11

Designing a Data Analysis Tool for Learners

Clifford Konold
University of Massachusetts Amherst

In this chapter, I describe ideas underlying the design of a software tool we developed for middle school students (Konold & Miller, 2005). The tool—*TinkerPlots*—allows students to organize data to help them see patterns and trends in data much in the spirit of visualization tools such as *Data Desk* (DataDescription, Inc.). But we also intend teachers and curriculum designers to use it to help students build solid conceptual understandings of what statistics are and how we can use them.

Designers of educational software tools inevitably struggle with the issue of complexity. In general, a simple tool will minimize the time needed to learn it at the expense of range of applications. On the other hand, designing a tool to handle a wide range of applications risks overwhelming students. I contrast the decisions we made regarding complexity when we developed *DataScope* 15 years ago with those we recently made in designing *TinkerPlots*, and describe how our more recent tack has served to increase student engagement at the same time it helps them see critical connections among display types. More generally, I suggest that in the attempt to avoid overwhelming students, too many educational environments managed instead to underwhelm them and thus serve to stifle rather than foster learning.

Before looking at the issue of complexity, I describe more general considerations that influenced our decisions about the basic nature of *TinkerPlots*. These include views about (1) what statistics is and where the practice of statistics might be headed and (2) how to approach designing for student learning.

OVERARCHING DESIGN CONSIDERATIONS

The Growing Role of Statistics

How can we teach statistics so that students better understand it? This was the primary question that 25 years ago motivated me and my colleagues to begin researching the statistical reasoning of undergraduate students. Our assumption was that if we could better understand students' intuitive beliefs, we could design more effective instruction. We researched student reasoning regarding concepts fundamental to the introductory statistics course. These included the concept of probability (Konold, 1989; Konold, Pollatsek, Well, Lohmeier, & Lipson, 1993; randomness (Falk & Konold, 1997); sampling (Pollatsek, Konold, Well, & Lima, 1984); and the Law of Large Numbers (Well, Pollatsek, & Boyce, 1990).

We find ourselves today concerned with a different set of questions. These include:

- What are the core ideas in statistics and data analysis?
- What are the statistical capabilities that today's citizens need, and what will they need 25 years from now?
- How do we start early in young peoples' lives to develop these capabilities?

Three interrelated developments are largely responsible, I believe, for this change of research focus. First, there has been an expansion of our view of statistical practice, a difference often signaled by use of the term *data analysis* in place of *statistics*. Much of the credit for enlarging our vision of how to analyze data goes to John Tukey. His 1977 book, *Exploratory Data Analysis*, in which he advocated that we look past the ends of our inferential noses, was in many ways ahead of its time.

Second, the United States, as have many other countries, has committed itself to introducing data analysis to students beginning as early as the first grade (NCTM, 2000). The reason often given for starting data analysis early in the curriculum is the ubiquity of data and chance in our everyday and professional lives. The objective has become not to teach statistics to a few but to build a data literate citizenry. Given that we never really figured out how to teach statistics well to college undergraduates, this is a daunting, if laudable, undertaking.

Finally, the enhanced capabilities and widespread availability of the computer has spawned a new set of tools and techniques for detecting patterns in massive data sets. These methods, sometimes referred to as "data visualization" and "data mining," take advantage of what our eyes (or ears, Flowers & Hauer, 1992) do exceptionally well. The parody of the statistician as a "number cruncher" is dated. A more fitting term for the modern version might be *plot wringer.*

Because of the ubiquity of data and their critical role across multiple disciplines and institutions, formally trained statisticians are now a thin sliver of those who work with data. Jim Landwehr, who was a candidate for the 2005 President-Elect of the American Statistical Association, made this observation in his ballot statement (http://www.intelliscaninc.com/amstat_90.htm#s02):

> I believe that a statistical problem-solving approach is an important, ingrained component of today's economy and society and will continue to thrive. It is not so obvious to me, however, that the same could be said of "core statistics" as a discipline or "core statisticians" as employees. With our diversity of topics and interests and with their importance to society, we statisticians face the dangers of fragmentation. Statistics can and will be done by people with primary training in other disciplines and with job titles that don't sound anything like "statistician." This is fine and we could not stop it even if we wanted to.

The growing stores of data along with the perception that we now have tools that permit us to efficiently "mine" them is helping to shape a heightened sense of accountability. As patients we view it as our right and obligation to examine the long-term performance of hospitals and individual doctors before submitting ourselves to their care. We expect that their recommendations are based on looking at past success rates of therapies and procedures. And this is not the case just in medicine. We expect nearly all our institutions—government, education, financial, business—to monitor and improve their performance using data. Where data exist, none of us are immune. Among the reasons Red Sox officials gave for firing manager Gracy Little at the end of the 2003 baseball season was his "unwillingness to rely on statistical analysis in making managerial decisions" (Thamel, 2003, p. 3). Public education will likely slip the noose of the No Child Left Behind legislation, but not without putting in place a more reasonable set of expectations and ways of objectively monitoring them. The information age is fast spawning the age of accountability.

It is critical that we consider these trends as we design data analysis tools and curricula for students. Our current efforts at teaching young students about data and chance are still overly influenced by statistical methods and applications of 30 or 50 years ago. This is not to suggest that we lay all bets on guesses about where the field might be headed. But we do need to imagine what skills today's students will likely be using 10 and 25 years from now and for what purposes. At the same time, we need to work harder to understand what the core ideas in statistics are and how to recognize them and their precursors in the reasoning of the 10-year-old. We can assume that these underlying concepts (e.g., covariation and density) will be evolving more slowly than the various methods we might use to think about or represent them (e.g., scatter-plot displays and histograms).

Bottom-Up Versus Top-Down Development

During the past few years, mathematics and science educators have been investigating how students reason about mathematical and scientific concepts and applying what they learn to improve education. This has led to methods of bottom-up instructional design, which takes into account not only where we want students to end up, but also where they are coming from. Earlier approaches, in contrast, emphasized a top-down approach in which the college-level course—taken as the ultimate goal—was progressively stripped down for lower grades. Figures 11.1, 11.2, 11.3, and 11.4 are a characterization of how statistics curriculum materials and tools have been produced and how, in contrast, we approached the design of *TinkerPlots*.

When they began several years ago to design statistics courses for the high school, educators patterned the curricula largely after the introductory college course in statistics, as if they were dropping a short rope down from the college level to students in high school (see Fig. 11.1). When, more recently, educators began developing statistics units for middle and elementary school students, they continued to lower the rope (Fig. 11.2), basically by removing from the college curricula the concepts and skills they considered too difficult for younger students. The objectives and content at a particular level are thus whatever was left over after subjecting the college course to this subtractive process. So grades 3–5 get line graphs and medians, grades 6–8 get scatter-plots and means, and grades 9–12 get regression lines and sampling distributions (see National Council of Teachers of Mathematics, 2000).

