A broad consensus exists that “students need to know about data analysis and related aspects of probability in order to reason statistically—skills necessary to becoming informed citizens and intelligent consumers” (NCTM, 2000). In light of these recommendations for greater emphasis on data analysis throughout the grades (AAAS, 1993; NRC, 2000), researchers at the National Center for Improving Student Learning and Achievement in Mathematics and Science (NCISLA) examined ways teachers can make statistics accessible to all students. The research studies highlighted here feature the ways that middle school students can come to understand statistics through carefully designed instruction and teacher professional development.

The approach taken by the NCISLA research team led by Paul Cobb and Kay McClain capitalized on the notion of distribution as a key concept in statistics and made generalization and justification an explicit focus of instruction. The research suggests that in order to support students' development of more advanced statistical reasoning, traditional middle school instructional strategies need to be reconsidered. The findings suggest ways teachers can develop lessons that are better suited to students' ways of thinking and learning. Through professional development seminars in two cities, Cobb, McClain, and their colleagues examined ways that alternative instructional designs can be implemented and sustained with only minimal external resources.

Students typically approached problems in a procedural manner, without considering the question or issue at hand. For example, students often calculated means to compare data sets even for cases in which the analysis of differences in range or variability of data were more appropriate.

To re-orient students' beliefs about what it means to do statistics in ways that resemble those of professional statisticians, the research team developed two instructional sequences—one for the seventh grade and one for the eighth grade—and computer tools, which they then tested in an experimental classroom and revised based on their analysis of student learning and understanding.

The researchers were particularly interested in developing instructional sequences that fostered students'—

- ability to analyze data in increasingly sophisticated ways.
- understanding of statistical inference.
- ability to design procedures for generating sound data.

The 2–year study involved 29 racially diverse students in the seventh grade and 11 of the same students as they continued into the eighth grade. The research team

---

1 Generalizations are general mathematical assertions concerning the structure, properties, or relationships that underlie key mathematical concepts. Students provide justification of their mathematical generalizations and analysis by developing mathematical arguments that are shared, discussed, and debated in class, in order to further student thinking and learning.

2 One team, located in the southern part of the United States, included 12 teachers; the other team, located in the western part of the United States, included 14 teachers.

3 The research team, led by Cobb and McClain, included Koeno Gravemeijer, Jose Cortina, Lynn Hedge, Maggie McGarr, Nira Shivani, and Carrie Tzou. Cliff Reid and Ernst Ruckel served as long-term consultants.

4 Although the intention was to continue with the 29 original seventh-grade students, the requirement that the cooperating school district follow a relatively prescriptive curriculum meant the researchers could not conduct the experiment during the regularly scheduled eighth-grade class. Thus, the students participating in the eighth-grade sequence volunteered to participate in an afternoon activity period. Of the 11 students, 7 were African American, there were White, and one was Asian American.
In the seventh–grade sequence, the team was able to reduce instructional time by two–thirds. In follow–up trials using the seventh–grade sequence, the team was able to reduce instructional time by two–thirds.

**Instructional Activities and Classroom Norms**

In the beginning of the sequence, a majority of the lessons engaged students in exploratory data analysis—an important activity that lays the groundwork for statistical inference. In exploratory data analysis, students draw informal conclusions by analyzing meaningful patterns in specific sets of data. Statistical inference takes the analysis of such patterns a step further by drawing conclusions about the population. Based on a data sample, statisticians assess the likelihood that those patterns reflect trends in a larger population, relying on notions of sampling distributions, confidence intervals, and significance tests to make probabilistic statements.

To promote student reasoning in these statistical activities, the researchers proposed using the notion of distribution as the overarching idea around which instruction would be organized. Specially designed computer tools allowed students to reorganize data in order to identify patterns in the way that the data were distributed. (See page 3 for more on Computer Tools That Supported Students’ Learning.) Students used the computer tools to represent their models and generalizations, which in turn provided a context and focus for class discussions (McClain & Cobb, 2001).

