IMPROVING DATA ANALYSIS

KAY McCLAIN, MAGGIE McGATHA, AND LYNN L. HODGE

THE NATIONAL COUNCIL OF TEACHERS of mathematics has been advocating the importance of effective communication in classrooms since the release of its *Standards* documents (NCTM 1989, 1991). This emphasis is echoed in Ricarhds's (1991) description of an inquiry classroom (see also, e.g. Ball [1993]; Cobb, Wood and Yackel [1991]; Lampert [1990]). In this setting, the teacher's role is to guide the negotiation of classroom norms to enable the teacher and student *together* to engage in meaningful mathematical discussions, which include asking questions, solving problems, posing conjectures, and formulating and critiquing mathematical arguments. And increased emphasis on communication in the mathematics classroom allows students the opportunity to discuss and validate mathematical ideas and to make and evaluate conjectures and arguments.

Computer tools would be essential

We find it particularly important to focus on communication when planning to teach statistical data analysis. In recent years, statistics has received increased attention in many education reform documents (cf. NCTM [1989,1991], Shaughnessy [1992]). Such discussions center on the increasingly prominent role of statistical reasoning in work, community, and home activities. The task for the students, then, is to participate in both developing and critiquing data-based arguments.

This article describes a classroom in which students developed statistical understanding of exploratory data analysis through mathematical argument. The session is taken from an eighth-grade classroom in which we worked with students on a data-analysis unit. The students had also participated with us in a data-analysis unit during the fall semester of the previous year. Our overall goal was to develop two instructional sequences, one for seventh grade and one for eighth grade, that focused on proactively supporting middle school students' development of statistical reasoning. As part of our efforts, we viewed the development of mathematical argument as a fundamental aspect of the classroom environment. As a result, the goal for the students' development was twofold—to support their ability to reason statistically as they developed ways to create and justify analyses of data.

Instructional Sequence

AS WE DESIGNED THE INSTRUCTIONAL SEquences, we tried to identify the "big ideas" in statistics. We hoped that these themes could guide the design of the sequence and allow us to avoid creating a list of separate, loosely related topics that characterize middle school curricula. In indentifying the big ideas, we focused on the notion of distribution. This emphasis let us treat such concepts as mean, mode, and median and other ideas, including skewness and spread, as characteristics of distributions. Our focus also allowed us to view various conventional graphs, such as histograms and box-and-whisker plots, as different ways of structuring distributions. The instructional goal was to support students' ability to view data as distributed along a space of possible values instead of as a set of individual data points.

In developing the instructional sequences, we were also guided by the idea that computer tools would be essential in supporting our goals. Such tools would let students organize, describe, and compare large sets of

---

KAY McCLAIN, kay.mcclain@vanderbilt.edu, MAGGIE McGATHA, mmcgatha@aol.com, and LYNN HODGE, lynn.l.hodge@vanderbilt.edu, are colleagues in teaching preservice courses at Vanderbilt University, Nashville, TN 37203. McClain is also involved in collaborative efforts with middle school teachers in the Nashville School System. Hodge is interested in diversity and equity issues in mathematics education. *The authors would also like to thank the students at Meigs Middle School. We would also like to acknowledge Paul Cobb's and Koeno Gravemeijer's roles in developing the computer tools and instructional sequence used. The research reported in this article was supported by the National Science Foundation under grant no. REC-9814898 and by the Office of Educational Research and Improvement through the National Center under grant no. R305A60007. The opinions do not necessarily ereflect the views of either the Foundation or the Center.*

data. We also acknowledged the importance of students' understanding how to reason with graphical representations, such as histograms and box-and-whisker plots, without concentrating instruction on the procedures for constructing these graphs. The computer tools that we designed offered a range of ways to structure data that supported students' ability to reason statistically while building toward conventional graphs.

