the seventh-grade classroom teaching experiment, our goal was to develop a coherent sequence that would tie together the separate, loosely related topics that typically characterize American middle-school statistics curricula. The notion that emerged as central from our synthesis of the literature was that of distribution. In the case of univariate data sets, for example, this enabled us to treat measures of center, spread, skewness, and relative frequency as characteristics of the way the data are distributed. In addition, it 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 the development of a single, multi-faceted notion that of distribution, rather than a collection of topics to be taught as separate components of a curriculum unit. A distinction that we made during this process which later proved to be important is that between reasoning additively and reasoning multiplicatively about data (Harel & Confrey, 1994; Thompson, 1994; and Thompson & Saldanha, 2000). Multiplicative reasoning is inherent in the proficient use of a number of conventional inscriptions such as histograms and box-and-whiskers plots.

As we began mapping out the instructional sequence, we were guided by the premise that the integration of computer tools was critical in supporting our mathematical goals. The instructional sequence developed in the course of the seventh-grade teaching experiment in fact 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. When data was 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 Figure 1).

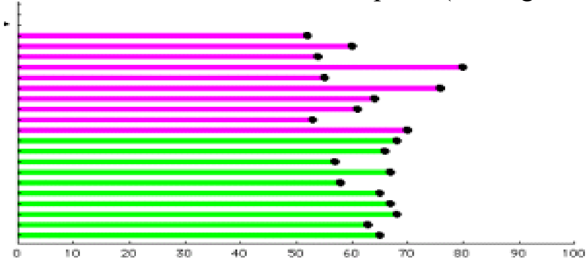

Figure 1. Data Displayed in First Minitool

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. 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. Students' activity with this tool supported the emergence of the first mathematical practice, that of *exploring qualitative characteristics of collections of data points* (Cobb, 1999).

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 a collection of dots located on an axis (i.e. an axis plot as shown in Figure 2). The tool offered a range of ways to structure data. The options included: (1) making your own groups, (2) partitioning the data into groups of a fixed size, (3) partitioning the data into equal interval widths, (4) partitioning the data into two equal groups, and (4) partitioning the data into four equal groups. 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. The second tool made possible the emergence of the second mathematical practice, that of *exploring qualitative characteristics of distributions* (Cobb, 1999).

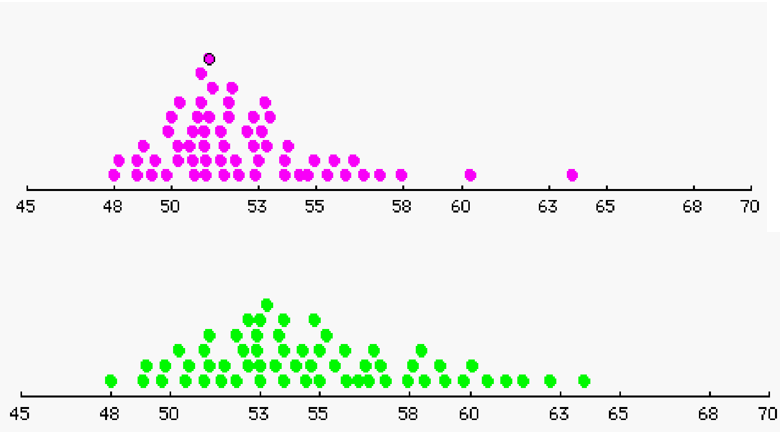

Figure 2. Data Displayed in Second Minitool

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 was 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.

METHODOLOGY

Detailed longitudinal analyses from the seventh-grade classroom teaching experiment had resulted in the specification of the sequence of classroom mathematical practices that emerged (for a detailed analysis of the emergence of these practices see Cobb, 1999). These classroom mathematical practices are the normative ways of reasoning, arguing, and symbolizing established by the classroom community while discussing particular mathematical ideas. An indication that a classroom mathematical practice has been established is that explanations about that practice are no longer necessary; they are beyond justification. (For a detailed analysis of the development of mathematical practices, see Cobb, Stephan, McClain, & Gravemeijer, 2001.) These classroom mathematical practices served as the initial conjectured learning trajectory for the teacher development experiment.

The particular lens that guided my analysis of the data was therefore a focus on the normative ways of solving tasks, or the classroom mathematical practices, which focuses on the collective mathematical learning of the classroom community (Cobb, Stephan, McClain, & Gravemeijer, 2001). It therefore enabled me to document the collective mathematical development of the classroom community over the period of time covered by the instructional sequence. This theoretical lens enables one to talk explicitly about collective mathematical learning. For this reason, the conjectured learning trajectory can be described as a conjectured sequence of mathematical practices along with the means of supporting the emergence of each successive practice from prior practices. The analysis then tests these conjectures against the learning trajectory to document the learning of the community.