The investigations typically required students to make a recommendation based on a comparison of two sets of data, for example data on patients who received two different medical treatments or on the braking distances of samples of two different models of cars. Students’ investigations of a particular problem set began with a discussion about the data–generation process. This was important because the students did not collect the data themselves and the researchers wanted to ensure that students’ analyses remained grounded in the problem or issue they were investigating. In these, often quite lengthy discussions, the teacher and students together—

- identified the particular phenomenon under investigation (e.g., two alternative treatments for AIDS patients).
- clarified why it was important to address the problem (e.g., relevance of determining the most effective medical treatment).
- identified relevant aspects that should be measured (e.g., T–cell counts).
- considered how those aspects might be measured (e.g., by taking blood samples).

A majority of classroom time was devoted to exploratory data analysis. Working first in pairs and then through whole–class discussions, students used the computer tools to structure data and present their analyses. Students were also required to write individual reports about their data analysis, outlining data–based recommendations for a target audience (e.g., the head medical officer of a hospital).

The researchers found that the computer tools served an important function in supporting student learning, as did the negotiation of norms or standards for what counted as an acceptable statistical argument. The reports required students to make explicit their reasoning and statistical arguments and, thus, also supported students’ learning of reasoned argumentation in statistics.

**Reorienting Student Thinking About Statistics.** Cobb and McClain reasoned that just as a proficient analyst searches for trends, patterns, and anomalies in relation to a research question, students need to conduct their data analysis in the context of problems they consider realistic and legitimate. Cobb and McClain found that by talking through the data collection process, students became engaged in the problem at hand and, ultimately, in finding ways to use statistical analysis to address the problem situation.

The research team found that if students had some familiarity with the phenomenon to be investigated, had an opportunity to talk through the data collection process, and recognized a broader purpose to the investigation, the teacher could foster and support students’ pragmatic interest in analyzing the data. Initially, students recounted personal stories that related to the problem scenario, rather than focusing on generating and analyzing data to come to practical decisions or judgments.

By the second week of the sequence, students were taking a more pragmatic approach to the investigation and increasingly posed questions about forms of data that would be helpful in their analysis.

As they worked, the teacher and students established norms for discussing their analysis and making statistical arguments: Students were expected to explain and justify their reasoning. In classroom discussions, students compared different analyses and clarified statistical issues, such as the most appropriate way to partition the data for comparisons. McClain and Cobb (2001) distinguished this form of conceptual discourse, in which students explain the rationale for their analysis, from calculational discourse, in which students merely explain how they analyze the data but do not explain why they choose a particular analytical method for addressing a problem.

These discussions, in which students explained the rationale for their analytical methods, gave the researchers important insights into their reasoning. The teacher was able to capitalize on this reasoning during classroom activities and discussions using the computer tools.

**Using Computer Tools to Enhance Learning.** In the seventh–grade sequence, students made the transition from analyzing data based on the absolute frequency of occurrences in particular intervals (e.g., 15 cars going faster than the speed limit and 30 going slower) to analysis based on the relative frequency (e.g., one–third of the cars going faster than the speed limit). Making this transition from additive to multiplicative reasoning proved important for students, especially in analyzing unequal data sets continued on page 4...
Computer Tools That Supported Students’ Learning

Statistical computer tools were designed to support development of students’ statistical thinking. The tools provide a basis for learning to interpret conventional representations used in statistics, such as histograms and box-and-whiskers plots. The computer tools are available on the CD-ROM included with the NCTM publication, *Navigating Through Data Analysis in Grades 6-8* (Bright, Brewer, McClain, & Mooney, 2003). Students can use the computer tools to—

- investigate trends and patterns in univariate data by ordering, partitioning, or otherwise organizing data points. **Tool 1** represents individual data points as separate bars and allows students to compare the number of data points within any intervals they choose.

- analyze increasingly complex univariate data sets in ways that more closely resemble standard ways in which professional statisticians structure data. **Tool 2** allows students to represent data as dots on an axis and then partition that data into groups of a specific size, with a specified interval width (a precursor to histograms), or two or four equal groups (a precursor to box plots). (See **FIGURE 1**.)