Designers of statistics software tools for young students have generally followed the same top-down approach, developing software packages that are fundamentally stripped-down professional tools (Biehler, 1997, p. 169). These programs provide a subset of conventional graph types and are simpler than professional tools only in that they have fewer, and more basic, options. More recently, Cobb, Gravemeijer, and their colleagues at Vanderbilt University and the Freudenthal Institute, have taken a different approach in designing the *Mini Tools* for use in the middle school. They incorporated into the *Mini Tools* a small set of graph types—case-value plots, stacked dot-plots, and scatter-plots. And in large part they decided what to include in the tool by building from a particular theory of mathematics learning and on

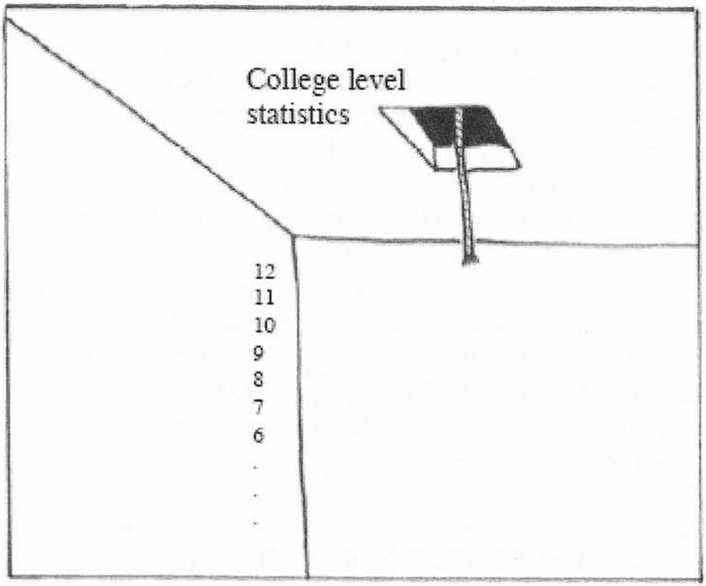

Figure 11.1 A top-down approach to developing tools and curricula for high school based on the college course.

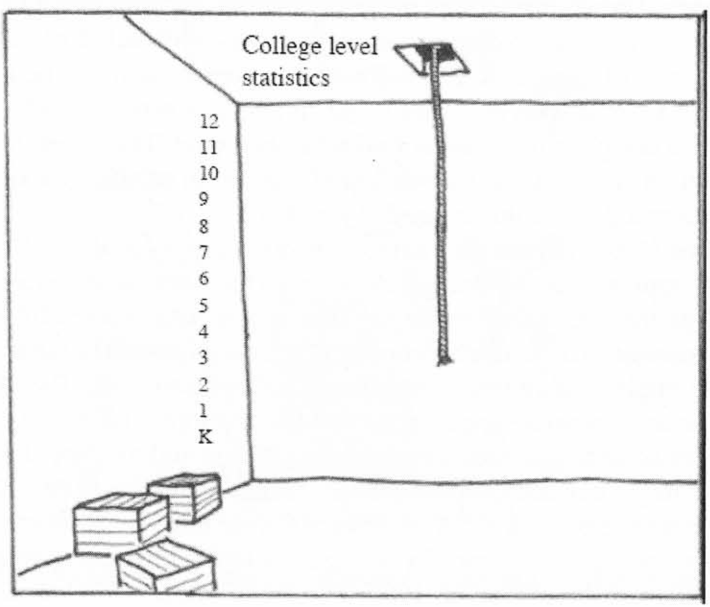

Figure 11.2 The top-down approach extended to development of materials for the lower grades.

research about student reasoning (Bakker, 2002; Cobb, 1999; Cobb, McClain, & Gravemeijer, 2003). In this way, their instructional units and accompanying software take into account not only where instruction should be headed; working from the bottom up, they also attempt to build on how students understand data and how they are prone, before instruction, to use data to support or formulate conjectures (Fig. 11.3).

In designing *TinkerPlots*, we undertook a more radical approach. First we assumed that we did not know quite where K–12 curricula should be headed and that, in any case, there were likely to be many ways for students to get there, a multiplicity of student understandings from which we could productively build up (Fig. 11.4). What held us in check, however, was an additional objective of designing software that was useful to curriculum designers writing materials that meet the NCTM Standards (2000) on data analysis. As I illustrate later, these two objectives pulled us at times in opposite directions, generating a set of tensions some of which were productive.

THE PROBLEM OF COMPLEXITY

Put five statistics educators in a room with the objective of specifying what should be in a data analysis tool intended for young students. The list of essential capabilities they generate is guaranteed to quickly grow to an alarming length. And no matter how many capabilities are built into a tool, teachers and curriculum developers—even students—will still find things they want to do, but cannot. If as a software developer you try to be helpful by including most of what everyone wants in a tool, it becomes so bloated that users then complain they cannot find what they want. Thus when it comes to the question of whether to include lots of features in a software tool, it's generally "damned if you do, damned if you don't."

Biehler (1997) refers to this as the *complexity-of-tool problem*. He suggests that one approach to addressing it is to design tools that become more sophisticated as the user gains expertise. This is just what successful computer games manage to do through a number of means (Gee, 2003), but it is hard to imagine implementing this in an educational software tool. The *Mini Tools* comprise three separate applications that the developers introduce in a specified order according to their understanding of how rudimentary skills in data analysis might develop over instruction. Perhaps the suite of *Mini Tools* is a simple example of the kind of evolving software Biehler had in mind.

In developing *DataScope* 15 years ago, we took a different approach to the complexity problem (Konold, 1995; Konold & Miller, 1994). *DataScope* is data-analysis software intended for students aged 14–17. We conceived of it

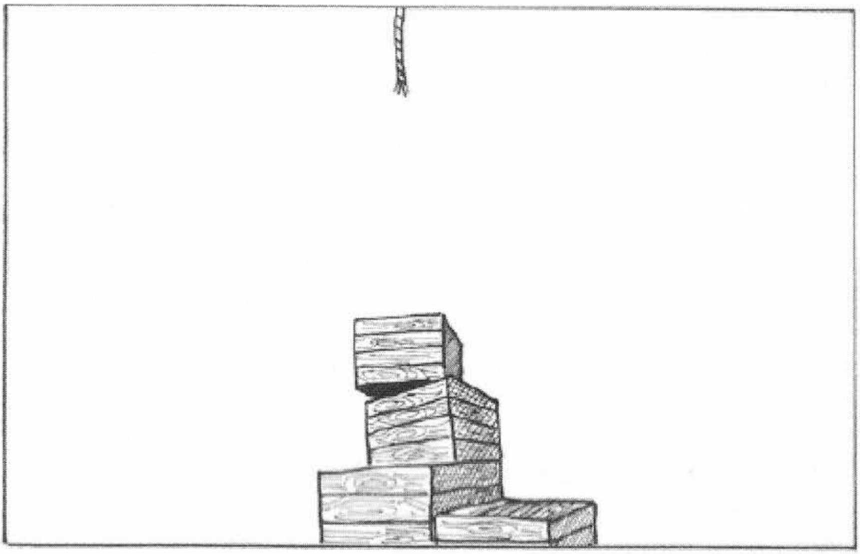

Figure 11.3 Using in conjunction a bottom up and top down approach, hoping to hook up in the middle.

Figure 11.4 A diversified bottom-up approach aimed at a moving or ambiguous target.

as a basic set of tools that would allow students to investigate multivariate data sets in the spirit of Exploratory Data Analysis (Tukey, 1977). To combat the complexity problem, we implemented only five basic representations: histograms (and bar graphs), box-plots, scatter-plots, one and two-way tables of frequencies, and tables of descriptive statistics. Our hope was that by limiting student choices, more instructional time could be focused on learning underlying concepts and data inquiry skills.

