4 Technology and mathematics education

An essay in honor of Jim Kaput

Cliff Konold
University of Massachusetts Amherst

Richard Lehrer[1]
Vanderbilt University

Technologies of writing have long played a constitutive role in mathematical practice. Mathematical reasoning is shaped by systems of inscription (i.e., writing mathematics) and notation (i.e., specialized forms of written mathematics), and systems of inscription and notation arise in relation to the expressive qualities of mathematical reasoning. Despite this long history, something new is afoot: Digital technologies offer a significant expansion of the writing space. In this essay, we begin with a view of the developmental origins of the coordination of reasoning and writing, contrasting notational systems to more generic forms of inscription. Then, following in the footsteps of Jim Kaput, we portray dynamic notations as a fundamental alteration to the landscape of writing afforded by digital technology. We suggest how dynamic notations blend the digital character of notation with the analog qualities of inscription, resulting in a hybrid form that is potentially more productive than either form in isolation. We conclude with another proposition spurred by Jim Kaput: Digital technologies afford new forms of mathematics. We illustrate this proposition by describing children's activities with *TinkerPlots*™ *2.0*, a tool designed to help students organize and structure data, and to relate their understandings of chance to these patterns and structures in data.

WRITING MATHEMATICS

Rotman (1993) portrays mathematics as a particular kind of written discourse; " a business of making and remaking permanent inscriptions…operated upon, transformed, indexed, amalgamated…" (Rotman, 1993, p. 25). By inscriptions, Rotman refers generally to marks on paper. He further suggests that inscriptions (signifiers) and mathematical ideas or objects are "co-creative and mutually originative" (Rotman, 1993, p. 33), so that, in his view, reasoning and writing mathematics co-originate. One does not first know and then symbolize, or first symbolize, and then know. DiSessa (2000) suggests, too, that writing mathematics can also spur new forms of mathematical experience:

> Not only can new inscription systems and literacies ease learning, as algebra simplified the proofs of Galileo's theorems, but they may also rearrange the entire terrain. New principles become fundamental and old ones become obvious. Entirely new terrain becomes accessible, and some old terrain becomes boring. (diSessa, 2000, p. 19)

Notations

Although mathematics utilizes a wide range of inscriptional forms, we are especially concerned with a form that Goodman (1976) describes as notational. He suggests heuristic principles to distinguish notational systems from other systems of inscription. These principles govern relations among inscriptions (signifiers-literal markings), objects (signified), character classes (equivalent inscriptions, such as different renderings of the numeral 7), and compliance classes (equivalent objects, such as Gaussian distributions or right triangles). Three principles govern the grammar of inscriptions that qualify as notational: (1) syntactic disjointedness, meaning that each inscription belongs to only one character class (e.g., the marking 7 is recognized as a member of a class of numeral 7's, but not numeral 1's; (2) inscriptional clarity, meaning that all variations in the marking of numeral 7 are treated as equivalent,); and (3) syntactic differentiation, meaning that one can readily determine the intended referent of each mark (e.g., if one used different lengths to mark different quantities, then the differences in lengths corresponding to the differences in quantities should be readily perceivable).

Recalling that compliance classes are the intended field of reference of the inscriptions, two other semantic principles regulate mappings between character classes and compliance classes. The first requirement is one of semantic consistency between the inscription and referent. It will not do if the reference shifts with context. Goodman (1976) terms this consistency semantic disjointedness. It implies a straightforward arrangement: an inscription, such as the numeral 7, always refers to the same quantity, although the members of the set can vary: 7 can refer to seven cats or seven apples but not at times, seventy cats or apples. A more subtle implication is that character classes, however inscribed, should not have overlapping fields of reference. This requirement rules out the intersecting categories of natural language, such as indicating that the numeral 7 refers to both a quantity and a "lucky number." A second principle of semantic differentiation indicates that every object represented in the notational scheme can be classified discretely (assigned to a compliance class)—a principle of digitalization which applies even to analog qualities. For example, although one can perceive the space between minutes on an analog watch, the system is notational if the teller of time consistently uses the minute markings. These markings carve the analog system at the joints, thus rendering it digitally. Of course, rendering analog to digital has become widespread with electronic technologies. It is easy for all but audiophiles to treat digitalization of music as transparent.

These features of notational systems afford the capacity to treat symbolic expressions as things in themselves, and thus to perform operations on the symbols without regard to what they might refer. This capacity for symbolically-mediated generalization creates a new faculty for mathematical reasoning and argument (Kaput, 1991, 1992; Kaput & Schaffer, 2002). For example, the well-formedness of notations makes algorithms possible, transforming ideas into computations (Berlinski, 2000). Notational systems provide systematic opportunity for student expression of mathematical ideas, but the systematic character of notation places fruitful constraints on expression (Thompson, 1992).

Learning to notate

Several studies elaborate on the relation between expression and constraint even in early childhood. Van Oers (2000, 2002) suggests that early parent–child interactions and play in pre-school with counting games set the stage for fixing and selecting portions of counting via inscription. When a child counts, parents have the opportunity to re-interpret that activity as referring to cardinality instead of mere succession. For example, as a child completes her count, a parent may hold up fingers to signify the quantity and repeat the last word in the counting sequence (e.g., 3 of 1, 2, 3). This act of notation, although perhaps crudely expressed as finger-tallies, curtails the activity of counting and signifies its cardinality. As suggested by Latour (1990), the word or tally (or numeral) can be transported across different situations, such as three candies or three cars, and so number becomes mobile as it is recruited to situations of "how many."

Pursuing the role of notation in developing early number sense, Munn (1998) investigated how preschool children's use of numeric notation might transform their understanding of number. She asked young children to participate in a "secret addition" task. Children first saw blocks in containers, and then they wrote a label for the quantity (e.g., with tallies) on the cover of each of four containers. The quantity in one container was covertly increased, and children were asked to discover which of the containers had been incremented. The critical behavior was the child's search strategy. Some children guessed. Others thought that they had to look in each container and try to recall its previous state. However, many used the numerical labels they had written to check the quantity of a container against its previous state. Munn found that over time, preschoolers were more likely to use their numeric inscriptions in their search for the added block, using inscriptions of quantity to compare past and current quantities. In her view, children's notations transformed the nature of their activity, signaling an early integration of inscriptions and conceptions of number.

Co-origination of mathematical thought and systems of inscription and notation

The co-creation of mathematical thought and inscription has been documented by researchers studying the mathematical activity of individuals across a diverse range of settings. Hall (1990, 1996), for example, investigated the inscriptions generated by algebra problem solvers (ranging from middle school to adult participants, including teachers) during the course of solution. He suggested that the quantitative inferences solvers made were not a simple result of parsing strings of expressions. Rather, the inferences sprang from "representational niches" defined by interaction among varied forms of inscription (e.g., algebraic expressions, diagrams, tables) and narratives. These niches or "material designs" helped participants visualize relations among quantities and stabilized otherwise shifting frames of reference. For example, when solving even simple algebraic word problems, participants typically employed sketches of the situation, tables and other representations that allowed them to iterate a model of the functional relations described in the problem, and only then employed the symbolism typically associated with algebra. Multiple inscriptions for developing representation of the problem were integral to mathematical solution.

Co-evolution of inscription and thinking was also prominent in Meira's (1995, 2002) investigations of (middle school) student thinking about linear functions that describe physical devices, such as winches or springs. His analysis focused on student construction and use of a table of values to describe relations among variables, such as the turns of a winch and the distance an object travels. As pairs of students solved problems, Meira (1995) noted shifting signification, in that students used marks initially representing weight to later represent distance. He also observed several different representational niches such as transforming a group of inscriptions into a single unit and then using that unit in subsequent calculation. This demonstrates a clear dependence of problem-solving strategies on qualities of the number tables and a lifting away from the physical devices to operations in the world of the inscriptions—a way of learning to see the world through inscriptions.

Izsak (2000) found that pairs of eighth-grade students experimented with different possibilities for algebraic expressions as they explored the alignment between computations on paper and the behavior of the winch featured in the Meira (1995) study. Pairs also negotiated shifting signification between symbols and aspects of device behavior, suggesting that interplay between mathematical expression and qualities of the world may constitute one genetic pathway for mediating mathematical thinking via inscriptions.

In their studies of student appropriation of graphical displays, Nemirovsky and his colleagues (Nemirovsky & Monk, 2000; Nemirovsky, Tierney, & Wright, 1998) suggested that learning to see the world through systems of inscription is more accurately described as a "fusion" between signifiers and signified. In their view, coming to interpret an inscription mathematically often involves treating the signifiers and the signified as undifferentiated, even though one knows that they can be treated distinctly. The roots of these capabilities are likely found in pretense and possibility (Lehrer & Lesh, 2003). In their studies of students' attempts to interpret graphical displays of physical motion, Nemirovsky et al. (1998) recount an instance of a teacher scaffolding by using "these" to simultaneously refer to lines on a graph, objects (toy bears), and a narrative in which the bears were nearing the finish of a race. This referential ambiguity helped the student create an interpretation of the inscription that was more consistent with disciplinary practice as she sorted out the relations among inscription, object, and the ongoing narrative that anchored use of the inscription to a time course of events.