- identify trends or patterns in data distribution of bivariate data sets. **Tool 3** provides students a variety of ways to investigate relations between two measures in bivariate data sets that are represented as scatter plots. (See **FIGURE 2**.)

The following example illustrates how the teacher used Tool 2 to support students’ learning as they investigated the effectiveness of two AIDS treatments. The number of patients in the two data sets differed significantly, challenging students to compare unequal data sets.

Using Tool 2 to plot the T-cell counts for the two treatments on two different axes, a number of the students decided to compare the data sets in terms of the number of data points falling above and below a particular data point (see **FIGURE 1a**). One group of students presented their analysis using this representation, concluding that the new treatment was more effective than the old treatment and noting that “the old treatment had 56 patients above 550, and the new one 37.” However, a fellow student pointed out that this argument was confusing because the students reported the number rather than the proportion of data points above 550 (i.e., absolute rather than relative frequency).

The analysis presented by another group of students, who used Tool 2 to divide the data into four equal groups, built off students’ decision to use percentages rather than absolute numbers to make the comparison (see **FIGURE 1b**). From this analysis, the students concluded that the new treatment was better than the old treatment “because the three groups for the new [treatment] group are above 525 and on the old they are below.” The data points in this graph are hidden so that students can focus on the ways the data are distributed when divided into intervals, each of which represent 25% of the data for a given set. This representation is a precursor to the box-and-whisker plot.

**FIGURE 1**
Univariate data organized using Tool 2 to partition data (a) at any given value and (b) into four equal groups, with individualized data points hidden.

---

*For more on the computer tools, also see McLain & Cobb, 2001; Cobb, McClain, & Gravemeijer, 2003.*
FIGURE 2
Bivariate data organized using the Grids option in computer Tool 3.
(comparing data sets with different numbers of data points), in which a focus on absolute frequencies can lead to erroneous conclusions. (See example described in Computer Tools That Supported Students’ Learning)

The teacher supported students’ transition from additive to multiplicative reasoning by encouraging them to think about a data set as a whole, rather than as a collection of individual data points.

By the end of the seventh–grade sequence, all 29 students came to reason about univariate data in terms of qualitative proportions. Nineteen out of 29 also could use the graphs representing distributions of the data sets (similar to histograms and box plots) to make sound data–based arguments, even when the data sets were unequal.

Importantly, an assessment of students’ data analysis at the beginning of the eighth–grade sequence indicated that none of the continuing students had regressed during the 9–month gap between the experiments.

In the eighth–grade experiment, distribution proved to be a crucial concept for students to understand when analyzing bivariate data. The research team wanted the students to be able to interpret a bivariate data set as distributed in a two–dimensional space that represented the change in one variable’s distribution when other variables changed (i.e., a distribution of univariate distributions). This focus on the distribution of bivariate data encompasses and deepens a concern about direction and strength of relationships between statistical variables.

The research team structured the activities in the eighth–grade sequence so that students could learn about key concepts, such as covariation. However, their analysis of student work suggested a need to establish student understanding of a distribution of univariate distributions before students could interpret scatterplots of bivariate data in terms of the line (representing the conjectured relationship of covariation) about which the data are distributed.

The research team found that when they used the grids option on Tool 3 (see Computer Tools That Supported Students’ Learning, FIGURE 2), students more easily interpreted bivariate data in terms of the shape of the different stacks (or slices) of univariate data. As a result of these discussions focused on shape, concepts such as “majority” and “median” took on new meanings that supported an understanding of bivariate distribution.

continued on page 5 . . .
Student understanding of bivariate data as distributed reflects a progression to a more sophisticated approach to statistical inference. Conceptually, understanding bivariate data sets as distributions is important for assessing which aspects of a data set are relatively stable. For example, if an analyst wants to repeat an experiment and collect new data, which characteristics does she examine to determine if she has a good representation of the population? Recognizing that the median of a univariate data set is relatively stable in comparison with extreme values is the first step toward statistical inference (i.e., making an inference about the population based on a given sample). Cobb and McClain proposed that this orientation to statistics can provide a starting point for understanding sampling distributions, a concept that has been problematic for most college students.