In many ways, we accomplished our goal with *DataScope*. Students took relatively little time to learn to use it, and it proved sufficiently general to allow them to flexibly explore multivariate data (Konold, 1995). However, one persistent pattern of student use troubled us. To explore a particular question, students would often select the relevant variables and then choose from the menus one of the five display options, often with only a vague idea of what the option they selected would produce. If that display did not seem useful, they would try another, and another, until they found a display that seemed to suit their purposes. If they were preparing an assignment or report, many students generated and printed out every possible display. There are undoubtedly several reasons for this behavior; Biehler (1998) reports similar tendencies among older students using software with considerably more options. However, it seemed clear that the limited number of displays in *DataScope* explained in part this trial-and-error approach, as there was little cost in always trying everything. Had this behavior been prevalent only among novice users, it would have not been of much concern. But, it persisted as students gained experience.

When we were field-testing *DataScope*, I had a fantasy that students would want to work with it outside of class—just for the fun of it, if you will. One day I walked into a class to discover that a student was already there. She had fired up the computer and was so engrossed that she didn't notice me. Trying not to disturb her, I quelled my excitement and tiptoed around her to see what data she was exploring. Alas, it was not the glow of *DataScope* lighting her face, but one of the rather mindless puzzles that early Macs included under the Apple menu. This was the closest I got in the *DataScope* days to realizing my fantasy.

It was this fantasy—of seeing students enjoying using a tool and using it with purpose—that drove many of the basic design decisions in *TinkerPlots*. The result was a tool that in ways is a complete opposite of *DataScope*. Rather than working to reduce the complexity of *TinkerPlots*, we purposely increased it. With rare exceptions, students are extremely enthusiastic with *TinkerPlots* and frequently ask to work with it outside of class. I believe that a big part of *TinkerPlots*' appeal has to do with its complexity. In what follows, I attempt to describe how we managed to build a complex tool that motivates students rather than overwhelms them.

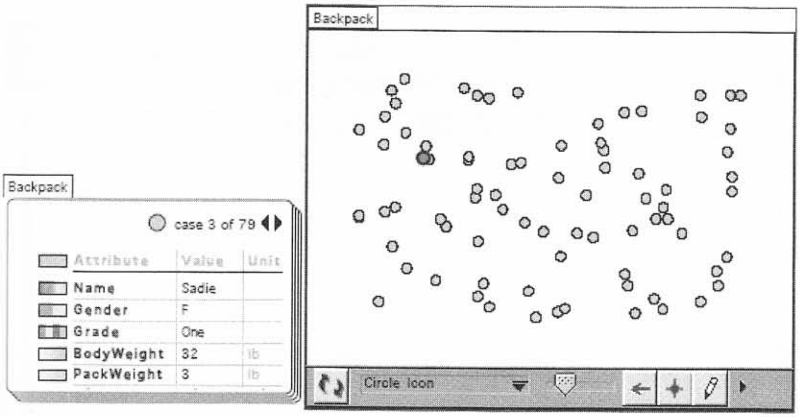

Figure 11.5 Information on 79 students along with their backpack weights displayed in *TinkerPlots*. Each case (student) is represented in a plot window (right) as a case icon. Initially, the case icons appear as shown here, randomly arranged. Clicking the Mix-up button (lower left of the plot window) sends the icons into a new random arrangement. The case highlighted in the plot window belongs to Sadie, whose data appears in the stack of Data Cards on the left.

Constructing Data Displays Using *TinkerPlots*

On first opening the plot window in *TinkerPlots*, individual case icons appear in it haphazardly arranged (see Fig. 11.5). Given the goal of answering a particular question about the data, the immediate problem facing students is how to impose some suitable organization on the case icons. *TinkerPlots* comes with no ready-made displays—no bar graphs, pie charts, or histograms. Instead, students build these and other representations by progressively organizing data icons in the plot window using basic operators including *order*, *stack*, and *separate*.

Figure 11.5 shows data I typically use as part of a first introduction to *TinkerPlots*. I ask the class whether they think students in higher grades carry heavier backpacks than do students in lower grades. I then have them explore this data set to see whether it supports their expectations. Figures 11.6 through 11.8 are a series of screen shots showing one way in which these data might be organized with *TinkerPlots* to answer this question.

In Figure 11.6, the cases have been separated into four bins according to the weight of the backpacks. This separation required first selecting the attribute PackWeight in the Data Cards and then pulling a plot icon to the

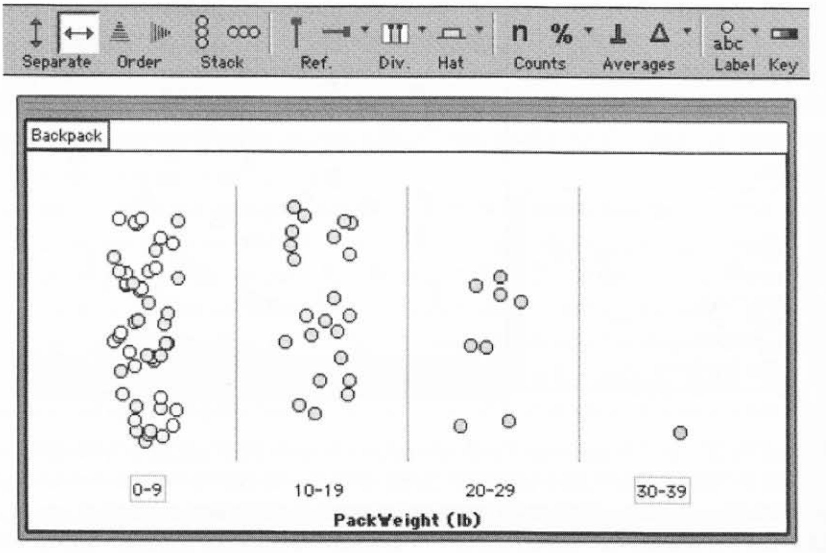

Figure 11.6 Plot icons separated into four bins according to the weight of students' backpacks. Shown above the plot window is a tool bar that includes various plotting options. When one of these buttons is pressed, it appears highlighted (as the horizontal Separate button currently is). Pressing that button again removes the effects of that operation from the plot.

right to form the desired number of bins. To progress to the representation shown in Figure 11.7, the icons were stacked, then separated completely until the case icons appeared over their actual values on a number line. Then the attribute Grade was selected, shown by the fact that the plot icons now appear in various shades of gray (in color, they appear red). With Grade selected, the Grade 5 students were separated vertically from the other grades. If we were to continue pulling out each of the three other grades one by one, we'd then see the distributions of PackWeight for each of the four grades in this data set (grades 1, 3, 5, and 7). We could go on to place *dividers* to indicate where the cases cluster, or to display the location of the means of all four groups (see Rubin, Hammerman, Campbell, & Puttick, [2005] for a description of the various *TinkerPlots* options that novices used to make comparisons between groups).