In order to conduct the analysis, it is important to focus on the diverse ways in which the teachers participate in communal classroom practices. For this reason, the participation of the teachers in discussions where their mathematical activity is the focus then becomes the data for analysis. This diversity in reasoning also serves as a primary means of support of the collective mathematical learning of the classroom community. An analysis focused on the emergence of classroom mathematical practice is therefore a conceptual tool that reflects particular interests and concerns (Cobb, et al., 2001).

ANALYSIS OF CLASSROOM EPISODES

The reader will recall that the first mathematical practice that emerged in the seventh-grade classroom was that of exploring qualitative characteristics of data points. The initial activities in support of the emergence of this practice involved the teachers analyzing data on the braking distances of ten each of two makes of cars, a coupe and a sedan. I introduced the task by first talking through the data creation process with the teachers. I then presented the data by giving the teachers paper copies of the data inscribed in the first minitool as shown in Figure 1. I asked the teachers to work at their tables to decide which make of car they thought was safer, based on this data. My decision to use printouts of the data was based on my own experience in working with students on these tasks. I had noticed that when students were asked to make initial conjectures based on analysis of the printouts, their activity on the computer tool seemed more focused. They used the features on the tool to substantiate their preliminary analysis instead of to explore the structures that resulted from the use of the features. In addition, they focused more on features of the data sets such as clusters. I was also curious to see if the tools we had designed offered the teachers the means of analyzing data that fit with their initial, informal ways of analyzing the data.

As the teachers began their analyses, most of them initially calculated the mean of each set of data. They subsequently judged that measure to be inadequate for making the decision and proceeded to find ways to structure the data that supported their efforts at analysis. In this process, they used vertical lines drawn in the data to create cut-points and to capture the range of each set. As an example, one teacher noted that all of the coupes took over 55 feet to stop whereas four of the ten sedans were able to stop in less than 55 feet. Other teachers focused on the “bunched-up-ness” of the coupes and reasoned that a consistent braking distance was an important feature.

I found these ways of reasoning significant for two reasons. The first was that the ways of structuring that they were creating with the drawn lines paralleled the features that we had designed on the tool. This implied that the tool would be a useful resource in supporting their analysis. The second was that their ways of reasoning about the data were consistent with the methods that we saw emerge in the seventh graders’ activity. It therefore appeared that the conjectured learning trajectory could guide the mathematical development of the teachers.

As the teachers discussed the results of their analysis in whole-group setting, I introduced the first computer minitool as a resource for sharing their ways of structuring the data. I used a projection system to make the data sets visible to the group and as the teachers explained their analysis, I used the features on the minitool to complement their explanations. As an example, as the teachers talked about the “bunched-up-ness” of the data sets, I activated the range tool so that they could identify the extremes in each data set. The teachers found this support helpful and were easily able to use the features on the tool to mirror their earlier activity with drawn lines.

In the remainder of the session, the teachers analyzed a second set of data on the longevity of two brands of batteries. Their analyses were again focused on cut points and clusters. These ways of reasoning parallel what we found in analysis of the seventh graders’ activity. In particular, Cobb (1999) notes that in the seventh-grade classroom

> [t]he characteristics of data sets that emerged as significant in this discussion and in the subsequent classroom sessions in which the first minitool was used included the range and maximum and minimum values, the number of data points above or below a certain value or within a specified interval (p. 17).

The significant difference between the seventh-grade students and the teachers was that the teachers were able to talk about the number of data points above or below a cut point in terms of percentages (or ratios) of the whole. Further, they could reason probabilistically such as arguing that *you have a 30% chance of getting a bad battery* with a Brand A. Their arguments therefore appeared to be multiplicative in nature.

A shift to the second minitool and a focus on the second mathematical practice — exploring qualitative characteristics of distributions — began with the introduction of the speed trap task. After a lengthy discussion of the data creation process, the teachers were shown printouts of data on the speeds of two sets of sixty cars (see Figure 2). The first were recorded on a busy highway on a Friday afternoon. The speeds were recorded on the first sixty cars to pass the data collection point. The second set of data was collected on a subsequent Friday afternoon after a speed trap had been put in place. The goal of the speed trap (e.g. issuing a large number of speeding tickets by ticketing anyone who exceeds the speed limit by even 1 mph) was to slow the traffic on a highway where numerous accidents typically occur. The task was to determine if the speed trap was effective in slowing traffic.