The classroom session described here used a computer tool that displays data along a horizontal axis as in an axis plot. The tool can display two data sets simultaneously, such as those in **figure 1**, allowing for a comparative analysis. The tool also includes a number of op-

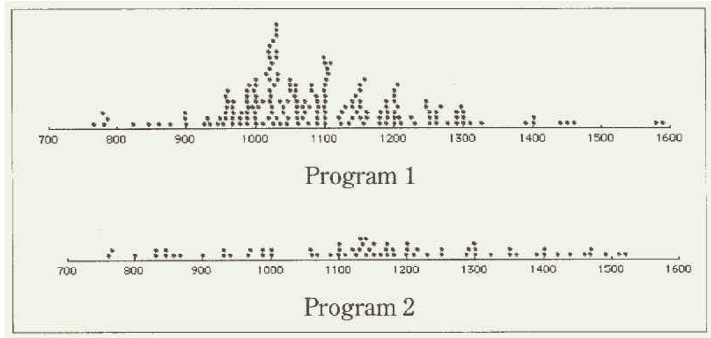

Fig. 1 SAT preparation program data

Photo Here

tions for structuring and analyzing data, some of which are precursors to conventional methods. For example, students in our classroom session used two options: (1) partitioning the data into four equal groups so that each group contains one-fourth of the data, as a precursor to creating the box-and-whisker plot; and (2) partitioning data into two equal groups so that each group contains one-half the data. These two options, in addition to the role of the computer tool, are described in greater detail in the later discussion of the class session.

In planning the instructional sequence, we reasoned that students needed to develop arguments based on the reasons for generating the data. They also needed to develop ways to analyze and summarize the data to substantiate their arguments. We anticipated that these goals could be achieved by developing instructional tasks that involved describing a data set or analyzing two or more data sets to make a decision.

Students were then asked to analyze SAT preparation course data

Classroom Session

THE CLASSROOM ROUTINE FOR OUR INSTRUCTIONAL sequences began with introducing a task by talking through the data-creation process. If students did not collect the data themselves, it was still important for them to discuss the reasoning behind the data-collection activity. Students would typically make conjectures and offer suggestions about the information necessary to make informed decisions. They would discuss what data they needed and how they might collect those data. These discussions were essential in grounding the students' recommendations or arguments in the context of the data-analysis activity. After this discussion, the students worked in pairs at their computer to conduct their analyses, writing justifications for recommendations and developing graphs of the data to support their recommendations. These justifications formed the basis for subsequent whole-class discussions. A projection system allowed students to display their data sets as they discussed their ways of organizing the data to support their decisions.

Whole-class discussions often involved critiquing these data-based arguments to emphasize the importance of designing adequate reports that were justified in terms of the task situation. Many discussions centered on developing adequate arguments and graphs to substantiate students' recommendations. These types of discussions were initially difficult for the students because of their lack of experience in engaging in mathematical argument. They had trouble distancing themselves from the classroom and understanding what someone outside their own investigations would need to know. As a result, their explanations often lacked detail; however, as students participated in such discussions more often, they began to support and articulate their arguments more effectively.

One of the task situations posed during the eighth-grade sequence involved members of the parent-teacher organization (PTO) in gathering data about two SAT preparation courses so as to recommend the more beneficial course. Students shared information about such programs and discussed why they might enroll. They then generated a list of information needed to make an informed decision about a program's success and discussed how to gather these data. Students were then asked to analyze data from two programs in which the SAT scores of the participants at he completion of the programs were shown on axis plots (see **fig. 1**). Students were asked to use their analyses to provide information that the PTO could use in formulating its recommendation. In their preliminary discussions, students reasoned that the "samples" taken from the two courses should have students of equal ability for the task to be reasonable; they operated under that assumption as they conducted their analyses and developed their arguments.

After the initial discussions, student pairs conducted their analyses using the computers. At the end of class, they were asked to turn in reports based on their analyses. The teacher reviewed these reports based on their analyses. The teacher reviewed these reports to decide how to orchestrate the whole-class discussion for the following day. To help support students' mathematical development and to assist them in formulating coherent arguments, the teacher began with one report that had the potential to foster a discussion about the adequacy of reports and graphs. The report selected was based on the notion of consistency, arguing that program 1 was better because the scores of its participants were more consistent. The report noted that most of the scores in program 1 fell in a range of about 300 points, whereas the scores in program 2 fell across a much wider range.

After the report was read in class, Barry made the following observation:

Barry. I'm not trying to put this way down, but I don't think it's a very good way to do on this particular thing because on this, we're not trying to pick out which one is more consistent–we want the highest grade possible...you'd rather have a higher chance of getting a higher grade. That's what you're looking for, not consistency.