Making these displays in *TinkerPlots* is considerably more complex than it would be in *DataScope*, *Mini Tools*, *Tabletop*, *Fathom*, or most any professional or educational tool. In almost all of these packages, one would simply specify

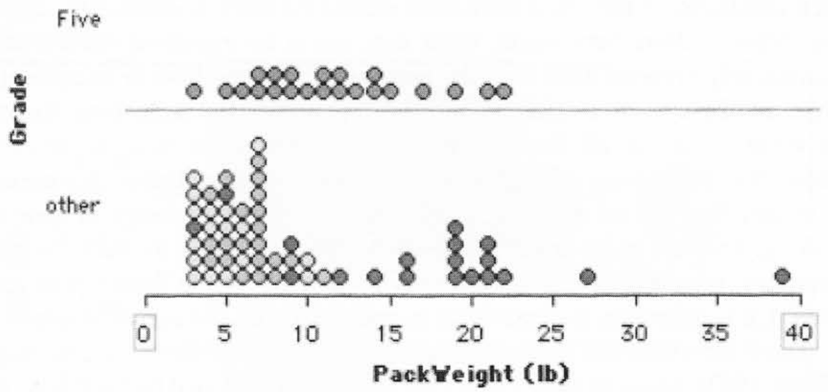

Figure 11.7 Cases have been stacked, then fully separated on the x axis until there are no bins. Then the grade 5 students have been separated out vertically, forming a new y axis. The cases are now colored according to grade, with darker gray (red in the actual program) indicating higher grade levels.

the two attributes and the appropriate graph type (e.g., stacked dot-plot). As we have seen, making such a stacked dot-plot in *TinkerPlots* requires perhaps 10 separate steps. What is important to keep in mind, however, is that the students, particularly when they are just learning the tool, typically do not have in mind a particular graph type they want to make as they organize the data. Rather, they take small steps in *TinkerPlots*, each step motivated by the goal of altering slightly the current display to move closer to their goal—in this case of being able to compare the pack weights of the different grades. Because each of these individual steps is small, it is relatively easy for students to evaluate whether the step is an improvement or not. If it is not a productive move, they can easily backtrack. The fact that with each step the icons animate into their new positions also helps students to determine the nature of, and evaluate, each modification.

There are a number of reasons we designed *TinkerPlots* as a construction set. A primary objective was that by giving students more fundamental choices about how to represent the data, they would develop the sense that they were making their own graphic representation rather than selecting from a set of preformed options. When I have students investigate the backpack data with *TinkerPlots*, I give them the task of making a graph that they can use to answer the question posed above. Having a specific task, especially when first learning *TinkerPlots*, is crucial. Without a clear goal, students would have no end to inch toward and thus no basis for evaluating their actions.

After about 30 minutes, most of the students have answered the question to their satisfaction. I then have them walk around the room to observe the displays that other students have made. What they see is an incredible variety, which immediately presents them with the problem of learning how to interpret these different displays, all of which are purportedly showing the same thing. But more importantly, seeing all these different graphs makes it clear to them that *TinkerPlots* is not doing the representational work for them. Rather, they are using it as they might a set of construction blocks to fashion a design of their own making. They are in the driver's seat, which means they have to make thoughtful decisions; mindlessly pressing buttons will most likely give them a poor result. Indeed, it is quite easy in *TinkerPlots* to make cluttered and useless displays.

There are numerous factors that affect the interpretability of a data display (Tufte, 1983). Many of these factors are ordinarily controlled by a software tool. In *TinkerPlots*, we chose to leave some rather fundamental display aspects under direct user control. Figure 11.8 shows the four levels of grade separated out on the y axis. But the plot icons are so large that they spill over the bin lines, and any subtle features of the four distributions are obscured. This sort of plot-crowding routinely occurs as students are making various graphs in *TinkerPlots*, and it is up to them to manually control the size of icons, which they quickly learn to do. It is a control they seem to enjoy exercising.

Note, too, in Figure 11.8 that the four levels of grade are not ordered sensibly. The current order resulted from the particular way each group was pulled out of the "other" category visible in Figure 11.7. In creating this data set, we intentionally entered the values of grade as text rather than as numbers so that students would tend initially to get a display like this, with values of grade not in an order ideal for comparing them. The ordering can be quickly changed, however, by dragging axis labels to the desired locations. Once ordered, students can sweep their eyes from bottom to top to evaluate the pattern of differences among the groups without having to continually refer back to the axis labels. In fact, it is this type of ordering from which graphic displays of data derive much of their power.

Leaving such details to the student further increases the complexity of the program. However, taking control of things like icon size, bin size, and the ordering of values on an axis helps students to become explicitly aware of important principles that underlie good data display. Furthermore, leaving these fundamental responsibilities to the student is yet another way of communicating to them that they, and not the software tool, are ultimately in control of what they produce. Finally, these are factors that most students seem to enjoy having direct control over. Part of this satisfaction undoubtedly comes from the fairly direct nature of the control and would be lost if instead we had used dialogue boxes.

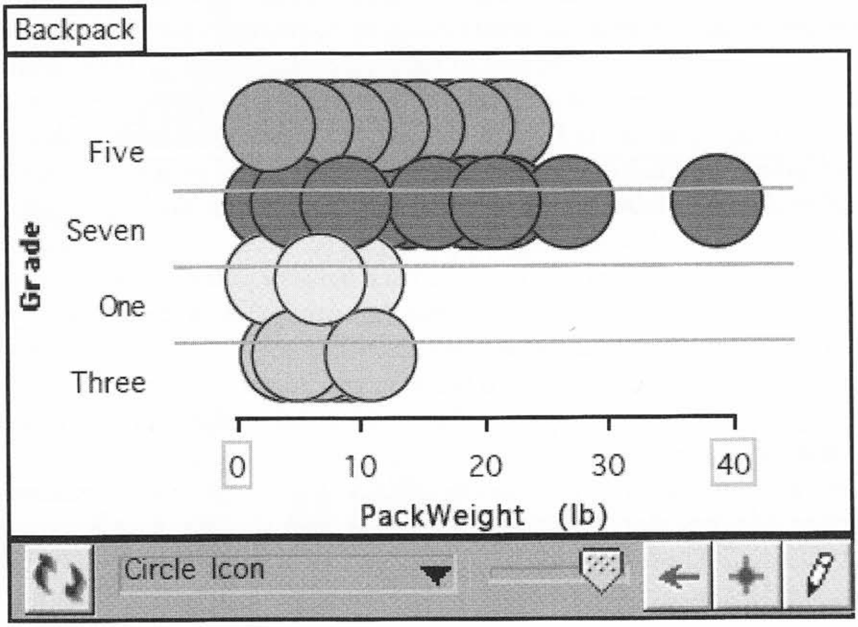

Figure 11.8 The plot icons in this graph are so large they obscure much of the data. Their size is under user control via the slider located on the tool bar below the plot.

Making the Complex Manageable

Certainly, it is not the complexity itself that makes *TinkerPlots* compelling, but the nature of that complexity. Indeed, one of the ways Biehler (1997) suggested to make a complex tool manageable is to build it around a "conceptual structure . . . which supports its piecewise appropriation" (p. 169). We chose the operators *separate*, *order*, and *stack* after having observed how students (and we ourselves) organized data on a table when it was presented as a collection of cards with information about each case on a separate card (Harradine & Konold, 2006). We then worked to implement these operations in the software in a way that would allow students to see the computer operations as akin to what they do when physically arranging real-world objects. This sense—that one already knows what the primary software operators will do—becomes important in building up expectations about how the various operators will interact when they are combined, because it is this ability to combine operators in *TinkerPlots* that makes it complex, and powerful.