Stevens and Hall (1998) further suggested that mathematical learning mediated by inscription is tantamount to disciplining one's perception: coming to see the inscription as a mathematical marking consistent with disciplinary interpretations, rather than as a material object consistent with everyday interpretations. That such a specialized form of perception is required is evident in the confusions that even older students have about forms of notation, such as the graph of a linear function. For example, a student's interpretation of slope in a case study conducted by Schoenfeld, Smith, and Arcavi (1993) included a conception of the line as varying with slope, y-intercept and x-intercept. The result was that the student's conception of slope was not stable across contexts of use. Stevens and Hall (1998) traced the interventions of a tutor who helped an eighth-grade student who was working on similar problems involving interpretations of graphical displays. Their analysis focused on the tutoring moves that helped reduce the student's dependence on a literal grid representing Cartesian coordinates. Some of the teacher's assistance included literal occlusion of grid, a move designed to short-circuit the student's reliance on the grid in order to promote a disciplinary focus on ratio of change to describe the line. Similarly, Moschkovich (1996) examined how pairs of ninth-grade students came to discipline their own perceptions by coordinating talk, gestures, and inscriptions of slope and intercept. Inscriptions helped orient students toward a shared object of reference, and the use of everyday metaphors like hills and steepness grounded this joint focus of conversation. Yet, ultimately, the relative ambiguity of these everyday metaphors instigated (for some pairs) a more disciplined interpretation, because meanings for these terms proved ambiguous in the context of conversation.

Other work surveys a broader landscape. Sherin (2001) explored the implications of replacing algebraic notation with programming for physics instruction. Here again, notations did not simply describe experience for students, but rather reflexively constituted it. Programming expressions of motion afforded more ready expression of time-varying situations. This instigated a corresponding shift in conception from an algebraically guided physics of balance and equilibrium to a physics of process and cause.

DIGITAL NOTATIONS

The chief contribution of electronic technologies to mathematics is the development of new forms of notational systems, often characterized as dynamic (Kaput, 1992, 1998). The manifestations of dynamic notations are diverse, but they share in common an expression of mathematics as computation (Noss & Hoyles, 1996). DiSessa (2000) suggested that computation is a new form of mathematical literacy, concluding that computation, especially programming, "turns analysis into experience and allows a connection between analytic forms and their experiential implications…" (p. 34). Moreover, simulating experience is a pathway for building students' understanding, yet is also integral to the professional practices of scientists and engineers. For instance, descriptions of continuous change in a natural system, such as changes in biological populations over time, were once the exclusive province of differential equations. These described rates of change at an aggregate level. These changes can now be redescribed computationally as ensembles of independent agents acting in parallel according to a comparatively simple set of relations (Resnick, 1996; Wilensky, 1996, 2003; Wilensky & Reisman, 2006; Wilensky & Resnick, 1999). The behavior of a system, such as a biological population, relies on the nature of the interactions among the agents. Change in the aggregate (population dynamics) is now viewed as emergent, rather than as the starting point of mathematical description, as it is for differential equations.

Dynamic notational schemes are not mere re-expression of familiar mathematics in new notational garb. In contrast, they suggest new mathematical visions of familiar systems, such as the agent-based alternatives to describing change. Other familiar examples include the now canonical view of the geometry of the plane as including new mathematical elements, such as "stretchy" line segments, and new operations, such as "drag," that were afforded by the development of dynamic geometry tools, such as Geometer's Sketchpad (Goldenberg & Cuoco, 1998). From the perspective of Goodman's treatment of the qualities of representational systems, dynamic notational schemes preserve the digital character of systems of notation, while offering the analog qualities of ordinary scribbling. For example, consider stretching a line segment or dragging to transform one figure into another. The action produces many intermediate states, much like the sweeping hand of an analog watch, but when the action is terminated, what remains is an object with distinct properties. In the sections that follow, we explore the affordances of a dynamic notational system for supporting students' exploration of chance and data. Our intention is to illuminate how dynamic notations can support new forms of mathematical activity in this realm as well. To set the stage for this exploration, we begin by describing several scenarios for production and interpretation of the "same" distribution of data. Our aim is to ground later exposition of students' mathematical activity with a dynamic software, *TinkerPlots*, in the different senses of data supported by these contextual variations.

INSCRIBING DISTRIBUTION

Figure 4.1 is a stacked dot plot display of a 150 numeric values plotted along a continuous axis. We have omitted the label on the axis because we will use this display in several different contexts to demonstrate basic tools statisticians and others use to extract information from inscriptions of data.

In this inscription, each circle has a well defined referent (or it will when we specify a context). A circle is a distinct case with a numeric value, the value being indicated by the placement of a case along the axis. If this were a distribution of people's heights, then a particular circle would reference a particular individual's height. Circles in the same stack have the same value. These two characteristics—syntactic disjointedness and syntactic differentiation—qualify this inscription as a notional system.

If the only purpose of this inscription had been to associate a case with a value, a simple listing of cases paired with their values would have sufficed. The cases have been displayed along an axis as a frequency distribution to highlight features of the collection as a whole. These features include the general shape of the distribution, where it is centered, and how spread out it is. These properties are signifiers, but they do not point to the individual cases. Note that shape, center, and spread are not properties that belong to or are shared by the individual cases that make up the distribution. So the question is, what real-world objects

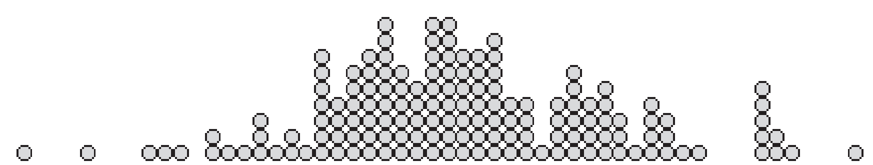

Figure 4.1 A distribution of values.

or phenomenon do the features of the distribution reference? The answer to this question depends very much on context, and contexts constrain and afford the interpretation of the distribution and its associated measures (i.e., statistics). We describe three contexts that we have employed with school-aged children, because although all use the same notational system, each offers different ways of thinking about distribution that we have found important for introducing students to statistics and data.

Three statistical contexts

Repeated measure

Consider the values in Figure 4.1 as 150 measurements of the circumference of a person's head, where a different person made each measurement. This context is inspired by an activity we have conducted several times with students in teaching experiments that we describe later. In this context, each individual measurement is an imperfect estimate of the value we want to know—the person's actual head circumference. We can therefore consider each measurement as comprising two components—the actual head circumference, and an error due to the measurement process.

Let us return to our question concerning the referential objects of the various properties of a distribution of values. In the case of repeatedly measuring a person's head circumference, the center of the distribution of measurements (the mean, say) references the actual head circumference, which is the quality we seek to establish. The spread of values around this average points to the error-prone process of various people measuring circumference with a particular tool. Errors are introduced for a number of reasons, including deficiencies in the measurement tool itself and in people's ability to use the tool and to correctly read values off of it. The relatively mound, or normal, shape of the distribution of measures tells us something about the measurement process as well. Knowing that measurement error is produced by a combination of independently acting factors, we expect the distribution of measures to be mound shaped. This is because mixed combinations of positive and negative error components are more likely than combinations comprising mostly positive (or negative) ones. If the distribution were relatively flat, we would be suspect of the data, perhaps concluding that they had been faked. If the distribution were skewed to one side or the other, it would suggest bias in the measurement method, which would lead us, again, to distrust our estimate of head-circumference based on the mean of the data.

In summary, statisticians see the distribution of measured values in this context as comprising two components: signal and noise (Konold & Pollatsek, 2002). The signal, perhaps best represented by the mean in this example, is an estimate of the person's true head circumference. The noise is the variability around this true value, which results from multiple sources of error in the measurement process. We use this particular distribution to make inferences about the process that produced it. The same process, if repeated, would result in a different collection of values, and different distribution. From this perspective, even the noise in this particular distribution can be considered a signal at another level. It suggests an emerging shape (here, a normal distribution), which we would expect would persist in future samplings and which supports our viewing these values as products of an unbiased measurement process.

A production process

Now imagine that the distribution in Figure 4.1 is made up of measurements of the diameter of bolts coming off of a factory's production line. In this case, what do the various properties of the distribution of these measures reference? Because we cannot measure anything precisely, some of the variability in the distribution of diameter measurements would still be due to measurement error. However, because it is impossible to produce anything exactly the same

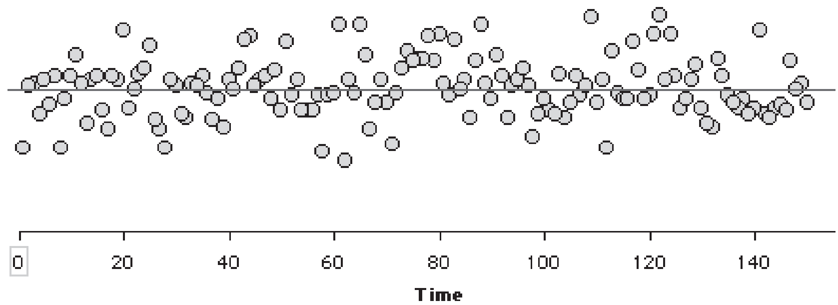

Figure 4.2 Time series display of 150 measurements from a production process. The line indicates the target specification of the product.

way twice, the bolts themselves differ in diameter. Assuming the measurement error is small compared to the actual differences in bolt diameters, the variability among values in our distribution primarily reference the inconsistency of the production process, a feature that manufactures of products are concerned with reducing and keeping in control. In this context, we would still expect the distribution of values to be basically mound shaped, again because of the multiple independent factors that are involved in their production. If this variability were too large, we might attempt to reduce it by fine-turning the manufacturing process, but we would expect this to narrow the distribution, not to alter its basic shape.