Teacher. So you're saying, in this particular case, for this particular problem, you're not sure that consistency would be what you would want to know about.

Barry. Correct.

In earlier analyses, the notion of consistency had emerged from the students' investigations. They had reasoned that in some instances, being able to predict what might happen was more important than having a "better score." This notion first emerged in analyzing braking distances of different cars. The students reasoned that having a car that might stop quickly one time but slowly another time was less preferable than having a car that had a consistent breaking distance. The task context provided a basis for their argument. In this task, however, Barry was challenging the notion of consistency as sufficient to justify a recommendation. For him, consistency was not as important as the "chance of getting a higher grade."

The teacher reproduced the graph from the report on the whiteboard (see **fig. 2**). The two students had drawn a line to show the shape of the data. Other students began to discuss the adequacy of the graph. One of the students who had worked on the report attempted to explain the graph by arguing that the data points were "under the line," and for program 1, "there's a ton right in the middle" but for program 2, "they're very spread out." He went on to add that with program 2, "you have a chance of getting really low or a chance of getting really high."

Next, Marcia offered her opinion about the adequacy of the report: "I think that way would work, but you have to make a lot of choices. You have to decide whether you want to go with one that has the higher score but also the lowest score, or if you want to go with the one that is more consistent." The other students agreed that it appeared from the graph that program 2 offers a greater chance of a higher score but also the lowest score. Program 1 offers more chance of having a score "in the middle." Because neither of these results gives information about which program is *better*, however, the graph and analysis offered what the students ultimately agreed was an "inadequate" way of making a recommendation.

At this point, the teacher asked Sam and Robert to share their recommendation. In monitoring their activity the previous day, the teacher saw that Sam and Robert had first looked at the data sets partitioned into four equal groups but had reasoned that the top 75 percent of each data set was in the same range, so structuring the data in that way was not useful. They then partitioned the data into two equal groups and found that the top 50 percent of program 2 was in a higher range than the top 50 percent of program 1. Sam and Robert used this justification for selecting program 2. After they explained their analysis, the teacher restated their recommendation and highlighted their justification by marking the data sets projected on the whiteboard as shown in **figure 3**.

Teacher. Their way was to do two equal groups. They found that way helpful because the upper 50 percent of this group [pointing to data on program 2] was in a range that was higher than the upper 50 percent of this group [pointing to data on program 1].

[Marcia raises her hand.]

Teacher. Marcia, did you want to raise an issue about their way?