Implementing these intuitive operators in the software was harder than we initially expected, however. In our first testable prototype, about half of the representations that students would make by combining operators were nonsensical. To remedy this, we had to reinterpret what some of the operations did in various contexts. *Stack*, for example, works as one might expect with the case icon style used in Figures 11.5 to 11.8. However, there are other icon styles where the stack operation behaves a bit differently so as to produce reasonable displays. For example, icons can be changed to *fuse rectangular*, a style used to make histograms (see bottom of Fig. 11.9). In this case, *stack* not only places case icons on top of one another, but also widens them so that they extend across the entire length of the bin they occupy. With the icon style *fuse circular*, case icons become wedges that fuse together into a circle (pie graphs). In this case, *stack* has no function and thus if it is turned on, it does nothing. In general, the user is unaware of these differences, but pays no price for this ignorance.

We avoid using error messages to instruct students, primarily because we worried that they would erode the attitude we are working hard to create—that the student, not the software, is in control. In some cases, applying an operator does nothing to the plot, and the button dims to indicate that it is in a suppressed state (as happens with *stack* in the context of pie graphs). Again, this goes mostly unnoticed.

However, whenever we can, we show some change in the plot, even if it is of only limited use. For example, when a numeric attribute is fully separated on an axis, students can click the median button to display the location of the median below the axis (see top of Fig. 11.9.) With a binned dot-plot, however, it would be misleading to show the median as a specific point on an axis. But rather than have nothing happen when students turn on the median in this state, we display the median as a line running the length of the interval in which the median occurs (middle graph in Fig. 11.9). Although it does not provide much information about the value of the median, this display does help communicate the fact that when we place different values into the same bin we are, for the moment, considering them to be the same. This binned dot-plot can be changed into a histogram by selecting the icon style *fuse rectangular* (see bottom graph of Fig. 11.9). Now the median symbol once again appears at a precise location on a continuous axis. The animation from the binned dot-plot to the histogram shows the cases growing in width to the edges of the bin lines, hinting at yet another change in how we are thinking of the values in a common bin.

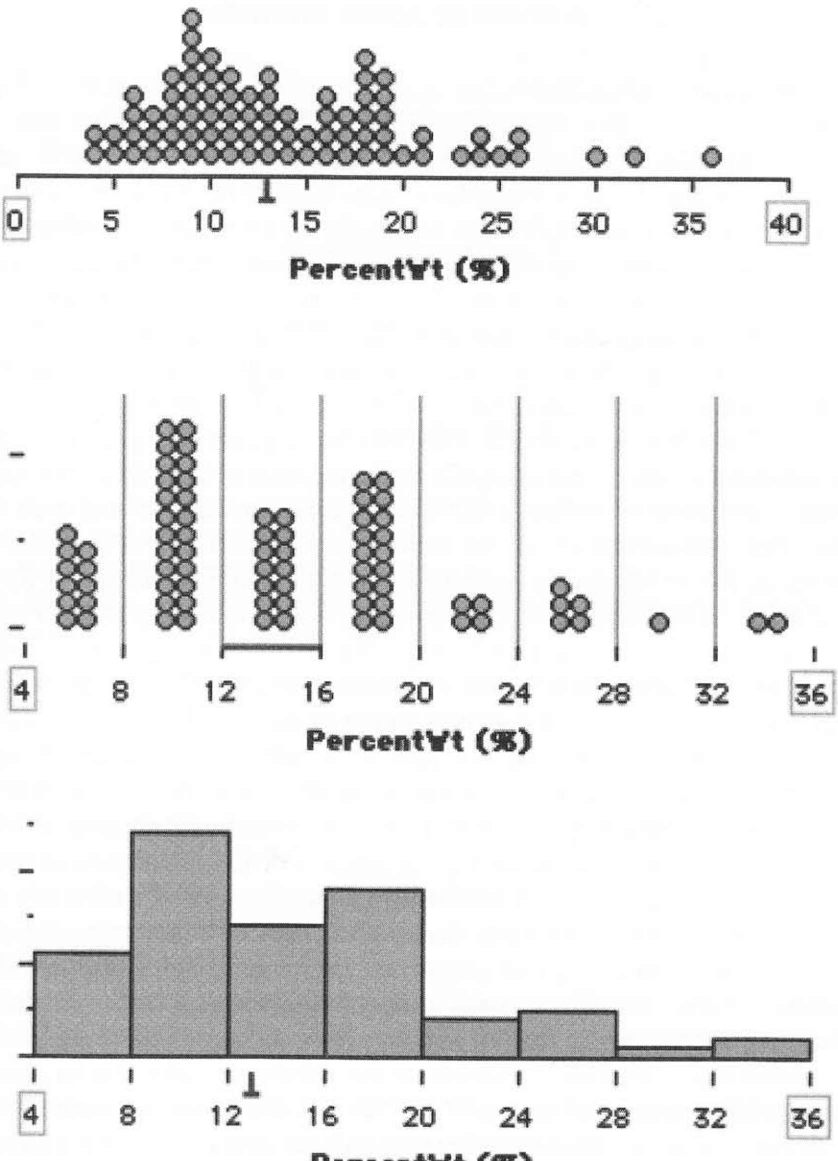

Figure 11.9 These graphs display the percentage the backpacks are of body weight. The top graph shows the location of the median (inverted T) at 13. In the binned dot-plot in the middle, the median now appears as a line below the bin, indicating that the median is in the interval 12–16. Changing the icon style to "fuse rectangular" makes a histogram, which now again displays the precise location of the median.

HAT-PLOT: A CASE OF BUILDING UP FROM STUDENT THINKING

In addition to building *TinkerPlots* around a conceptual structure involving the operators *stack*, *order*, and *separate*, we also included features that were inspired by what we and others had observed students wanting to do in making and reading data displays. These features include the *reference line*, which students use to mark salient features and to determine the precise value of case icons, and *dividers* for partitioning of data into subgroups. In this section, I describe the *hat-plot*, a new type of display we introduce in *TinkerPlots*.

One way to think of a hat-plot is as a generalization of Tukey's (1977) box-plot. Figure 11.10 shows percentile hat-plots for the weights of students in four different grades. Each hat is composed of two parts: a "brim" and a "crown." The brim is a line that extends to the range for each group; the crown is a rectangle that, in this case, shows the location of the middle 50% of the data—the Interquartile Range (IQR). The particular hat-plots in Figure 11.10 are thus constructed using the same basic conventions of the box-plot. Whereas the whiskers of a box-plot are drawn through the center of the IQR rectangle, in the hat-plot the corresponding line is drawn along the bottom of this rectangle. We think that locating the central rectangle on top of the whiskers, or range line, helps emphasize what the center rectangle is depicting—the location of a central clump of the data. I say more about this later. In addition, our own sense is that the hat-like display that results makes it easier for students to notice and describe general differences in the shapes of the distributions. Note in Figure 11.10, for example, the striking difference in appearance between the hat for the weights of first-graders, with its relatively tall crown, and the hat for the seventh-graders, with its relatively short, but spreading, crown. The more skewed a distribution is, the more its hat-plot will appear as something like a baseball cap. We also think students will take quickly to the idea of summarizing distributions with hats, in part because they will already have a rich vocabulary for doing so.