What does the mean of this distribution of measurements point to? In this case, it is not the actual measurement of a bolt. As we said, individual bolts are of slightly different diameters. We might, however, consider the mean to be the target diameter that the current manufacturing process is shooting for. And if this average is close to the specs the designers of the bolt established, and the variability around that value is also within acceptable tolerances, then we would regard the manufacturing processes as being in adjustment. If, on the other hand, the average of the distribution of values was relatively far away from the designers specs, or the variability was too large, then we might interrupt the manufacturing process and call in the maintenance crew. Indeed, in this context, seeing the diameters displayed as above in a frequency graph would not allow us to easily detect such adjustment problems as they occurred. Better would be a "control chart" display as shown in Figure 4.2, which shows the measurements over time. If the bolt manufacturing process went out of adjustment, we would eventually observe the average drifting way from the ideal, or the spread in the values increasing.

Natural variation

Finally, what do the distribution features reference if the measurements in Figure 4.1 are the heights of individual plants of the same species? In this case, some of the variability among the individual values is again due to measurement error. But as in the manufacturing context above, much of the variability reflects real differences in the heights of individual plants, differences that we could easily perceive if we could look at all the plants together. Compared to these real differences in heights, the differences among individuals that are due to measurement error are likely to be miniscule and thus of comparatively little import. What does the mean of this distribution reference? In this case, it is difficult to point to anything in the world that the mean corresponds to. It is neither a true score, nor a specification set by some designer.

As we move from the context of measurement error to that of individual differences, it becomes more difficult to link statistical characteristics of the distributions to real-world referents (Konold & Pollatsek, 2002; Lehrer & Schauble, 2007). This observation is borne out in the history of the development of statistics. In the context of repeated measures, the mean was first used in the late 1500s as a way of estimating the actual position of a star

(Plackett, 1970). But it was over 200 years later before it was used with individual difference data (Quetelet, 1842). To provide a conceptual bridge between the idea of mean of a bunch of error-prone measurements and the mean of a group of individuals who differed, Quetelet invented the construct of the "average man," and interpreted the mean as describing this imaginary entity. In contrast, a focus on variability in this context affords a prospective entrée to population thinking in biology. Population thinking emphasizes variability as the building block of biological diversity, as in Darwin's proposal of natural selection operating to bias the direction of variability. Variability in turn emerges from processes of genetic recombination (predominantly) during reproduction, so that evolutionary-based explanations rely on coordination between generation of variability and its alteration via selection. This explanation relies on reasoning about distribution, so that unlike the previous two contexts, characterizing variability and changes in variability constitute the cornerstone of explanation. Variability is necessary, not nuisance. (A lack of variability inevitably leads to extinction.)

SUPPORTING CONCEPTUAL DEVELOPMENT WITH DYNAMIC SOFTWARE

In this section, we describe how use of dynamic software can mediate student learning in each of these contexts. Our description relies on a series of classroom teaching studies, where we have introduced students to a sequence of problems and to dynamic software as a tool for implementing runnable models that offer both solutions and suggest explanations. In our view, the objects that students build with this tool, and the inscriptions they create to organize and explore the output, are dynamic forms of mathematical expression which give rise to and facilitate their thinking about the domain, and that these ideas would not be readily available to them if they were restricted to purely written symbolic forms of mathematics.

Repeated measure

In a number of classroom teaching experiments (Petrosino, Lehrer, & Schauble, 2003; Lehrer, Konold, & Kim, 2006; Lehrer, Kim, & Schauble, 2007), we have introduced students to data analysis through repeated measure. As we have explained, contexts of repeated measure facilitate the conceptual separation of signal (the actual measure of the object) from noise (measurement error). In these contexts, the signal in a distribution of measures corresponds to a measure of center, while noise corresponds to the variability among individual measurements.

In the example we discuss here, fifth graders measured the circumference of their teacher's head. Each student measured first with a 15 cm. ruler and then with a 75 cm. ruled tape measure (Lehrer, Konold, & Kim, 2006). The pedagogical intention was to create a prospective bridge between qualities of activity (i.e., measuring with the two different tools) and qualities of the resulting distribution. Hence, qualities of distribution, such as the greater variability obtained with the less precise tool, could prospectively be related to one's experience using this tool. After collecting these measurements, students input them into *TinkerPlots*, a data analysis tool designed specifically for upper elementary and middle school students (Konold & Miller, 2004; Konold, 2007).

A unique feature of *TinkerPlots* is that it has no straightforward way of making standard displays. Rather, students organize the data in small steps using operators including "stack," "order," and "separate." Figure 4.3 shows the 66 student-collected measurements of head circumference as they initially appeared in *TinkerPlots* after entering them. On the left is a stack of 66 data cards into which the students recorded their data. Each card holds an individual measurement of *Circumference*. The additional attribute named *Tool* indicates whether the measurement was made with the precise instrument (the tape measure) or the crude one

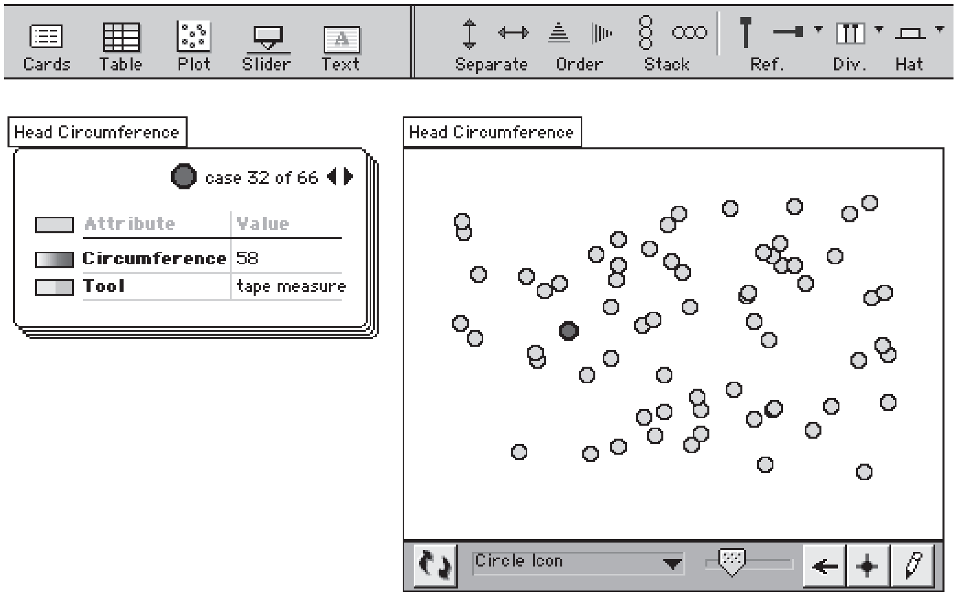

Figure 4.3 TinkerPlots representation of 66 repeated measurements of the circumference of the teacher's head. The data card on the left shows the values of the case that is highlighted in the plot object on the right.

(the 15 cm. ruler). In the plot object at the right of Figure 4.3, each case appears as a small circle. Initially, the cases appear as they do here, haphazardly arranged. It is the students' task to organize the icons into a display that helps them answer their question. In this activity, the primary objective for the students was to use the various measurements to estimate the circumference of the teacher's head. We anticipated that students would develop indicators of the center of the respective distributions as a way to estimate from the data the actual head circumference. A second objective was to develop a measure of the precision of the measurements that would allow students to compare quantitatively the precision of the two different measurement techniques.

There are multiple ways students might proceed to produce displays in *TinkerPlots* that would help answer these questions. Figures 4.4–4.8 show one set of possibilities. To get the graph in Figure 4.4, the icons were first colored according to their value on the attribute *Circumference*. The darker hue an icon (in color these would appear blue), the larger the value for *Circumference*. Coloring the icons in this way sets the stage for separating or ordering them.

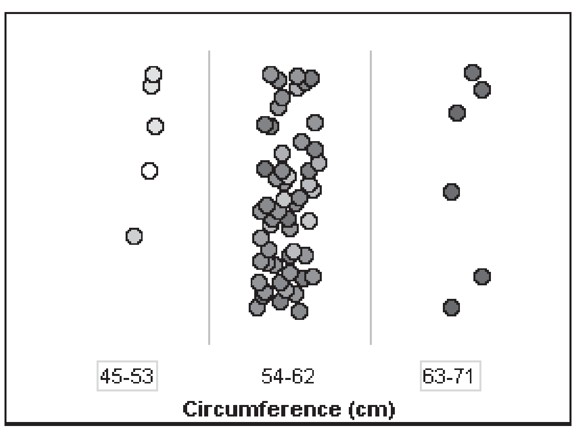

Figure 4.4 Representing magnitude with color hue and saturation. Here, the darker an icon, the larger its value on the attribute, *Circumference*. Dynamic separation of cases affords thinking about groups.

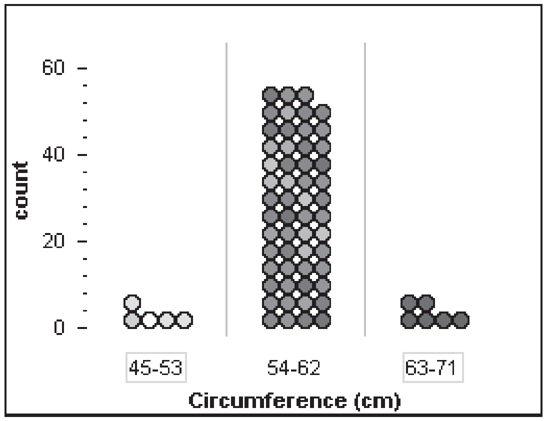

Figure 4.5 Stacking cases facilitates comparing bin sizes and emphasizes aggregate structure.