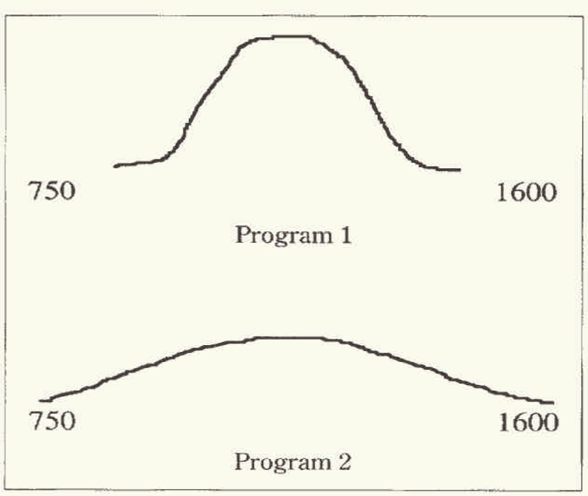

Fig. 2 Student graph of SAT preparation program data

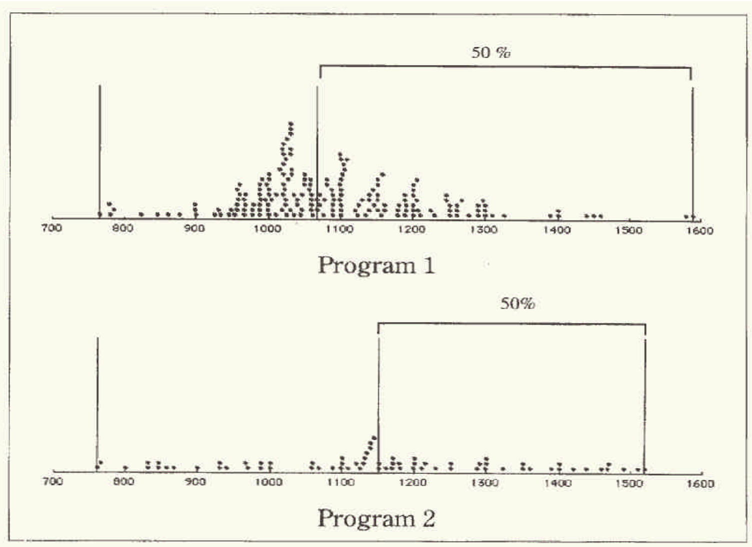

Fig. 3 SAT preparation program data organized in two equal groups

Photo Here

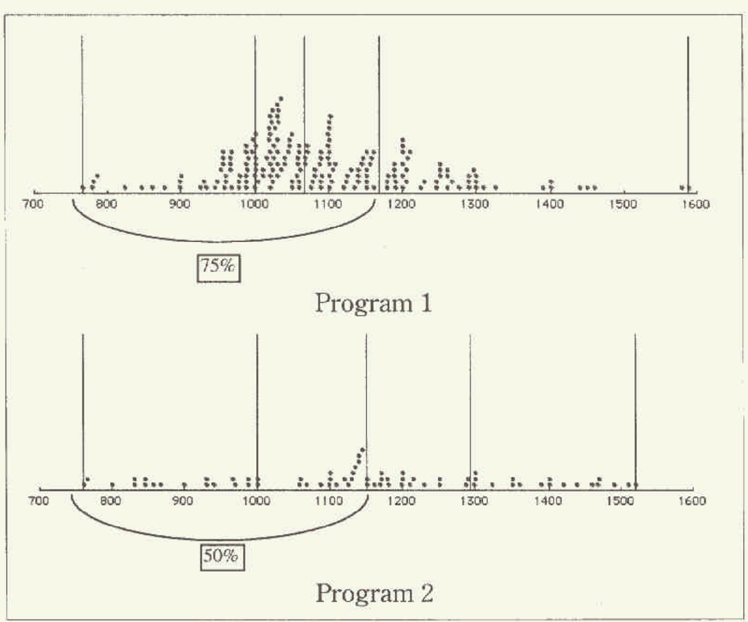

Fig. 4 SAT preparation program data organized in four equal groups

Marcia. The way they had it with four equal groups, he said he didn't get anything out of the four equal groups....but if you look at it this way [going to the board and drawing on the diagram, as shown in fig. 4] you have 50 percent up here and 75 percent right here. I'm saying look at it the other way. Instead of using the 75 percent where they are the same, flip it over and do it where they are different. That's what was bothering me when he said he didn't get anything out of this.

Marcia's reasoning used the data as they would be structured in a box-and-whisker plot. She justified why program 2 should be recommended over program 1, by reasoning about the proportions of each set of data that fell in a certain range. From Marcia's point of view, the lower 75 percent of program 1 was in the same range as only the lower 50 percent of program 2; therefore, program 2 had the better results. As Marcia spoke, the teacher drew on the whiteboard the ranges shown in **figure 5**. When Marcia finished, other students in the class agreed with her analysis. Their comments included, "Oh, wow!" "Yeah, I see!" "Oh, now I see." "I got what she is saying." Unfortunately, the bell indicating that class was over ended the discussion.

Concluding Remarks

IN THIS CLASS SESSION, THE DISCUSSION FOcused not only on the results of the students' analyses but also on what counted as an acceptable explanation. This type of discussion supported the students' development of strong mathematical arguments. As students discussed their justifications, they also developed ways of reasoning to support their analyses. The students came to understand that merely showing a graph and making a recommendation are insufficient. The recommendation must be justified in the context of the data.

In addition, the students developed language to use in formulating arguments, as evidenced in the discussion of consistency. This significant shift required the students both to make a recommendation and to justify the recommendation on the basis of their analysis. The discussions emphasized the ways of structuring the data to support the argument. It is interesting to note that students often changed their initial judgments on the basis of the whole-class discussions–their reasoning was constantly being challenged and modified in light of others' arguments.