Whereas the median is an inherent part of the box-plot display, it is not automatically displayed as part of the hat-plot. But using a separate control, you can display its location below the axis as shown in Figure 11.10. The result is that a box-plot divides the data into four parts, whereas the hat-plot divides it into three. We anticipate that this will have some pedagogical advantages as students already have a strong tendency to view many distributions as comprising three groups (see, e.g., Bakker & Gravemeijer, 2004).

A more fundamental difference between hat-plots and box-plots is that with hat-plots, you can change the setting for the brim edges to represent percentiles other than the box-plot's 25th and 75th percentiles. By clicking and

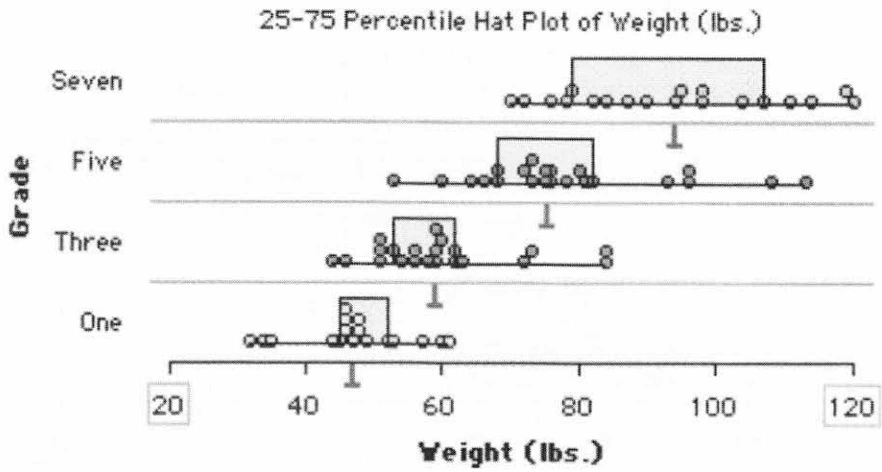

Figure 11.10 Superimposed percentile hat-plots for weight (in pounds) of students in grades 1, 3, 5, and 7. The edges of the hat crowns show the location of the 25th and 75th percentiles. The inverted T's under each group indicate medians. Thus these particular hat-plots show the same basic information as would box-plots, except that rather than retracting to display outliers, hat-plot brims extend to the minimum and maximum values.

dragging the edge of a brim, you can change these into, for example, 20th–80th percentile hats. Furthermore, each brim edge is adjusted independently, so you could set them to display 20th–76th percentiles. This allows students to make hat-plots that are initially tailored to a particular group of data, which they later can test on other groups.

You can also switch the metric of the brim from percentiles to ranges, average deviations or standard deviations. In Figure 11.11, I've used the range metric to construct hat crowns that extend from one third of the range to two thirds of the range.

Some have questioned our choice to include in *TinkerPlots* a display that is not among those listed in the NCTM Standards for the middle grades. There seems to be an implicit assumption in this question—that if something is not in the Standards, we should not include it in our learning objectives, curricula, or student tools. Although there are some grounds for taking this stance, we should reject it as a guiding principle.

The Standards are not a sacred canon but rather "a resource and guide" (NCTM, 2000, p. ix). NCTM describes the Standards as "part of an ongoing process of improving mathematics education" and believes that for the Standards "to remain viable, the goals and visions they embody must periodically be

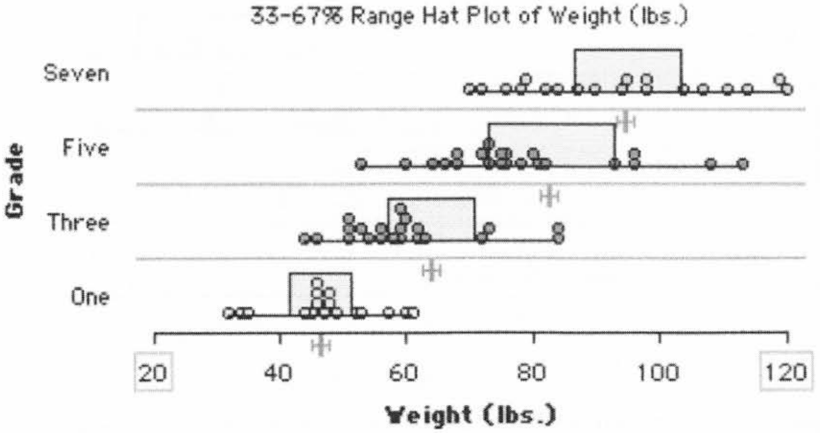

Figure 11.11 Range hat-plots for the weight of students (in pounds) in grades 1, 3, 5, and 7. These plots divide the range for each grade into equal thirds. The symbols below the bin lines indicate the midranges.

examined, evaluated, tested by practitioners, and revised" (p. x). Thus, it is contrary to the spirit of the Standards to use them to justify and support a rigid orthodoxy about what and how we should teach. As curriculum writers, researchers, and educators, we should be pushing ourselves to test and refine our vision of how students might learn to analyze data. Put another way, we should be thinking at least as much about what the next version of the Standards should say as we do about what this version says.

Instructional Role of Hat-Plots

In the following, I briefly present the rationale for including hat-plots in *TinkerPlots* and the reasons why I think they could play a helpful role in middle school data-analysis curricula. I also offer some tentative, and admittedly vague, ideas about how they might be used in a sequence of instruction, confident only in the fact that these ideas will change as a result of more thought and of trying them out in classrooms.

From the beginning, we struggled with the question of whether and how to implement box-plots in *TinkerPlots*. One of our objectives was to avoid using plot types as operational primitives. Our aim was to have standard displays, such as scatter-plots and histograms, be among the many possibilities that emerged as students progressively organized data using the basic operators *order*, *stack*, and *separate*. And as previously mentioned, the more general principle that informed the design of *TinkerPlots* was that, to the extent possible, we should build instruction on what students already know and are inclined to do.

Box-plots posed a problem to both these principles. First, the operational primitives in *TinkerPlots* worked well for producing most of the traditional displays included in the current middle school data analysis curricula, but they did not get us to box-plots. (This is another indication, by the way, of just how different box-plots are from most other statistical graphs. See Bakker, Biehler & Konold, 2005.) Second, regardless of how we imagined implementing box-plots in *TinkerPlots*, we could see no clear way of introducing box-plots to students other than as a convention some statisticians find useful. It should be clear at this point that despite what I said above about not allowing the Standards to co-opt educational choices, we as developers certainly feel a great deal of pressure to accommodate them. The fear is that if we do not, various decision-makers including publishers, school boards, and teachers, who themselves are often under the mandate to adopt "Standards-based" approaches, will elect not to use what we have developed. In this regard, the Standards get used both to facilitate and to stifle reform and innovation.