**Implications: Supporting Students’ Learning of Statistics**

Cobb and McClain’s research reveals ways that middle school students’ learning of statistics can be better supported and suggests ways to design instruction and pursue instructional reform in middle school mathematics education. The research shows that through the seventh- and eighth-grade sequences, students learned to—

- develop sound designs for generating data that involve both issues of sampling and of controlling extraneous variables.
- effectively develop and communicate statistical arguments.
- make recommendations that were justified in the context of the data.
- structure the data in different ways to support their arguments and to reason statistically.
- take a more active role in statistical analysis.

As a result, students developed a more realistic perspective about what it means to do statistics.

Although NCTM’s *Principles and Standards for School Mathematics* (NCTM, 2000) does not propose introduction of the topic of statistical covariation until high school, Cobb and McClain found that middle school students were capable of reasoning about data in more sophisticated ways than is generally thought possible. Cobb and McClain’s findings constitute the first apparent documented instance of middle school students’ reasoning about bivariate data in such a sophisticated manner.

**Next Steps**

The instructional sequence and computer tools developed by Cobb and McClain now serve as resources for middle school teachers, who are adapting them to achieve their instructional goals (Cobb, McClain, de Silva Lamberg, & Dean, 2003; Cobb & McClain, in press). As previous studies have shown, implementation of an innovation is less likely if teachers do not have an understanding of the underlying ideas that help them adapt the innovation to their own settings (Ball, 1989, 1993; Bransford, Brown, & Cocking, 2000; Grossman, 1990; Sowder & Schappelle, 1995). Cobb and McClain’s research indicates that teacher professional development that is premised on students’ learning trajectories can help teachers design goal-directed instruction that builds on and advances students’ ways of thinking.

(For more detail about reforming instruction and teacher professional development see Teacher Practice insert and Professional Development insert.)
Based on analyses of students’ reasoning and learning trajectories (see page 4, Why Focus on Students’ Learning Trajectories?), Cobb and McClain’s research team developed computer tools and instructional sequences that support students’ learning of statistics. These tools were subsequently used in researcher-led professional development in two urban areas with 26 teachers.

One challenge of planning a lesson based on students’ learning trajectories is anticipating what students will understand and the ways students will approach the instructional activities. Through their research and professional development experience, Cobb and McClain found that teachers can design lessons responsively and effectively along three dimensions—

- instructional activities.
- tools and resources students use for analysis.
- classroom discourse.

Through these studies, the teachers and researchers identified the following issues as critical in designing effective instruction—

**Fostering Student Interest in Statistics.** Cultivating student interest was considered a crucial factor in ensuring all students had the opportunity to learn statistical ideas. The extent to which students and their families see learning mathematics paying off (in terms of future educational and economic opportunities) varies as a consequence of family history, race or ethnic history, and class structure. By making the development of students’ interest in mathematics an explicit goal of the instructional sequence, participating teachers were better able to provide all students with a reason to engage in statistics.

Specifically, the researchers found the types of activities selected for investigation mattered for cultivating student interest. Students were less likely to engage in analysis of the data if they had little knowledge of the phenomenon to be investigated. To develop students’ pragmatic interest in the investigation and to familiarize students about a phenomenon, teachers led discussions about the data-generation process.

**Allowing for Contrasting Methods of Analysis.** Fruitful statistical discussions occurred when students shared different approaches to analyzing data. The initial data sets were designed to allow for contrasting methods of analysis that resulted in different recommendations or pragmatic arguments. Later, the teachers introduced data sets that purposefully had a significantly different number of data points to lead students to make comparisons in relative rather than in absolute terms. Classroom discussions were then organized, and specific tools were incorporated, in order to reach other instructional goals.

By circulating around the room during group work, the teacher was able to assess student thinking. He or she could then incorporate the students’ contrasting solutions into classroom discussions in order to highlight statistically significant issues. In student interviews, one student expressed this aspect of the lesson as her favorite because “a lot of times people talked about ways that I hadn’t thought of. I liked trying to figure out their way or seeing if their thing is better.”