In this example, they were then separated into three bins by pulling one of the icons to the right.

In Figure 4.5 cases have been stacked. In Figure 4.6 the data appear as a stacked dot plot as a result of continuing to separate them into smaller and smaller intervals until the icons are positioned over their exact value of *Circumference* on a number line. To explore whether one set of measurements are more accurate than the other, the icons were then colored according to their values on the attribute *Tool*. In Figure 4.7, they have then been separated along the vertical dimension into these two groups to facilitate visual comparison.

By using these operators in combination, students express notationally what they otherwise would do with their hands. It may appear that *TinkerPlots* is a simple instantiation of the sort of physical actions one uses when building graphs from physical objects representing cases (Harradine & Konold, 2006). But because actions in *TinkerPlots* are separated into their constituent components and are explicitly named, graphing is lifted out of the plane of sheer activity. As a result, students can specify differences between two representations in terms of the actions they would use to change one representation into the other. The dynamics of pulling on cases and re-arranging them are complemented by a notational result: the displays represent a snapshot of the results of an activity, such as stacking.

Looking at the graphs in Figure 4.7, students quickly pointed out salient features and difference. They described the data collected with the ruler as being more "spread out," and pointed out that in each set of measurements, most of the values were around 55 to 60 centimeters. To help clarify what in the graphs they were noticing, students quickly learned

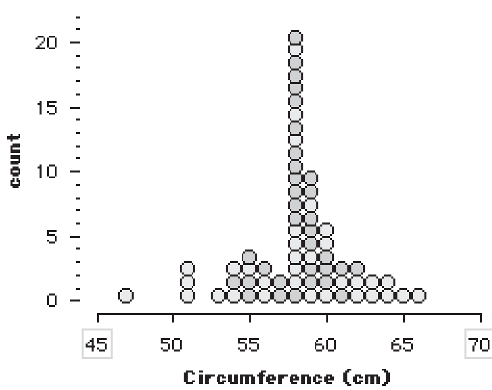

Figure 4.6 Continued dynamic pulling on one of the icon results in a continuous plot. The different colors (which appear here as different values of grey) represent two different measurement tools.

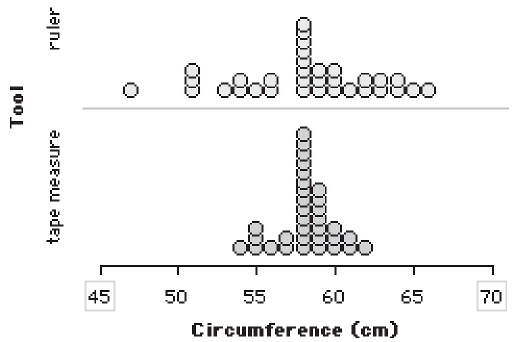

Figure 4.7 Pulling up separated into groups vertically, revealing a difference in the variability of the measurements made with the two different tools.

to use a number of additional display options in *TinkerPlots*. The dividers were perhaps the most commonly used enhancement of displays of this sort (see Figure 4.8). Our sense is that students key on these, because dividers are well suited to their proclivity to perceive and talk about pieces of distributions (Konold, Higgins, Khalil, & Russell, 2004; Lehrer & Schauble, 2002, 2004). With them, students can indicate precisely the location of subsets of the data they think are important. To the divisions they create they can add the count (and/or percent) of the cases within each subsection, sparing them the task of counting them (which they will do).

Though students begin using the tools to mark and communicate about subsets of the data, the tools also support students as they begin to notice and talk about aggregate features, such as the proportion of the data represented in an interval. Dividers become an especially powerful way for students to begin communicating more explicitly about where they perceive the data as being centered and eventually lead them to using measures such as the median and mean with an intuitive sense of what these measures are representing about the data (Konold, Kazak, Lehrer, & Kim, 2007; Petrosino, Lehrer, & Schauble, 2003).

Measuring signal and noise

Because students perceived that the attribute being measured has a "real" value, we asked students to invent a measure that used the data in some way to estimate this value (Lehrer, Konold, & Kim, 2006). Most student inventions were oriented toward the center clump of

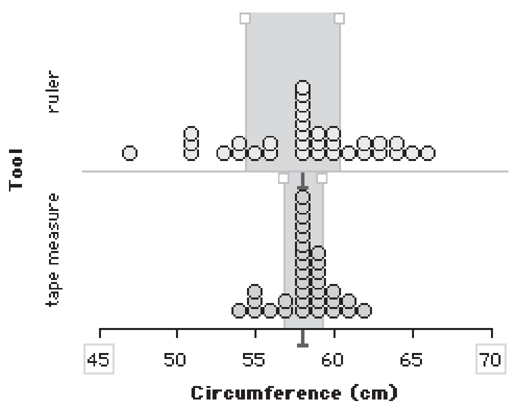

Figure 4.8 Adjustable dividers support reasoning about relative density within and between distributions. Inverted T's indicate the location of sample medians.

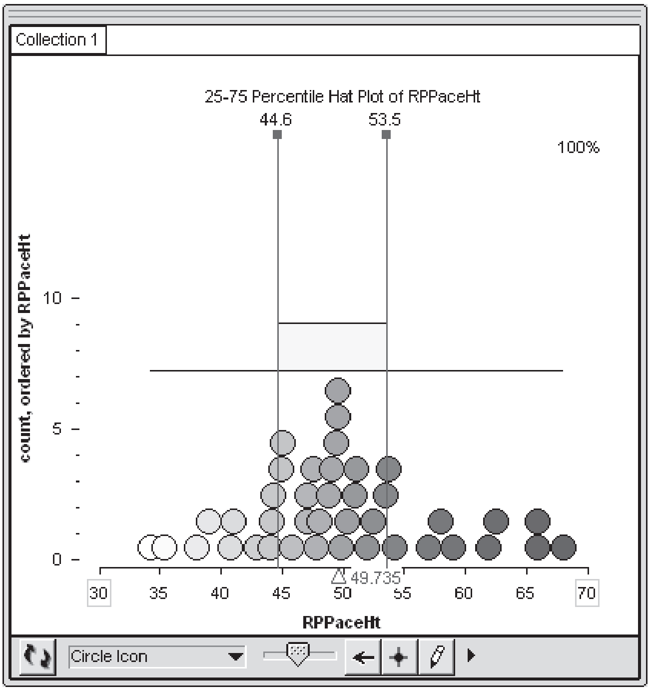

Figure 4.9 A 25–75 percentile hat plot along with references lines used by a sixth grade student to indicate the precision of a collection of measurements as the range of the middle 50%. The triangle under the axis shows the location of the mean at 49.7.

values, such as those suggested in Figure 4.7 and Figure 4.8. It was fairly commonplace for at least one group of students in every iteration of the design to relate this clump to their sense that middle means "half." Half is an important step toward inventing the median or a closely related measure of central tendency.

Students also considered noise. For example, Figure 4.9 displays one sixth-grader's approach to characterizing the precision of measurements conducted by his class to measure the height of the school's flag-pole (in feet). Henry attempted to measure precision by considering proximity of the observed measurements to the center. After first exploring the data with the flexible dividers displayed in Figure 4.8, he used an innovation introduced by *TinkerPlot*—a "hat plot" (see Konold, 2007). His invented statistic works in the sense that it allows Henry to compare the relative precisions obtained with different tools (his class used manufactured and home-made tools). Furthermore, it corresponds to conventions in the field, although Henry was not aware of these conventions when he invented this measure.

Modeling signal and noise

Because students can readily associate qualities of the distribution to measurement processes they experienced, and to their previous efforts to invent measures of signal and noise, the repeated measurement context affords an introduction to modeling chance inherent in the measurement process. We engaged students in a discussion of the nature of the "mistakes" made as they measured. With this as background, students developed descriptions of sources of error, their likelihood, and their relative magnitudes. For example, when measuring the circumference of their teacher's head, many students noticed that they had difficulty bending the ruler to conform to the shape of the head. Due to this and the length of the ruler, students had to iterate the ruler to obtain a measurement. Errors of overestimation occurred when the ruler "lapped" so that some regions of the head were measured twice. Errors of under-estimation resulted when the ruler "gapped" so that some regions of the head were not

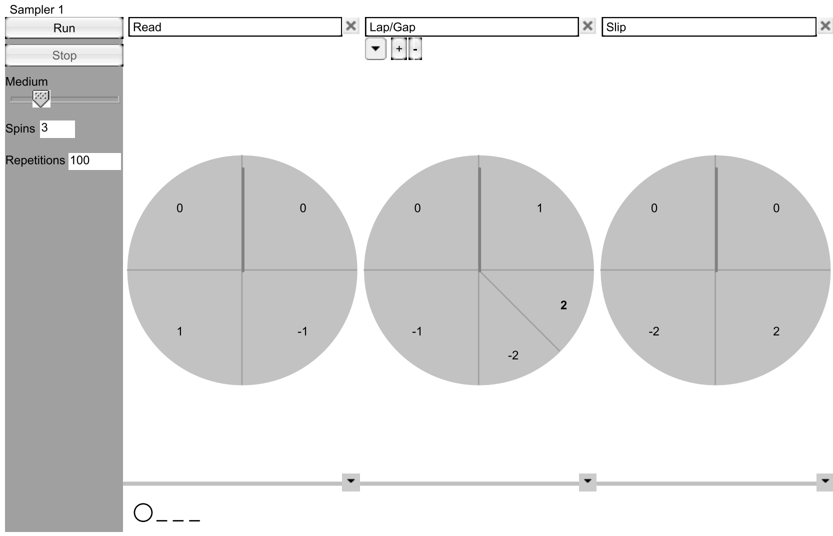

Figure 4.10 *A* facsimile of a model of errors of measure developed by a pair of fifth-grade students. "Read" are positive and negative errors due to inaccuracies in reading values of the ruler. "Lap/Gap" are errors due to difficulties iterating the ruler. "Slip" are errors due to slippery contact of the ruler with the head. When the student presses the Run botton, a value from each spinner is randomly sampled. The sum of these three values gives the measurement error for a single measure. The model is set up to produce 100 such measurements.

included in the measurement. Gapping and lapping were errors that no student was able to control (although some tried mightily), and we exploited this understanding to introduce the notion that each source of error could be modeled by the behavior of a chance device, such as a spinner (Lehrer, Konold, & Kim, 2006). Conceptually, the distribution of measures could be approximated by adding the contributions of different sources of error to the "best guess" of the true value of the biometric measure.