One of our early solutions to implementing box-plots was to adopt the approach used in the *Mini Tools* (see Cobb, 1999). In *Mini Tool 2*, students can overlay various types of groupings on top of stacked dot-plot displays. Creating four, equal-count groups is one of several options, and one that the Vanderbilt learning trajectory made special use of, because it could lead naturally into box-plots. This solved the first problem for us, in that box-plots could emerge in *TinkerPlots* as they did in the *Mini Tools* from the more basic act of dividing into groups.

However, there were aspects of teaching box-plots to students that we struggled with. In particular, it was not clear how to motivate students to make groups composed of roughly equal numbers of cases, and furthermore to make four such groups (rather than three or five or twenty). As Arthur Bakker put it in an e-mail exchange with us, "I have never found any activities with data sets that really begged to be organized by four equal [count] groups …"

These reservations eventually led us to think about how we might build box-plot-like displays on the well-known tendency of students to summarize single, numeric distributions using "center" or "modal" clumps. These are ranges of values in the heart of a distribution that students use to indicate what is "typical" or "usual." Among the researchers who have reported the use of these sorts of "hills" or "clumps" are Bakker (2001); Cobb, (1999); Konold and Higgins (2003); Konold, Robinson, Khalil, Pollatsek, Well, Wing, et al. (2002); and Noss, Pozzi, and Hoyles (1999).

Inspired by this research, we included in *TinkerPlots* a divider tool that displays two lines on top of a distribution of values that students can freely

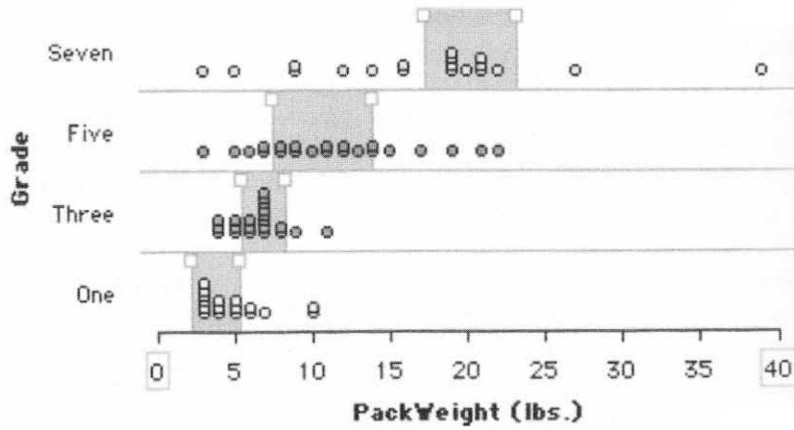

Figure 11.12 Stacked dot-plots of backpack weight (in pounds) of students in grades 1, 3, 5 and 7. Overlaid on each distribution is a pair of adjustable dividers that students can use to mark where data tends to be centered. As an option, they can display the number (and/or percent) of cases contained in each of the three divisions.

adjust (see Fig. 11.12). Our primary intention was that students would use these dividers to mark the location of modal clumps and, optionally, to display the number (or percent) of cases both inside and outside these clumps. Many researchers have commented on the difficulty students have using normative averages, including the mean and median, in meaningful ways (for a review, see Konold & Higgins, 2003). Modal clumps may provide a way for students to begin by using an average-like construct that does make intuitive sense to them. Furthermore, Cobb, Gravemeijer and their colleagues have described teaching sequences directed towards encouraging students to use these to compare groups by noting the relative positions of the "hills" in two distributions.

Figure 11.13 shows the added possibility in *TinkerPlots* of simultaneously seeing modal clumps and the location of means and medians. Seeing these together provides opportunities for helping students to develop more intuitive views of the standard measures of center. Freed from the role of representing what's average about the data, modal clumps might then provide students an informal way of describing the "average" variability in the data. Research by Konold et al. (2002) suggests in fact that the width of students' modal clumps is remarkably close to those of IQRs.

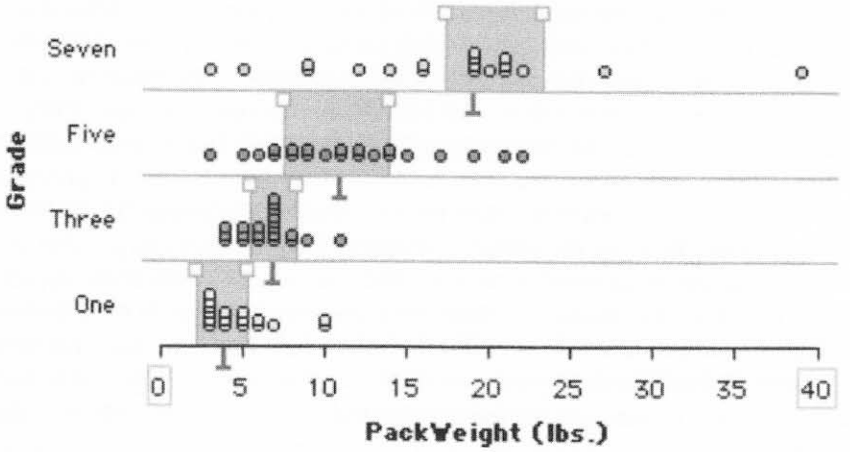

Figure 11.13 The same displays as shown in Figure 11.11 with the addition of the medians.

Our idea for hat-plots then emerged as a way for students to formalize their idea of modal clump, so that rather than fitting modal clumps *ad hoc* on the basis of how data happened to be distributed in a particular display, they could set them according to more objective, and previously agreed-to, criteria. Thus, our overall intention is that in analyzing data, students will naturally take to using the dividers provided in *TinkerPlots* as a way to indicate and communicate to others what they perceive as typical values in a distribution. Hat-plots will offer students a way of formalizing these modal clumps as part of establishing agreed on and objective criteria for using modal clumps to decide, for example, if two groups are different. That hat-plots divide a distribution into three components, just as dividers do, should facilitate the transition from dividers to hat-plots. It also fits with the observations of several researchers that students often initially perceive a distribution of values as comprising low, middle, and high values (e.g., see Bakker & Gravemeijer, 2004).

One reason for providing different metrics for hat-plots (e.g., where the three components split the distribution into different fractions of the range or into multiples of the average or standard deviation) is to encourage students to view hat-plots as a more general method for representing data. Another is to allow exploration of the relative strengths of various metrics. For many students, the range is a salient feature of distributions. We therefore expect that many students will initially choose to construct hat-plots by specifying fractions of the range, which is why originally we used the range metric as the default setting. Our hope was that with some exploration, students would discover some of the drawbacks of using the range. For example, it will often be the case that a range hat-plot that looks reasonable for the group on which it was constructed will not do a good job on other groups. Note that in Figure 11.11, the range hat-plots for students in grades 7 and 1 seem reasonable as summaries of the modal clumps; those for grades 5 and 3 do a fairly poor job. In contrast, the percentile hat-plots for the same data in Figure 11.10 fit the modal clumps of all of the groups reasonably well. Students might also discover using the range metric how one value can drastically affect the appearance of the hat-plot. Indeed, by dragging a single case on one of the extremes using the change-value tool in *TinkerPlots*, they can watch the entire hat slide upwards or downwards in pace with that case, making the range metric perhaps too fussy or sensitive for use in comparing groups. As an aside, we later changed the hat-plot default setting to percentiles when we observed many teachers using the hat-plot's range default setting but assuming that they were percentiles.