**Structuring Class Discussion Around Students’ Analyses of Data.** Cobb and McClain designed the computer tools as resources for teachers to use to achieve their instructional goals by capitalizing on their students’ reasoning. In each investigation, which often spanned two or more class sessions, students engaged in a whole-class discussion of the data-generation process, individual or small-group activity in which the students usually analyzed the data at computers, and a whole-class discussion of their analyses, often with the use of a computer projection system to display the computer tools. Students also learned to communicate their critical reasoning and to address the broader issues of the investigation through writing reports.

The computer tools and classroom discourse norms played important roles in all of these activities. Classroom norms helped delineate what was considered appropriate for students to contribute to a statistical discussion. Teachers and students could negotiate what Yackel and Cobb (1996) described as sociomathematical norms. Such norms included delineating what counted as a contrasting mathematical solution, what counted as a sophisticated solution, what was considered an acceptable line of reasoning and what counted as an efficient solution. Ideally, students justified their statistical reasoning by explaining how the method they used to structure the data resulted in insights into the pragmatic issues relevant to the investigation. The classroom norms evolved as students reorganized their own thinking based on classmates’ contributions and growing statistical knowledge.

**Capitalizing on Student Contributions.** In order to focus discussions on statistical issues, the teacher often needed to ask students to clarify their statements or expand on comments made by other students. For example, when one student referred to the “majority of numbers,” the teacher asked what he meant by “majority.” Clarifying and establishing student understanding of terminology was a critical part of refining student thinking and engaging in statistical discussions.

---

6All the teachers had at least three years of experience in teaching mathematics in a middle school.

7The computer tools are available on the CD-ROM included with the NCTM publication, Navigating Through Data Analysis in Grades 6–8 (Bright, Brewer, McClain, & Mooney, 2003).

The tools are also available online at http://peabody.vanderbilt.edu/dept/tandl/mted/MiniTools/MiniTools.html.
**Designing Statistics for Middle School**

**PROFESSIONAL DEVELOPMENT IMPLICATIONS**

Teacher professional development is central to an instructional approach to statistics that builds on students’ learning trajectories. (See page 4, Why Focus on Students’ Learning Trajectories?) Working over the long-term in two teaching communities (12 and 14 teachers respectively), Paul Cobb and Kay McClain examined ways to develop and sustain professional teaching communities. In these teaching communities, the researchers and teachers implemented a professional development program that enabled teachers to—

- identify student learning trajectories.
- establish instructional goals in light of students’ learning trajectories.
- propose benchmarks to assess students’ progress towards instructional goals.
- examine videotaped classroom sessions to investigate student learning and to evaluate the effectiveness of sequences in relation to the instructional goals.
- plan instruction by adapting sequences to fit local instructional situations and learning goals.
- engage in and facilitate statistical analysis using the computer tools.
- generate evidence of students’ mathematical reasoning to make the complexity of teaching and learning visible to school leaders and the educational community.
- collaborate on assessment tasks that could be used across contexts to examine student thinking.

Based on their research and professional development experience, Cobb and McClain worked to establish teaching communities that build teachers’ content knowledge, pedagogical skill, and collaborative strength. Teachers commented that the professional development required significant commitment on the part of teachers, support from administrators was essential. The researchers observed that if districts become focused on a narrow range of tasks assessed through tests, teachers might likewise feel compelled to narrow their instructional focus and professional development, consequentially missing opportunities to recognize and capitalize on their students’ thinking and reasoning. Teachers and administrators can foster an environment that supports the development of a professional teaching community—one that builds teachers’ content knowledge and pedagogical skills—by collaboratively defining a vision that recognizes students’ capacities to engage in statistical reasoning and then invest in professional development to that end. (See Gamoran et al., 2003, for more detail about organizational supports for instructional change.)

**Establishing “Buy-in.”** Cobb and McClain found that teachers were willing to invest time and energy in changing their instructional practices when they were provided concrete evidence of students’ misconceptions about statistics (often grounded in traditional instructional approaches that focused on calculating but not understanding the median, mean, mode, and range). Videos of classroom sessions, during which students’ reasoning and misconceptions were evident, provided teachers a framework for analyzing their own practice.

