

Chapter 7

LEARNING TO REASON ABOUT DISTRIBUTION

Arthur Bakker and Koeno P. E. Gravemeijer
Freudenthal Institute, Utrecht University, the Netherlands

OVERVIEW

The purpose of this chapter is to explore how informal reasoning about distribution can be developed in a technological learning environment. The development of reasoning about distribution in seventh-grade classes is described in three stages as students reason about different representations. It is shown how specially designed software tools, students' created graphs, and prediction tasks supported the learning of different aspects of distribution. In this process, several students came to reason about the shape of a distribution using the term bump along with statistical notions such as outliers and sample size.

This type of research, referred to as “design research,” was inspired by that of Cobb, Gravemeijer, McClain, and colleagues (see Chapter 16). After exploratory interviews and a small field test, we conducted teaching experiments of 12 to 15 lessons in 4 seventh-grade classes in the Netherlands. The design research cycles consisted of three main phases: design of instructional materials, classroom-based teaching experiments, and retrospective analyses. For the retrospective analysis of the data, we used a constant comparative method similar to the methods of Glaser and Strauss (Strauss & Corbin, 1998) and Cobb and Whitenack (1996) to continually generate and test conjectures about students' learning processes.

DATA SET AS AN AGGREGATE

An essential characteristic of statistical data analysis is that it is mainly about describing and predicting aggregate features of data sets. Students, however, tend to conceive a data set as a collection of individual values instead of an aggregate that has certain properties (Hancock, Kaput, & Goldsmith, 1992; Konold & Higgins, 2002; Ben-Zvi & Arcavi, 2001; Ben-Zvi, Chapter 6). An underlying problem is that middle-grade students generally do not see “five feet” as a value of the variable

“height,” but as a personal characteristic of, say, Katie. In addition to this view, students should learn to disconnect the measurement value from the object or person measured and consider data against a background of possible measurement values. They should furthermore develop a notion of distribution, since that is an organizing conceptual structure with which they can conceive the aggregate instead of just the individual values (Cobb, 1999; Petrosino, Lehrer, & Schauble, 2003).

These learning goals formed the motivation to explore the possibilities for students in early secondary education with little or no prior statistical knowledge to develop an informal understanding of distribution. Such understanding could then be the basis for more formal statistics in higher grades. The main question in this study is therefore: How can seventh-grade students learn to reason about distribution in an informal way?

DISTRIBUTION

To answer this question, we first analyze the relation between data and distribution. Distinguishing between data as individual values and distribution as a conceptual entity, we examine aspects of both data sets and distributions such as center, spread, density, and skewness (Table 1). Measures of center include mean, median, and midrange. Spread can be quantified with, for instance, range, standard deviation, and interquartile range. The aspects and measures in the table should not be seen as excluding each other; outliers and extreme values, for instance, influence skewness, density, spread, and even most measures of center.

Table 1. Between data and distribution

distribution (conceptual entity)			
center mean, median, midrange, ...	spread range, standard deviation, inter- quartile range, ...	density (relative) frequency, majority, quartiles	skewness position majority of data
data (individual values)			

This structure can be read upward and downward. The upward perspective is typical for novices in statistics: Students tend to see individual values, which they can use to calculate, for instance, the mean, median, range, or quartiles. This does not automatically imply that they see mean or median as a measure of center or as representative of a group (Mokros & Russell, 1995; Konold & Pollatsek, Chapter 8). In fact, students need a notion of distribution before they can sensibly choose

between such measures of center (Zawojewski & Shaughnessy, 2000). Therefore, students need to develop the downward perspective as well: conceiving center, spread, and skewness as characteristics of a distribution, and looking at data with a notion of distribution as an organizing structure or a conceptual entity. Experts in statistics can easily combine the upward and downward perspectives. We might say that the upward perspective leads to a frequency distribution of a data set. In the downward perspective, we typically use probability distributions such as the normal distribution to model data.

The table shows that the concept of distribution has a complex structure, but this concept is also part of a larger structure consisting of big ideas such as variation and sampling (Reading & Shaughnessy, Chapter 9; Watson, Chapter 12). Without variation, there is no distribution, and without sampling there are mostly no data. We therefore chose to deal informally and coherently with all these big ideas at the same time with distribution in a central position. As Cobb (1999) notes, focusing on distribution as a multifaceted end goal of instruction might bring more coherence in the statistics curriculum. The question is how. Our answer is to focus on the informal aspects of shape.