As the teachers worked on the printouts of the data, most of them created cut points at the speed limit and reasoned about the number of drivers exceeding the speed limit both before and after the speed trap. They used a range of strategies including ratios and percentages. Further, none of the teachers calculated the mean. One teacher focused on the shape of the two data sets and noted that at first it *looked like a Volkswagen Beetle* and then it *flattened out like a large Town Car.* I found this particularly significant because it was the first occasion where a teacher found a way to describe the shape of the distribution. I capitalized on her solution and recast it in terms of "hills." My reason for doing so was grounded in my earlier work in the seventh-grade classroom where a similar incident had occurred with this same data set. For the students, the notion of the hill was pivotal in shifting their reasoning from viewing data sets as collections of points to distributions (Cobb, 1999; McClain, Cobb, Gravemeijer, 2000; and McClain & Cobb, in press).

The significance of this shift became apparent in the next session when I introduced a task containing data sets with unequal numbers of data points. In the task, two sets of AIDS patients were enrolled in treatment protocols — a traditional treatment program with 186 patients and an experimental treatment program with 46 patients. T-cell counts were reported on all 232 patients (see Figure 3).

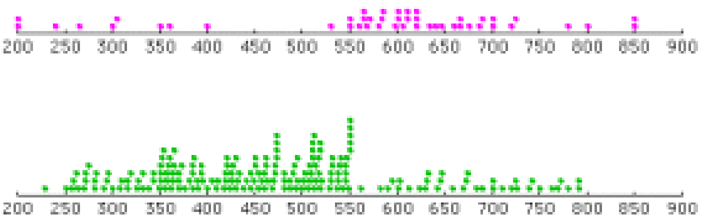

Figure 3. AIDS Data Displayed in the Second Minitool

As the teachers worked on their analysis, they initially noted that the clump, cluster, or hill of the data shifted between the two groups. This was a significant aspect of the distributions and one on which they focused as the continued their analysis using the second minitool. In their work at the computers, they found ways to characterize this shift including creating cut points and reasoning about the percentage of patients in each group with T-cell counts above the cut point. They, like the seventh grade students, noted that that cluster of T-cell counts in the traditional treatment program was below the cut point whereas the cluster of T-cell counts in the experimental was above. Arguments such as these indicate their ability to reason about characteristics of distributions.

A RETURN TO THE CONJECTURED LEARNING TRAJECTORY

Throughout the analysis, I have documented the development that occurred as the teachers and I together engaged in a series of tasks from the statistics instructional sequence. In doing so, I have pointed to the teachers' developing ways understanding their own and others' mathematical activity. In particular, I have shown a progression from exploring qualitative characteristics of collections of data points to exploring qualitative characteristics of distributions. However, I have noted a difference in the learning of the teachers compared to that of the seventh-grade students in that the teachers were able to reason multiplicatively from the outset.

As a result of these parallel developments, it appears that the underlying learning trajectory that guided the seventh-grade students' development offers the means of supporting teachers' development of understandings of statistical data analysis. This is an important finding in that the broad strokes of the significant mathematical issues that emerge from investigating ways to support teachers' understanding can be painted with the statistics instructional sequence, its tools and its underlying learning trajectory. This finding also has broader implications — that the broad base of research conducted in classrooms can feed forward to inform efforts at supporting the development of teachers' content knowledge.

ACKNOWLEDGEMENTS

I would like to thank the special group of teachers discussed in this paper for allowing me to participate with them in their professional endeavors. The research reported in this paper was supported by the National Science Foundation under grant REC-9814898 and by the Office of Educational Research and Improvement under grant R305A6007.

REFERENCES

Cobb, P. (1999). Individual and collective mathematical learning: The case of statistical data analysis. *Mathematical Thinking and Learning*, 5-44.

Cobb, P., Stephan, M., McClain, K., & Gravemeijer, K. (2001). Participating in classroom mathematical practices. *Journal for the Learning Sciences, 10(1&2),* 113-164.

Harel, G. & Confrey, J. (1994). *The development of multiplicative reasoning in the learning of mathematics*. NY: SUNY Press.

McClain, K., & Cobb, P. (in press). Supporting students' ability to reason about data. *Educational Studies in Mathematics*.

McClain, K., Cobb, P., & Gravemeijer, K. (2000). Supporting students' ways of reasoning about data. In M. Burke (Ed.), *Learning mathematics for a new century: 2001 Yearbook of the National Council of Teachers of Mathematics*. Reston, VA: NCTM.

Simon, M.A. (2000). Research on mathematics teacher development: The teacher development experiment. In A. Kelly and R. Lesh (Eds.), *Handbook of research design in mathematics and science education.* Hillsdale, NJ: LEA.

Thompson, P.W. (1994). The development of the concept of speed and its relationship to concepts of speed. In G. Harel and J. Confrey (Eds.), *The Development of Multiplicative Reasoning in the Learning of Mathematics* (pp. 179-234). NY: SUNY Press.

Thompson, P., & Saldanha, L. (2000). *To understand post counting numbers and operations*. White paper prepared for the National Council of Teachers of Mathematics.