**Allocating Time and Resources.** In team meetings, middle school mathematics teachers served as resources for each other, sharing observations of students’ thinking and collaboratively planning or adapting lessons. Teachers analyzed students’ reasoning and solution strategies and discussed ways they could revise problems in order to correct common misunderstandings and advance student thinking. Classroom episodes documented by Cobb and McClain during visits to teachers’ classrooms were useful for analyzing instructional practices and students’ reasoning in these school-supported teacher seminars.

**Supporting Teachers’ Work on Improving Instructional Sequences.** In monthly seminars, teachers became actively involved in the process of improving and adapting the instructional sequences. A typical seminar often involved teachers in working through a lesson and then discussing the different ways that they analyzed the data. In subsequent seminars, teachers revisited the lesson and examined the ways it played out in his or her classroom, provided specific examples of students’ work, discussed students’ learning trajectories, and suggested ways to improve the lessons to further enhance students’ engagement and learning. This iterative process not only improved lessons but also strengthened teachers’ practice and content knowledge.

One teacher expressed this change: “I’m going through [the seminar] I had tools, for example, I’ve never seen a box and whiskers plot. It’s become my best friend. I can’t imagine looking at data—most data—without using a box-and-whiskers.”

**Additional Resources**


---

**NCISLA • Wisconsin Center for Education Research • School of Education, University of Wisconsin–Madison • 1025 West Johnson Street • Madison, Wisconsin 53706**

**Phone: (608) 263-7582 • Web site: http://www.wcer.wisc.edu/ncisla**

**WINTER 2004**
References


For More Information

This issue of *in Brief* is based on research by Paul Cobb, Kay McClain, and Keono Gravemeijer that is reported in a forthcoming book, *Understanding Mathematics and Science Matters*, edited by T. Romberg, T. Carpenter, and F. Dremock, and in other journal articles (see References). For additional publications by Paul Cobb, Kay McClain and Keono Gravemeijer see:


About in Brief

This publication and the research reported herein were supported under the Educational Research and Development Centers Program, FR/Award Number R305A960007, as administered by the Office of Educational Research and Improvement, U.S. Department of Education. The opinions expressed herein are those of the featured researchers and authors and do not necessarily reflect the views of the supporting agencies.

Permission to copy is not necessary. This publication is also available for free download at: http://www.wcer.wisc.edu/ncisla/publications.

About NCISLA

NCISLA was a university-based research center focusing on K–12 mathematics and science education. During 1995–2003, Center researchers collaborated with schools and teachers to create and study instructional approaches that support and improve student learning and understanding of mathematics and science. Through research and development, the Center identified new professional development models and ways that schools can better support teacher professional development and student learning.

NCISLA Director: Thomas P. Carpenter
NCISLA Associate Directors: Jim Stewart & Paul Cobb
NCISLA Communication Director: Susan Smetzer-Anderson
Writer: Valerie Farnsworth
Senior Editor: Fae Dremock
Designer: Linda Endlich, [IMDC], SOE, UW-Madison

A FINAL NOTE

This is the last *in Brief* to be published by the National Center for Improving Student Learning and Achievement in Mathematics and Science. Center researchers will continue to publish—in books, journals, and papers—their research and professional development findings.

For additional information about Center researchers’ work, we invite you to consult the extensive bibliography posted at www.wcer.wisc.edu/ncisla/publications.

A book, *Understanding Mathematics and Science Matters* (Romberg, Carpenter, & Dremock, Eds., in press), is forthcoming from Lawrence Erlbaum Associates. This book discusses the Center’s research on learning with understanding and provides a vision for reform in mathematics and science instruction. You may find it a useful resource along with the multimedia product you will receive this spring—courtesy of the Center—*Powerful Practices in Mathematics and Science*.

The collaborative work of Center researchers has provided important insights into mathematics and science instruction and teachers’ professional development. We were glad to have the opportunity to share it with you over the years and hope it has a positive impact in school and policy arenas.

Sincerely,

Thomas P. Carpenter
DIRECTOR, National Center for Improving Student Learning & Achievement in Mathematics & Science (1995-2004)