A new version of *TinkerPlots* under development (version 2.0) supports this form of modeling by allowing students to design and run virtual random devices (spinners and urns). Figure 4.10 displays a facsimile of a model of random error developed by a pair of fifth-grade students who measured the circumference of their teacher's head with a 15 cm. ruler. They decided to represent three sources of error with spinners for each. These corresponded to various misadventures with using a ruler to measure head circumference. For example, "Lap/Gap" represented the difficulties of trying to iterate a ruler when the surface was curved, related to the earlier classroom discussions about gapping and lapping that we noted previously. Positive magnitudes indicated overestimates, where the observed value was more than the true value. Negative magnitudes indicated underestimates, where the observed value was less than the true value. The areas of the spinners correspond to student estimates of the relative likelihood of each magnitude.

The results of one of their simulations of 100 measurers is displayed in Figure 4.11. These values were obtained by summing the simulated errors and adding to this error the true value of the height. The true value was approximated by the sample median of the class measurements (58 cm.). Having built this model, students used it repeatedly to create potential distributions of measurements. They noted basic features that remained consistent over successive runs of their model—notably that most distributions of 100 simulated measurements had the same basic shape, and that the centers were fairly consistent as well. Based on these observations, the students judged this model as "good" because it produced distributions of measurements whose shape resembled that of the distribution of their actual measurements.

To test their understanding of how the relative magnitude and direction of error played a role in creating the shape of the distribution, their instructor challenged them to create a "bad" model of the observed measurements. The results of one of their efforts, dubbed a truly

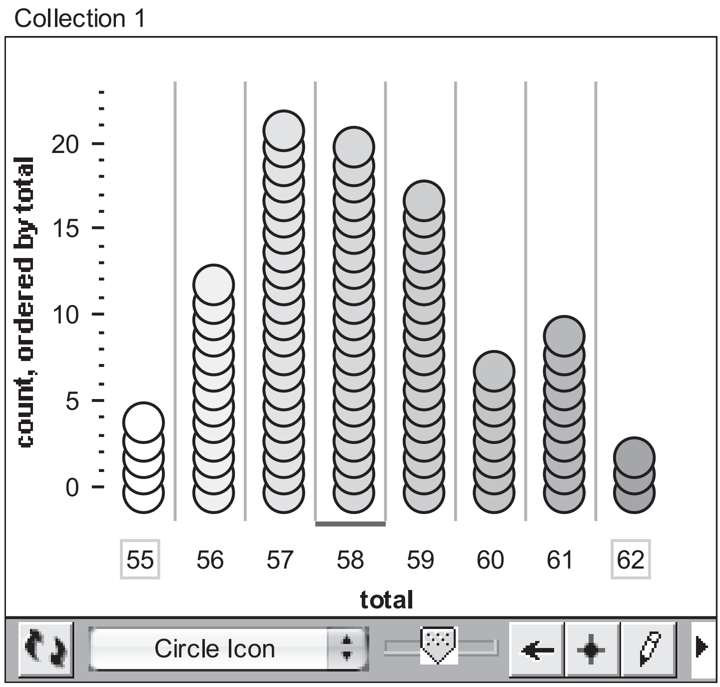

Figure 4.11 One run of a simulation of observed measures expressed as the sum of true score (estimated by the sample median) and the sum of the errors of measure from each source of error in the model shown in Figure 10. The bar at the bottom of the 58 bin indicates the location of the median in this simulated run.

"rotten" model, is displayed in Figure 4.12. It was accomplished by emphasizing a greater chance of errors of comparatively large magnitude along with a relatively small chance of errors of comparatively small magnitudes. Again, the basic shape of the distribution produced by this bad model was determined not from a single running of the model, but from multiple runnings.

Reprise

We employ a context of repeated measure to introduce students to data and statistics. Students use the dynamic notations available in *TinkerPlots* to structure and represent data. As a result, they typically notice that their measurements, despite their variability, tend to cluster

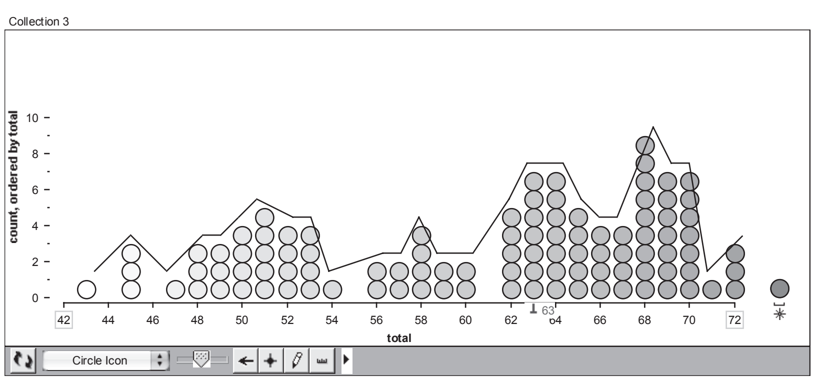

Figure 4.12 The results of one simulation of a "bad" model of the repeated measure of the circumference of the teacher's head. The bad model produced measurements that tended to bunch up at the higher end. The students who built the model added a line to the plot of the data which connected the tops of the stacks to help emphasize the distribution's characteristic shape.

in roughly the middle of the distribution of values. We help students relate these data characteristics to their experiences measuring, and ask them to invent measures corresponding to what conventionally are called statistics of center and spread. In this context, measures of center are estimates of the true value of the measured attribute, and measures of spread are estimates of the precision of measure. These measures (i.e., statistics) are supported by *TinkerPlots* tools that allow students to readily partition and repartition the data, and to think relationally about these partitions (generally, by displaying percent or proportions of data within particular regions). Grappling with the data in this way puts students in a position to develop additional explanatory accounts of these data: Observed measures are compositions of signal and random error. To model this composition, a development version of *TinkerPlots* 2.0 mimics the behavior of random devices, represented iconically by spinners or urns. These devices allow students to simulate chance processes and to observe the resulting distribution of outcomes. Hence, the context of repeated measures affords students a gentle introduction to central concepts in statistics, especially the ideas of distribution and their characteristics (including measures of center and spread) and ideas of chance variation.

Designed objects

We suggested earlier that a manufacturing context provides a rich conceptual middle ground between variability in repeated measures and variability of naturally occurring objects. Along with Anthony Harradine, we recently conducted a weeklong teaching experiment using such a context with students aged 13–14 at Prince Alfred College, a boys-only private school in Adelaide, Australia. For 5 days we turned their classroom into a factory producing "fruit sausages," small cylindrical pieces of Play Dough that ideally were to measure 5cm. in length and 1 cm. in diameter.

On the first day, each of the 12 students made five sausages, rolling them by hand to the desired diameter, cutting them to the desired length, and then weighing them on a balance beam scale. Analysis of the data using *TinkerPlots* showed the weights and lengths to be quite variable, which set the stage for introducing a pressing device which students could use to squeeze out a 1 cm. diameter length of material. Using this device, they repeated the production process and then analyzed their results.

During the analysis of both production processes, the need for describing the centers and the spread of their data naturally emerged. For describing centers, they began using center clumps and eventually gravitated to medians. To motivate a precise, and agreed upon, measure of spread, we offered a prize to the group who had produced the most consistent product. With guidance, they finally settled on using the absolute values of distances from medians to individual observations, a measure which turned out to fit well with their informal judgments of variability based on observing graphs of the data.

We have yet to systematically analyze the videotapes made of these classroom sessions, but our sense is that we replicated most of the findings reported with repeated measures above. In particular, having participated themselves in the process of making the sausages, and watching others in the classroom using slightly different techniques, the students were in a position to provide rich accounts of, and explanations for, the variability in sausage sizes. Furthermore, they spontaneously used the centers of their distributions to judge how well their process was performing relative to the specifications they were given.

The production process may have several additional points of pedagogical leverage. First, in measurement contexts, although students have readily identified individual sources of error, estimating the magnitudes of the effects of each source of error is more difficult. For example, to estimate the error due to gaps and laps, students iterated a ruler a number of times and estimated about the number of cm. of error that might typically result from a gap or a lap. The quantification of error in a production process is simpler in that one can repeat an undesirable aspect of the process and determine its magnitude. Second, the measurement context presumes that students have a firm grasp of the nature of scale and unit, and of course, many students do not, despite years of schooling (Lehrer, 2003). Finally, the manufacturing context, because it produces multiple individuals that vary, seems to us to be conceptually closer to the context of naturally occurring objects and thus may provide for an easier transition to viewing such data statistically.