In discussing the expectations and obligations of this class, many students remarked that they felt as if they were "all sort of learning together." Other comments showed the importance of discussions; one student noted, “[I] liked the discussions because I can share ideas and get some back. Even when people disagree, it makes you think.” In this setting, the students agreed that “everybody is smart.” They commented, “Most of the time, teachers don’t care about what you think, but in here we get to share our opinions–that’s the whole discussion.” For these students, learning mathematics was not an individual effort but depended on the contributions of everyone in the class. Their work was not complete after they had conducted their analyses. They then had to work to understand others’ analyses and question the recommendations made on the basis of those analyses. Students also viewed the role of the teacher differently: “It’s like the teacher teaches you, but you also teach the teacher.” In this way, the students and the teacher contributed to establishing classroom norms for explanations and justifications.

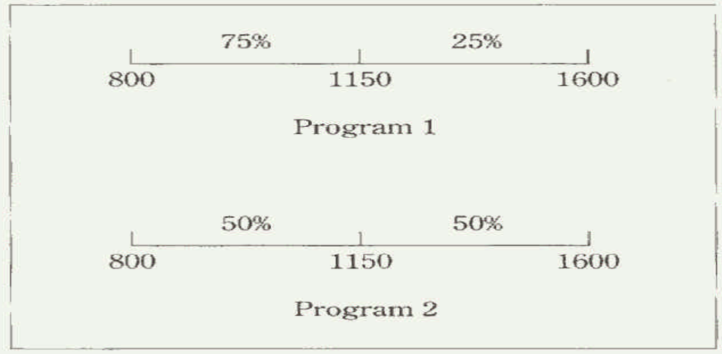

Fig. 5 Marcia’s view of the SAT preparation program data

Engaging students in productive mathematical discussions often involves changing the students’ ideas about what it means to know and do mathematics–it involves engaging them in the process of change. Instead of the teacher’s assuming the primary responsibility for classroom discussions, students play an essential role in this process. This statement is not to imply that the teacher is no longer responsible for the mathematics. The teacher’s role is expanded to include facilitating discussions in which students explain and justify their mathematics. This outcome is often accomplished when the teacher is aware of students’ different solution methods and can orchestrate a discussion in which fundamental mathematical ideas and ways of reasoning form the basis for argument. In this way, students’ communicating their mathematics becomes the basis for learning that is shared by all members of the classroom community.

References

Ball, Deborah. “With an Eye on the Mathematical Horizon: Dilemna of Teaching Elementary School Mathematics.” *Elementary School Journal* 93 (1993): 373-98.

Cobb, Paul, Terry Wood, and E. Yackel. “A Constructivist Approach to Second Grade Mathematics.” In *Radical Constructivism in Mathematics Education*, edited by Ernst von Glaserfeld, 157-76. Dordrecht, Netherlands: Kluwer, 1991.

Lampert, Magdalene. “When the Problem Is Not the Question and the Solution Is Not the Answer: Mathematical Knowing and Teaching.” *American Educational Research Journal* 27 (1990): 29-63.

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

_____. *Professional Standards for Teaching Mathematics*. Reston, Va.: NCTM, 1991.

Richards, John. “Mathematical Discussion.” In *Radical Constructivism in Mathematics Education*, edited by Ernst von Glaserfeld, 13-51. Dordrecht, Netherlands: Kluwer, 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, 465-94. New York: Macmillan, 1992.

HOT TOPICS !

Mathematics Teaching in the Middle School (MTMS) is interested in re-ceiving manuscripts on the following topics. Over 70 percent of recently surveyed members indicated that these topics most concern them:

- Algebraic thinking and algebra for all
- Mathematical reasoning
- Problem solving

If you have interesting ideas or classroom-tested approaches surrounding the big ideas for any of these three topics, please share them by writing a manuscript for the journal.

Submit five copies of such a manuscript to *Mathematics Teaching in the Middle School*, NCTM, 1906 Association Drive, Reston, VA 20191-9988. If you would like hints on how to prepare a manuscript for publication, request our “*MTMS* Writer’s Packet” from NCTM at the same address or send an e-mail to Kathleen Lay at klay@nctm.org.