The shape of a distribution is influenced by various statistical aspects. A high peak, for example, is caused by a high frequency of a certain class and long tails on the left or right with the hill out of center indicate skewed distributions. This implies that by reasoning with informal terms about the shape of a distribution, students may already reason with aspects of that distribution. And indeed, students in this study used informal words to describe density (crowded, empty, piled up, clumped, busy), spread (spread out, close together), and shape (hill, bump). If students compare the height distributions of two different grades, they might realize that the graphs have the same shape but are shifted in location (Biehler, 2001). And they might see that samples of different sizes still have similar shapes. We envisioned that reasoning with shapes forms the basis for reasoning about distributions.

METHODOLOGY AND SUBJECTS

To answer the main question of how students can develop a notion of distribution, we carried out developmental research, which is also called design research (Freudenthal, 1991; Gravemeijer, 1994; Edelson, 2002; Cobb & McClain, Chapter 16). Design research typically involves the design of instructional materials, teaching experiments, and retrospective analyses. In line with the principles of Realistic Mathematics Education (Freudenthal, 1991; Gravemeijer, 1994) and the National Council of Teachers of Mathematics (NCTM) Standards (2000), we looked for ways to guide students in being active learners dealing with increasingly sophisticated means of support.

To assist students in exploring data and developing the concept of distribution, we decided to use some specially designed Minitools (see Cobb, 1999). These web applets were developed by reasoning backward from the intended end goal of reasoning about distribution to possible starting points. One aspect of distribution,

shape, can be inferred from stacked dot plots. To understand what dots in a dot plot stand for, students need to realize that a dot represents a value on some variable. One way to help students develop this insight is to let them start with case-value bars, which range from 0 to the corresponding value on the horizontal axis. We presume that bars representing values are closer to students' daily life reality than dots on an axis, because they are used to bar graphs and because horizontal bars are natural ways to symbolize certain variables such as the braking distance of cars, the life span of batteries, or the wingspan of birds. For that reason, each case in Minitool 1 (Figure 1) is signified by a bar whose relative length corresponds to the value of the case, and each case in Minitool 2 (Figure 2) is signified by a dot in a dot plot.

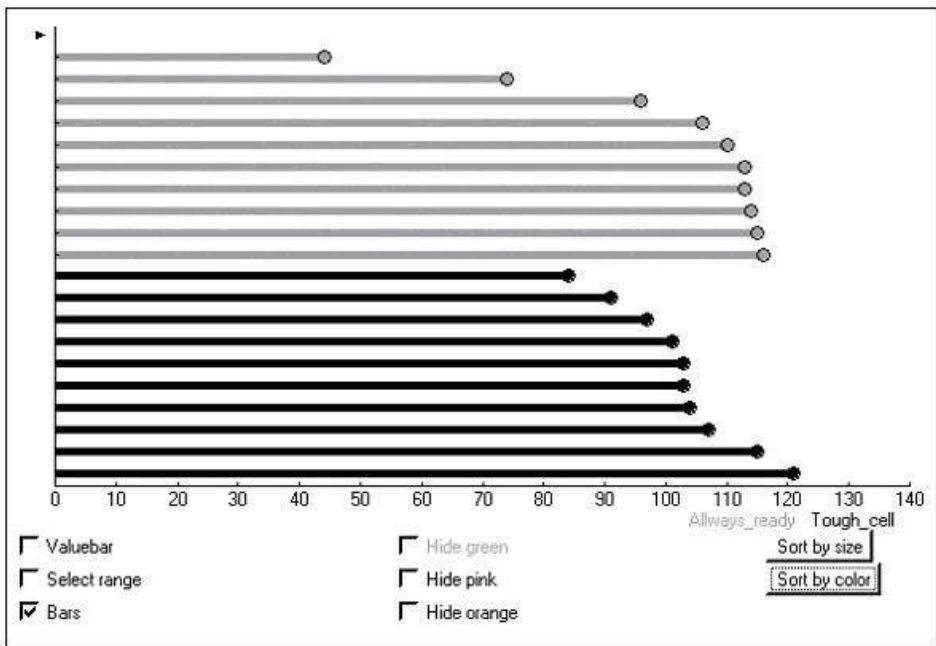


Figure 1. Minitool 1 (sorted by size and color).