Natural variation

Most of the important questions we want to answer with statistics involve neither contexts of repeated measurement nor of control processes in industry. We might want to know how tall a particular type of tree tends to grow, whether one medical procedure is better than another, or if scores on NAEP have increased over the past four years. In these contexts, it is only with reluctance that many students use formal averages to summarize or compare groups (Gal, Rothschild, & Wagner, 1990; Roth & Bowen, 1994; Watson & Moritz, 1999; Bright & Friel 1998; Konold & Higgins, 2003). We believe that the major reason for this is that in these contexts it is conceptually difficult to make sense of averages and of group variability.

Over the past several years, many researchers have been studying how we might help students think about such data and to use, when appropriate, measures of fit and spread to summarize such data or to compare groups. For example, Lehrer and Schauble (2004) described the emergence of population thinking in a class of fifth-grade students who observed changes during the course of a plant's life cycle in the distribution of the heights of a sample of 63 plants. Students constructed variations of bootstrap samples to consider questions about what might happen if plants of the same species were grown again. Repeated sampling proved an especially effective entrée to considering sampling distribution of statistics, such as sample medians. Intuitions that students developed about sample-to-sample variation later guided their inferences about the effects of light and fertilizer on plant height. For example, some students used the results of sampling studies to suggest an informal band of confidence around the median of plants grown with "regular" fertilizer and then made an inference about the effect of higher levels of fertilizer by locating relative to that band the median height of plants grown with more fertilizer. (They were disappointed that their conjecture that more fertilizer would lead to greater plant heights was not confirmed.) However, students with prior experience with repeated measures also seemed to recapitulate the historic difficulty of mapping interpretations of distribution from measure to natural variability (Lehrer & Schauble, 2007). As one student noted, an interpretation of the median or mean as a "typical" plant was not sensible, because "they're all typical!"

In light of this difficulty, we have been exploring how we might employ the modeling approach, which we have used successfully in repeated measures context, to helping student build a statistical view of naturally occurring data as composed of separable components. In these investigations, we have worked to support the development of aggregate (or global) methods of groups comparison (Ben-Zvi & Arcavi, 2001; Cobb, 1999; Lehrer & Romberg, 1996; Lehrer & Schauble, 2004; Rubin & Hammerman, 2006) that would allow students to decide whether and by how much two groups differed by comparing the locations of the distributions' centers. But rather than have students work primarily with samples of real data, where we can never know for sure whether two groups differ, we employ computer-generated data and pose problems as mysteries to solve and for which, afterwards, they can see whether or not they were correct.

Below, we present snapshots of the reasoning of three students, two sixth-grade boys (Nelson and Byron), and an eighth-grade girl (Erin). The students were participating in a yearlong teaching experiment conducted during an after-school program in Holyoke, Massachusetts. As part of the after-school program, these students, along with the other nine participants in the after-school program, had been using the development version of *TinkerPlots* 2.0 to model a range of probabilistic situations. They had routinely built computer models of various situations,

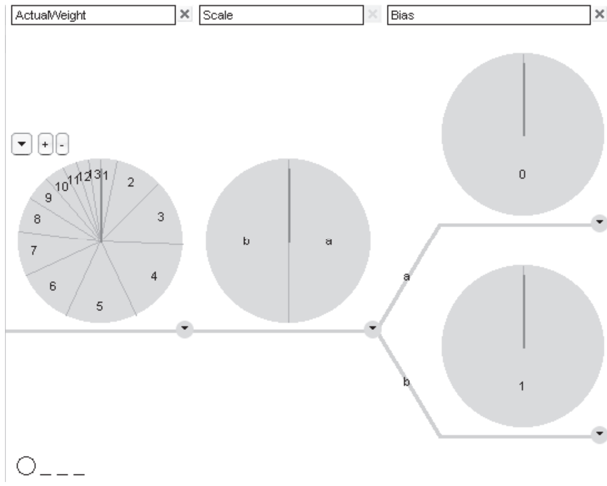

Figure 4.13 A model of the weights of letters weighed on two scales at a post office. A letter is first assigned an actual weight, ranging from 1 to 13 ounces, from the spinner on the far left. It then goes to either scale *a* or scale *b* (center spinner) to be weighed. Letters weighed on scale *a* are accurately weighed (bias of 0). Letters weighed on scale *b*, however, have an additional ounce added on (bottom right spinner).

graphed and analyzed the results of repeated experiments, and discussed the variability and persistent trends resulting from those experiments (see Konold, Kazak, Lehrer, & Kim, 2007).

Near the end of the after-school program, we involved these three students in three one-hour sessions, spread over consecutive weeks. The sessions were designed to help them come to view one distribution of data as a linear shift of another distribution, where the degree of that shift could be estimated by computing the differences between the groups' centers. This basically is how such data are viewed from the statistical perspective of the General Linear Model.

It was on the third day of the intervention that we introduced students to the additive shift idea. This began with a class discussion of weighing letters at a post office. After exploring a single distribution of letter weights, which ranged from 1 to 13 ounces, we told students that at the post office in question, there were two scales. One scale weighed accurately, but the other was not properly adjusted and added one ounce to the actual weight of the letter. As a class, we explored the *TinkerPlots* version of this situation in which the biased scale was represented as a spinner that added the constant +1 to the actual letter weight and the fair scale was a spinner that added a zero (see Figure 4.13).

Students looked at the resultant distributions of weight to see if they could detect how the distributions were different for the two scales. We then explored what those distributions would look like if the biased scale added 2, 4, and finally –4 ounces. Following this introduction, we progressed to a problem about pumpkin weights, presented as a mystery factory within *TinkerPlots* (see Figure 4.14). In this case, which was isomorphic in structure to the problem about letter weights, we kept hidden from students the value of the constant that was being added or subtracted to a group of fertilized pumpkins. The students' task was to guess the mystery constant from the samples we drew.

After the students had worked individually and written responses to the question, we displayed the graphs in *TinkerPlots* on a Smart Board and asked the students "How many more pounds do you think that the fertilized pumpkins tend to weigh than the non-fertilized?" Nelson gave the answer of 8 pounds, explaining that "from 15 to about … 23 is the average [of the fertilized group]" while the non-fertilized group "would be around 17 to 11." As he gave these ranges, we placed dividers at those locations on the plot. To explain how he got 8 from these, he took control of *TinkerPlots* and showed that 8 was the distance between the two dividers on the top graph. As he said this, however, he also adjusted the lower dividers

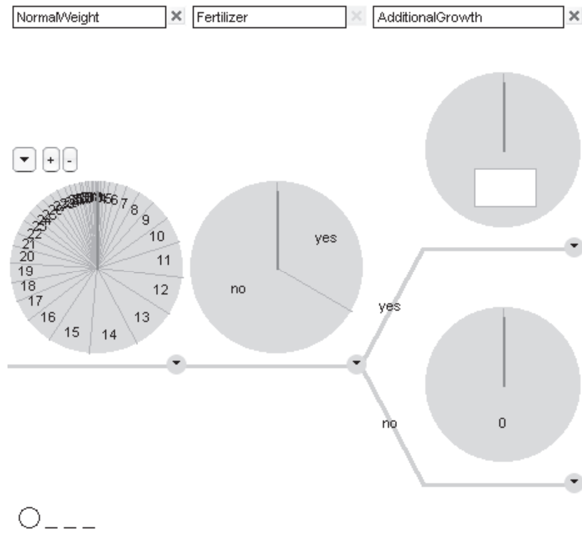

Figure 4.14 A model of weights of two types of pumpkins where students must guess the effect of fertilizing the pumpkins by analyzing data from the model. Pumpkins are randomly assigned a weight in pounds (left spinner). One third of these are then fertilized (middle spinner). The fertilized pumpkins received an unknown but constant number of pounds (spinner value masked). The students' task is to guess the value in the mystery spinner based on analysis of data drawn from the Sampler.

on the fertilized group, making it equal to the upper divider in the non-fertilized group (see Figure 4.15). After making this adjustment, he said the difference was about 6 pounds. Note that this corresponds to the difference between the right-hand dividers in each group. Our conclusion is that Nelson used a valid global method to compare the groups.