We have included in *TinkerPlots* both dividers and hat-plots in the hope that they may provide means for allowing students to build on intuitive ideas that they have about distributions. Our reasons for thinking these tools might be useful are based on recent research that has explored student reasoning and investigated various approaches to instruction that build on students' intuitive ideas. In the near future, we expect to learn from the curriculum developers, who are working in collaboration with the *TinkerPlots* development team, whether hat-plots and dividers are indeed useful, and what modifications or enhancements might make them more so.

CONCLUSIONS

In helping students learn a complex domain such as data analysis, we inevitably must find effective ways to restructure the domain into manageable components. The art is in finding ways to do this that preserve the essence and purpose of the pursuit. It is all too common in classrooms to find students succeeding at learning the small bits they are fed, but never coming to see the big picture nor experiencing the excitement of the enterprise. Of course, *TinkerPlots* by itself cannot change this, and much depends on how teachers and curriculum developers put it to use. Just as I have watched in frustration as students in traditional classrooms spend months learning to make simple graphs of single attributes and never get to a question they care about, I now have had the experience of watching students work through teacher-made worksheets to learn *TinkerPlots* operations one at a time, "mastering" each one before moving on to the next. This is despite the fact that the parts cannot be mastered in isolation or out of context.

After class, I spoke with the teacher who had created the worksheets and gently offered the observation that students could discover and learn to use many of the commands he was drilling them on as a normal part of pursing a question. He informed me that they didn't have time in their schedule to have students "playing around." Although his response added to my despair about the direction education in the United States seems to be heading under the pressures of the testing/accountability movement, I also took it as another indicator that we succeeded with *TinkerPlots* in developing the tool we had hoped to—that in the absence of the strict regime of a worksheet, students seem to actually enjoy using it to explore data.

ACKNOWLEDGMENTS

I thank Amy Robinson for her comments on portions of this chapter and Daniel Konold for drawing Figures 11.1 through 11.4. Portions of this chapter are from my article "Handling Complexity in the Design of Educational Software Tools," to appear in the *Proceedings of the Seventh International Conference on Teaching Statistics*. *TinkerPlots* is published by *Key Curriculum Press* and was developed with grants from the National Science Foundation (ESI-9818946, REC-0337675, ESI-0454754). Opinions expressed here are my own and not necessarily those of the Foundation.

REFERENCES

Bakker, A. (2002). Route-type and landscape-type software for learning statistical data analysis. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching of Statistics*. Voorburg, The Netherlands: International Statistical Institute.

Bakker, A., Biehler, R., & Konold, C. (2005). Should young students learn about box-plots? In G. Burrill & M. Camden (Eds.), *Curricular development in statistics education: International Association for Statistical Education 2004 Roundtable* (pp. 163–173). Voorburg, the Netherlands: International Statistical Institute.

Bakker, A., & Gravemeijer, K. P. E. (2004). Learning to reason about distribution. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 147–168) Dordrecht: Kluwer.

Biehler, R. (1997). Software for learning and for doing statistics. *International Statistical Review*, 65(2), 167–189.

Biehler, R. (1998). Students—statistical software—statistical tasks: A study of problems at the interfaces. In L. Pereira-Mendoza, L. S. Kea, T. W. Kee, & W. Wong (Eds.), *Proceedings of the Fifth International Conference on Teaching Statistics* (pp. 1025–1031). Singapore: International Statistical Institute.

Cobb, P. (1999). Individual and collective mathematical development: The case of statistical data analysis. *Mathematical Thinking and Learning, 1*(1), 5–43.

Cobb, P., McClain, K., & Gravemeijer, K. (2003). Learning about statistical covariation. *Cognition and Instruction, 21*, 1–78.

Falk, R., & Konold, C. (1997). Making sense of randomness: Implicit encoding as a basis for judgment. *Psychological Review, 104,* 301–318.

Flowers, J. H., & Hauer, T. A. (1992). The ear's versus the eye's potential to assess characteristics of numeric data: Are we too visuocentric? *Behavior Research Methods, Instruments, & Computers, 24(2),* 258–264.

Gee, J. P. (2003). *What video games have to teach us about learning and literacy*. New York: Palgrave Macmillan.

Harradine, A., & Konold, C. (2006). How representational medium affects the data displays students make. In A. Rossman & B. Chance (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics (ICOTS)* [CD-ROM]. Voorburg, The Netherlands: International Statistical Institute.

Konold, C. (1989). Informal conceptions of probability. *Cognition and Instruction, 6*, 59–98.

Konold, C. (2002). Teaching concepts rather than conventions. *New England Journal of Mathematics, 34(2),* 69–81.

Konold, C. (1995). *Datenanalyse mit einfachen, didaktisch gestalteten Softwarewerkzeugen für Schülerinnen und Schüler* [Designing data analysis tools for students]. *Computer und Unterricht, 17*, 42–49.

Konold, C., & Higgins, T. L. (2003). Reasoning about data. In J. Kilpatrick, W. G. Martin, & D. Schifter (Eds.), *A research companion to Principles and Standards for School Mathematics* (pp. 193–215). Reston, VA: National Council of Teachers of Mathematics

Konold, C., & Miller, C. (1994). *DataScope*. Santa Barbara: Intellimation Library for the Macintosh.

Konold, C., & Miller, C., D. (2005). *TinkerPlots: Dynamic data exploration.* Emeryville, CA: Key Curriculum Press.

Konold, C., Pollatsek, A., Well, A. D., Lohmeier, J., & Lipson, A. (1993). Inconsistencies in students' reasoning about probability. *Journal for Research in Mathematics Education, 24,* 392–414.

Konold, C., Robinson, A., Khalil, K., Pollatsek, A., Well, A., Wing, R., et al. (2002). Students' use of modal clumps to summarize data. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching of Statistics*. Voorburg, The Netherlands: International Statistical Institute.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: NCTM.

Noss, R., Pozzi, S. & Hoyles, C. (1999). Touching epistemologies: Meanings of average and variation in nursing practice. *Educational Studies in Mathematics, 40(1),* 25–51.

Pollatsek, A., Konold, C., Well, A. D., & Lima, S. D. (1984). Beliefs underlying random sampling. *Memory & Cognition, 12*, 395–401.

Rubin, A., Hammerman, J., Campbell, C., & Puttick, G. (2005). The effect of distributional shape on group comparison strategies. In K. Makar (Ed.), *Reasoning about distribution:*

A collection of current research studies. Proceedings of the Fourth International Research Forum on Statistical Reasoning, Thinking, and Literacy (SRTL-4). Brisbane: University of Queensland.

Thamel, P. (2003). Baseball; Red Sox' Checklist: Manager (Yes), Rodriguez (Maybe). *The New York Times*, December 5, 2003. Sec. D, p. 3.

Tufte, E. R. (1983). *The visual display of quantitative information*. Cheshire, CT: Graphics Press.

Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.

Well, A. D., Pollatsek, A., & Boyce, S. J. (1990). Understanding the effects of sample size on the variability of the mean. *Organizational Behavior and Human Decision Processes*, *47*, 289–312.