To identify a baseline of what Dutch seventh-grade students already know about statistics and how easily they would solve statistical problems using the two Minitools, we interviewed 26 students about these issues. The students had encountered no statistics before except the arithmetic mean and bar graphs. They had almost no problems in reading off values from the Minitools, but they focused on individual data values (Section 2). We then did a small field test and conducted teaching experiments in 4 seventh-grade classes, which worked through a complete sequence of 12 to 15 lessons of 50 minutes each. The experiments were carried out during the school year 1999–2000, in a public school in a small town near Utrecht (the Netherlands) that prepared about 800 students for university (vwo) or higher

vocational education (havo). At that time about 15% of the Dutch students went to the vwo level, 20% to the havo level, about 40% to the mavo level (for middle vocational education), and the remaining 25% to lower vocational education (in the meantime the last two levels have been merged). These percentages indicate that the learning abilities of the vwo and havo students of our teaching experiments were above average.

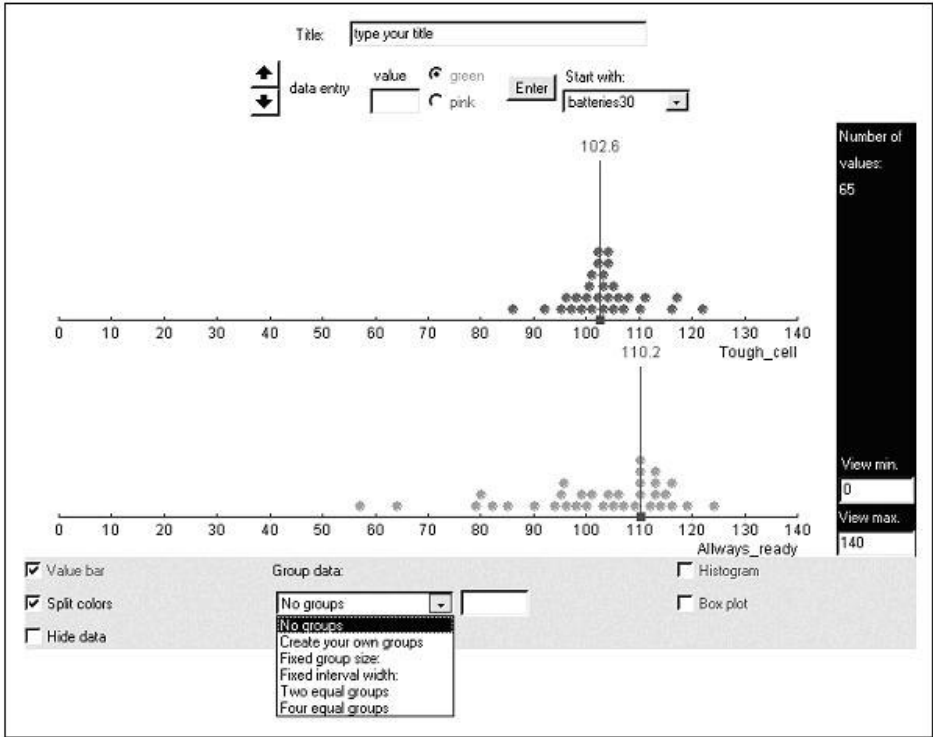


Figure 2. Minitool 2 (split colors and with vertical value bars).

The collected data include audio recordings, student work, field notes, and final tests in all classes, as well as videotapes and pretests in the last two experiments (see Table 2). The pretests were meant to find out if students already knew what we wanted them to learn (they did not).

An essential part of the data corpus was a set of mini-interviews that were held during lessons. Mini-interviews varied from about 20 seconds to 4 minutes and were meant to find out what concepts and graphs meant for the students. We realize that this influenced their learning, because the mini-interviews often stimulated reflection. In our view, however, the validity of the research was not in danger: Our aim was to find out how students could learn to reason with distribution, not whether teaching the sequence in other seventh-grade classes would lead to the same results.

For the retrospective analysis of the fourth teaching experiment, we have read the transcripts, watched the videotapes, and formulated conjectures on students'

learning based on the transcript and video episodes. The generated conjectures were being tested at the other episodes and the rest of the collected data (student work, field observations, and tests) in the next round of analysis (triangulation). Then the whole generating and testing process was repeated. This method resembles Glaser and Strauss's constant comparative method (Strauss & Corbin, 1998; Cobb and Whitenack, 1996). Important transcript fragments, including those in this chapter, have been discussed with colleagues (peer examination).