Byron then came to the board to help explain why he thought Nelson's answer of 6 made sense. He added a reference line to each distribution, placing each almost precisely in the middle of the range of each modal clump (see Figure 4.15). This was a use of reference lines

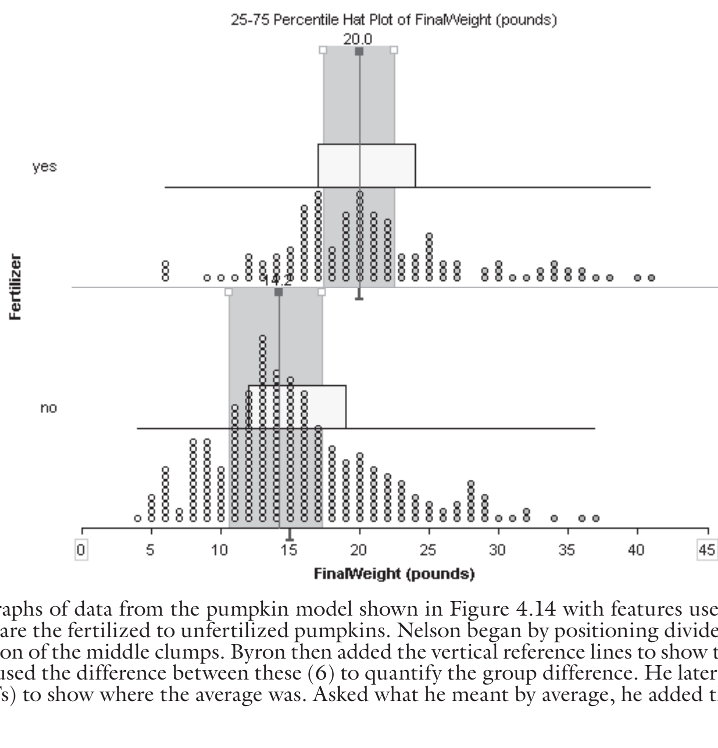

Figure 4.15 Graphs of data from the pumpkin model shown in Figure 4.14 with features used by Nelson and Byron to compare the fertilized to unfertilized pumpkins. Nelson began by positioning dividers (grey areas) to show the location of the middle clumps. Byron then added the vertical reference lines to show the "center of the averages" and used the difference between these (6) to quantify the group difference. He later added the medians (inverted Ts) to show where the average was. Asked what he meant by average, he added the hat plots.

we had not demonstrated to students nor seen them use before. He then counted from the lower reference line to the higher one to get a difference of 6. This was different than the answer of 4 pounds he had written on his worksheet, which he had gotten by determining the distance between the modes of the two groups. Asked to explain the placement of the reference lines, Byron responded:

Byron: Because it's the middle of the average.…
I: What's the average, then?
Byron: [Adds hat plots to the graph.] That.
I: What does the hat plot show?
Byron: Sort of where the middle is, kind of.

Byron, we believe, was here conceiving the modal clump as an average and was using dividers and hat plots interchangeably to mark the locations of these averages. The idea of marking the "middle of the average" with the reference line was presumably motivated by the need to get a single-value difference between the two groups.

Erin then came to the board and demonstrated her solution, which used a cut point at 19. She expressed the difference as a range, which she got by computing the differences from the cut point of 19 to the minimum of the lower group and to the maximum of the upper group. She explained that "if you take one of the smaller numbers [points to values on the lower end of the non-fertilized group], right, and then add it to sev…, you get one of these numbers [points to values above cut point in fertilized group]." Unfortunately, she had made an error in subtracting which gave her 7 rather than 17 as the lower range. But we think she was employing a shift idea here, imagining what it would take to shift values below the cut point in the non-fertilized group to values above the cut point in the fertilized group.

Following Erin's presentation, Byron asked that we add the median to the plots he had made. When we did so, the students were impressed by how close the medians were to the reference lines Byron had placed "in the middle of the averages" (see Figure 4.15). Nelson responded, "That's so cool. He was right on it for the first one (fertilized group)." Byron, we believe, had constructed an understanding of a median he had not previously held. It was no longer simply the result of a procedure of finding the middle score, but an indicator of the center of a distribution. Its meaning, we believed, derived from linking it conceptually to the modal clump, which he had been using previously as an indicator of center.

To capitalize on Byron's move, we concluded by saying that "some people" compare groups by comparing the medians, and get the difference between the two groups by taking the difference in the medians. We then computed the difference in the medians to get 5. We also showed that the difference in the upper crown edges of each hat plot was 5, as was the difference between the lower crown edges. We then opened the mystery spinner to reveal a value of 5. We had two objectives in mind in comparing a variety of features. The first was to deter the students from coming to regard the median as some magical indicator. The other was to build on the idea of distribution shift—that adding a 5 in our model shifted not only the median, but the entire distribution.

After this discussion, students worked individually at the computer using *TinkerPlots* to analyze simulated data on a parallel problem involving the number of minutes it took for two different headache remedies to have an effect. The task was to decide whether one pill formula worked faster than the other, and if so by how much. All three students used aggregate methods in addressing these questions.

Nelson, for example, constructed two graphs of data from the model, placed dividers to show modal clumps, added medians to each group, and aligned split references lines with the medians to help display their values. Thus, he used everything except the hat plots, which he had just seen Byron use in the classroom discussion. As he was writing his conclusion based on the representations he had made, we probed:

I: Explain to me what you've got.
Nelson: I got the time difference showing, 'cause I used the median between the thing. This is the old pill. This is the new pill.

We might view Nelson as simply repeating what he had seen Byron do in front of the class. But our sense is that he was not simply mimicking—that he was appropriating what he had seen because it made sense to him. The prior classroom discussion had included the presentation and discussion of several approaches to this problem. Nelson was choosing from these which to adopt. Furthermore, in his verbal responses to our queries, Nelson showed a reasonable understanding of the comparison task and how the tools he was using addressed them.

I: And why did you use the medians?
Nelson: To find where the middle would be and to find the difference between the two.

Another indication of use with understanding is his elaborated conclusion, something we had not worked on in a concerted way.

I: And what do you conclude?
Nelson: [as he is writing] I conclude that the new pill is about [stops writing and looks at graphs] six minutes more, the new pill would take six minutes earlier than the old pill.

Based on follow-up interviews with these students, we are encouraged about the viability of this modeling approach. Note that it begins by reversing the direction of the inference that students of statistics are usually asked to make when they encounter group differences. Rather than reasoning backwards from observed differences to unknowable population parameters, we introduce it as reasoning from a given factory set up (as in the simulation of weighing letters), to samples drawn from that factory. Furthermore, when it comes time for students to reason from samples to population parameters (in this case represented as a constant in a spinner), the value they are trying to guess is ultimately revealed. This allows them to evaluate the effectiveness of their reasoning and build up confidence in it. A third feature of the contexts we used was that the averages of the single distributions were not, of themselves, of particular interest or importance. Rather, they were useful only in trying to estimate what the constant difference between the groups was. In a prior session, students had discovered that centers of distributions (as indicated by the location of the modal clump or the median) were more stable properties of samples than, say, extreme values. Thus, we had prepared them to view centers as reasonable things to compare.

We should also add that we carefully selected the contexts we presented to students so that it was relatively easy to imagine one distribution as being a linear shift of another (e.g., applying fertilizer might cause all pumpkins to grow larger by a certain amount). We have not tested how well students reason about situations when the idea of distribution shift must be applied only metaphorically. For example, it seems a stretch to describe male heights as being female heights shifted to the right. Given that most contexts are more like the male/female heights than they are like the biased/fair scale, an important next step in determining the viability of this approach is to test it with such contexts.

In summary, we expect that the data factory metaphor will have a profound effect on how students think about real data and thus how they construe the activity of inference. Typically, when we analyze data, we make guesses about the mechanisms (or population parameters) that may have created them. We notice a small gender difference in a sample and ask ourselves whether such a difference exists in the population. Thus, we must reason backwards from effects to possible causes. We believe that the ability to create virtual data puts students in a more natural orientation, allowing them first to attend to causal mechanisms (or parameters), and then afterwards explore how those mechanisms influence the resultant data. We expect that with sufficient experience operating on data from this perspective that students will come to view real data in the same way, as having been produced by a collection of independently acting mechanisms (or a parent population) that it is their job to make guesses about based on the data they have.

CONCLUSION

In all three of the contexts that we have described, dynamic notations alter the conceptual landscape of data modeling. In the cases that we have described, the digital character of the notational system allowed students to build mathematical objects that served to model the behavior of the systems they investigated. The objects could be inspected and modified, and these inspections and modifications supported the development of new forms of mathematical reasoning. The dynamic character of the software enabled students to mimic physical activity, the kind that occurs continuously in space and time. This coupling of digital and analog meant that students now had the means to dynamically generate chance phenomena and to produce variations of these phenomena by altering the generating system, including both chance and non-chance components.

This capability alters the traditional dynamic between model and world, where models are generally developed as accounts of the world. Instead, as suggested by Kaput (1998), it is possible to envision a role reversal, in which simulation leads and "reality" follows. Such deliberate simplification lies at the heart of the epistemology of modeling. Our narrative is but an exemplar of a more general proposition: A new way of writing mathematics affords a new way of reasoning about a mathematical system. Here, students have new means to investigate chance, data, and their coordination.

NOTE

1. We both contributed equally to this work, which was supported by grants from the National Science Foundation, REC 0337675 and ESI 0454754. The views expressed do not necessarily represent those of the Foundation.

REFERENCES

Ben-Zvi, D., & Arcavi, A. (2001). Junior high school students' construction of global views of data and data representations. *Educational Studies in Mathematics*, *45*, 35–65.

Berlinski, D. (2000). *The advent of the algorithm. The 300-year journey from an idea to the computer.* New York: Harcourt.

Bright, G. W., & Friel, S. N. (1998). Helping students interpret data. In S. P. Lajoie (Ed.), *Reflections on statistics: Learning, teaching, and assessment in grades K–12* (pp. 63–88). Mahwah, NJ: Erlbaum.

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

diSessa, A. (2000). *Changing minds. Computers, learning, and literacy.* Cambridge, MA: MIT Press.

Gal, I., Rothschild, K., & Wagner, D. A. (1990). Statistical concepts and statistical reasoning in school children: Convergence or divergence. Paper presented at the annual meeting of the American Educational Research Association, Boston, MA.

Goldenberg, E. P., & Cuoco, A. (1998). What is dynamic geometry? In R. Lehrer & D. Chazan (Eds.), *Designing learning environments for developing understanding of geometry and space* (pp. 351–367). Mahwah, NJ: Erlbaum.

Goodman, N. (1976). *Languages of art.* Indianapolis, IN: Hackett Publishing.

Hall, R. (1990). Making mathematics on paper: Constructing representations of stories about related linear functions. Unpublished doctoral dissertation, University of California at Irvine.

Hall, R. (1996). Representation as shared activity: Situated cognition and Dewey's cartography of experience. *The Journal of the Learning Sciences, 5*(3), 209–238.

Harradine, A., & Konold, C. (2006). How representational medium affects the data displays students make. In A. Rossman & B. Chance (Eds.) *Proceedings of the 7th International Conference on Teaching Statistics (ICOTS)* [CD-ROM]. Voorburg, The Netherlands: International Statistical Institute.

Izsak, A. (2000). Inscribing the winch: Mechanisms by which students develop knowledge structures for representing the physical world with algebra. *The Journal of the Learning Sciences*, *9*(1), 31–74.

Kaput, J. (1991). Notations and representations as mediators of constructive processes. In E. von Glasersfeld (Ed.), *Constructivism and mathematics education* (pp. 53–74). Dordrecht, The Netherlands: Kluwer.

Kaput, J. (1992). Technology and mathematics education. In D. A. Grouws (Ed.), *Research on mathematics teaching and learning* (pp. 515–556). New York: Macmillan.

Kaput, J. (1998). Representations, inscriptions, descriptions and learning: A kaleidoscope of windows. *Journal of Mathematical Behavior, 17*(2), 265–281.

Kaput, J., & Shaffer, D. (2002). On the development of human representational competence from an evolutionary point of view. In K. Gravemeijer, R. Lehrer, B. van Oers, & L. Verschaffel (Eds.), *Symbolizing, modeling and tool use in mathematics education* (pp. 277–293). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Konold, C. (2007). Designing a data tool for learners. In M. Lovett & P. Shah (Eds.), *Thinking with data* (pp. 267–291). New York: Taylor & Francis.

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., Higgins, T., Khalil, K., & Russell, S. J. (2004). *Data seen through different lenses.* Amherst, MA: University of Massachusetts. http://www.umass.edu/srri/serg/papers/index.html.

Konold, C., Kazak, S., Lehrer, R., & Kim, M-J. (2007). *To understand a distribution, try building it from scratch.* Paper presented at the Fifth International Research Forum on Statistical Reasoning, Thinking, and Literacy (SRTL-5), Warwick, UK.

Konold, C., & Miller, C. (2004). *TinkerPlots™ Dynamic Data Exploration* (Version 1.0). Emeryville, CA: Key Curriculum Press.

Konold, C., & Pollatsek, A. (2002). Data analysis as the search for signals in noisy processes. *Journal for Research in Mathematics Education, 33*(4), 259–289.

Latour, B. (1990). Drawing things together. In M. Lynch & S. Woolgar (Eds.), *Representation in scientific practice* (pp. 19–68). Cambridge, MA: MIT Press.

Lehrer, R. (2003). Developing understanding of measurement. In J. Kilpatrick, W. G. Martin, & D. E. Schifter (Eds.), *A research companion to principles and standards for school mathematics* (pp. 179–192). Reston, VA: National Council of Teachers of Mathematics.

Lehrer, R., Kim, M., & Schauble, L. (2007). Supporting the development of conceptions of statistics by engaging students in modeling and measuring variability. *International Journal of Computers in Mathematics Education, 12*(3), 195–216.

Lehrer, R., Konold, C., & Kim, M. (2006, April). *Constructing data, modeling chance in the middle school.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Lehrer, R., & Lesh, R. (2003). Mathematical learning. In W. Reynolds & G. Miller (Eds.), *Comprehensive handbook of psychology* (Vol. 7; pp. 357 –391). New York: Wiley.

Lehrer, R., & Romberg, T. (1996). Exploring children's data modeling. *Cognition and Instruction, 14*(1), 69–108.

Lehrer, R., & Schauble, L. (2002). Symbolic communication in mathematics and science: Co-constituting inscription and thought. In E. D. Amsel & J. Byrnes (Eds.), *Language, literacy, and cognitive development. The development and consequences of symbolic communication* (pp. 167–192). Mahwah, NJ: Erlbaum.

Lehrer, R., & Schauble, L. (2004). Modeling natural variation through distribution. *American Educational Research Journal*, *41*(3), 635–679.

Lehrer, R., & Schauble, L. (2007). Contrasting emerging conceptions of distribution in contexts of error and natural variation. In M. Lovett & P. Shah (Eds.), *Carnegie symposium on cognition: Thinking with data* (pp. 149–176). New York: Erlbaum.

Meira, L. (2002). Mathematical representations as systems of notations-in-use. In K. Gravemeijer, R. Lehrer, B. van Oers, & L. Verschaffel (Eds.), *Symbolizing, modeling and tool use in mathematics education* (pp 87–103). Dordrecht, The Netherlands: Kluwer Academic.

Meira, L. (1995). The microevolution of mathematical representations in children's activity. *Cognition and Instruction*, *13*, 269–313.

Moschkovich, J. N. (1996). Moving up and getting steeper: Negotiating shared descriptions of linear graphs. *The Journal of the Learning Sciences, 5*(3), 239–277.

Munn, P. (1998). Symbolic function in pre-schoolers. In C. Donlan (Ed.), *The development of mathematical skills* (pp. 47–71). East Sussex, UK: Psychology Press, Taylor & Francis.

Nemirovsky, R., & Monk, S. (2000). "If you look at it the other way...": An exploration into the nature of symbolizing. In P. Cobb & E. Yackel & K. McClain (Eds.), *Symbolizing and communicating in mathematics classrooms. Perspectives on discourse, tools, and instructional design* (pp. 177–221). Mahwah, NJ: Erlbaum.

Nemirovsky, R., Tierney, C., & Wright, T. (1998). Body motion and graphing. *Cognition and Instruction, 16(*2), 119–172.

Noss, R., & Hoyles, C. (1996). *Windows on mathematical meaning.* Amsterdam, The Netherlands: Kluwer Academic.

Petrosino, A., Lehrer, R., & Schauble, L. (2003). Structuring error and experimental variation as distribution in the fourth grade. *Mathematical Thinking and Learning, 5*(2, 3), 131–156.

Plackett, R. L. (1970). The principle of the arithmetic mean. In E. S. Pearson & M. G. Kendall (Eds.), *Studies in the history of statistics and probability* (pp. 121–126). London: Charles Griffen & Company.

Quetelet, M. A. (1842). *A treatise on man and the development of his faculties.* Edinburgh: William and Robert Chambers.

Resnick, M. (1996). Beyond the centralized mindset. *The Journal for the Learning Sciences, 5*, 1–22.

Roth, W.-M. & Bowen, G. M. (1994). Mathematization of experience in a grade 8 open-inquiry environment: An introduction to the representational practices of science. *Journal of Research in Science Teaching*, *31*, 293– 318.

Rotman, B. (1993). *Ad infinitum.* Stanford, CA: Stanford University Press.

Rubin, A., & Hammerman, J. K. (2006). Understanding data through new software representations. In G. Burrill & P. C. Elliott (Eds.), *Thinking and reasoning with data and chance: 68th NCTM Yearbook* (pp. 241–256). Reston, VA.: National Council of Teachers of Mathematics.

Schoenfeld, A. H., Smith III, J. P., & Arcavi, A. (1993). Learning: The microgenetic analysis of one student's evolving understanding of a complex subject matter domain. In R. Glaser (Ed.), *Advances in instructional psychology* (pp. 55–175). Hillsdale, NJ: Erlbaum.

Sherin, B. L. (2001). A comparison of programming languages and algebraic notation as expressive languages for physics. *International Journal of Computers for Mathematical Learning*, *6*, 1–61.

Stevens, R., & Hall, R. (1998). Disciplined perception: Learning to see in technoscience. In M. Lampert & M. L. Blunk (Eds.), *Talking mathematics* (pp. 107–149). Cambridge, UK: Cambridge University Press.

Thompson, P. W. (1992). Notations, conventions, and constraints: Contributions to effective use of concrete materials in elementary mathematics education. *Journal for Research in Mathematics Education, 23*(2), 123–147.

van Oers, B. (2000). The appropriation of mathematical symbols: A psychosemiotic approach to mathematics learning. In E. Y. P. Cobb & K. McClain (Eds.), *Symbolizing and communicating in mathematics classrooms: Perspectives on discourse, tools, and instructional design* (pp. 133–176). Mahwah, NJ: Erlbaum.

van Oers, B. (2002). The mathematization of young children's language. In K. Gravemeijer, R. Lehrer, B. van Oers, & L. Verschaffel (Eds.), *Symbolizing, modeling, and tool use in mathematics education* (pp. 29–58). Dortrecht, The Netherlands: Kluwer Academic.

Watson, J. M., & Moritz, J.B. (1999). The beginning of statistical inference: Comparing two data sets. *Educational Studies in Mathematics*, *37*, 145–168.

Wilensky, U. (1996). Modeling rugby: Kick first, generalize later? *International Journal of Computers for Mathematical Learning, 1*, 125–131.

Wilensky, U. (2003). Statistical mechanics for secondary school: The GasLab multi-agent modeling toolkit. *International Journal of Computers for Mathematical Learning, 8,* 1–41.

Wilensky, U., & Reisman, K. (2006). Thinking like a wolf, a sheep, or a firefly: learning biology through constructing and testing computational theories—An embodied modeling approach. *Cognition and Instruction, 24*(2), 171–209.

Wilensky, U., & Resnick, M. (1999). Thinking in levels: A dynamic systems approach to making sense of the world. *Journal of Science Education and Technology, 8*(1), 3–19.