Table 2. Overview of subjects, teaching experiments, data collection, number of lessons, and levels of education

Subjects (grade 7)	Type of Experiment	Data Collection	No. of Lessons	Level
26 students (1999)	Exploratory interviews (15 minutes for two students)	audio	—	mavo, havo, vwo
Class A (25)	Exploratory field test	student work, final test, field notes, audio	4	havo
Class F (27)	First teaching experiment		12	vwo
Class E (28)	Second teaching experiment		15	vwo
Class C (23) (2000)	Third teaching experiment	idem plus pretest and video	12	havo
Class B (23)	Fourth teaching experiment		12	havo
12 classes (2000–2002)	Implementation	e-mail reports of two teachers, field notes from incidental visits	144	havo and vwo

Furthermore, we have identified patterns of student answers that were similar in all teaching experiments, and categorized the evolving learning trajectory in three stages according to students' reasoning with the representations used. The sections describing stages 1 and 2 describe observations that were similar for all four observed classes. In the first stage, students worked with graphs in which data were represented by horizontal bars (Minitool 1, Figure 1). In the second stage, from lesson 5 to 12, students mainly worked with dot plots (Minitool 2, Figure 2). In the third stage students used both Minitools and came to reason with bumps; the examples stem from the second teaching experiment. The students in this class had good learning abilities (vwo) and had 15 lessons—three more than in the other classes. The specific stages began to overlap each other when we started to stimulate comparison of different graphs during the last two teaching experiments.

STAGE 1—DATA ARE REPRESENTED BY BARS

The aim of the first activities was to let students reason about different aspects of distributions in an informal way such as about majority, center, extreme values,

spread-out-ness, and consistency. In the second lesson, for example, students had to prepare reports to Consumer Reports (a consumers' journal) on the quality of two battery brands. They were given a data set of 10 battery life spans of two brands in Minitool 1; using different computer options, they could sort the data and split the data, for instance of the two brands. In the beginning they used the vertical value bar (Figure 3) to read off values, but later sometimes to estimate the mean visually.

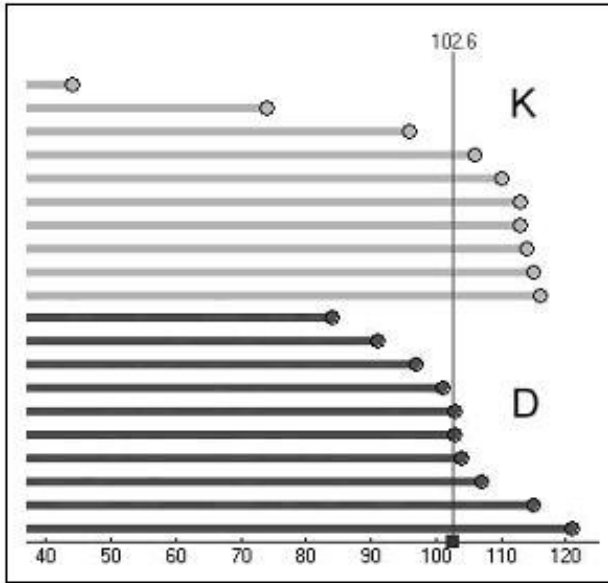


Figure 3. Estimating the mean of brand D with the movable vertical value bar (life span in hours).

During this battery activity, students in all teaching experiments could already reason about aspects of distributions. “Brand K has outliers, but you have more chance for a good one,” was one answer. “Brand D is more reliable, since you know that it will last more than 80 hours,” was another. This notion of reliability formed a good basis for talking about spread. Our observations resemble those of Cobb (1999) and Sfard (2000), who analyzed students’ spontaneous use of the notion of “consistency.”

The activities with Minitool 1 afforded more than informal reasoning about majority, outliers, chance, and reliability; they also supported the visual estimation of the mean (Figures 3 and 4). After this strategy had spontaneously emerged in the exploratory interviews, we incorporated instructional activities to evoke this strategy in other classes as well (Bakker, 2003). Minitool 1 supported the strategy with the movable vertical value bar. Students said that they cut off the longer bars, and gave the bits to the shorter bars. Several students in different classes could explain that this approach was legitimate: The total stays the same, and the mean is the total

divided by the number. When students said that brand D is better because its mean is higher, they used the mean to say how good the brand is. In that case, the mean is not just a calculation on a collection of data, but refers to a whole subset of one brand. As we intended, they learned to use the mean as a representative value for a data set and to reason about the brand instead of the individual data values.

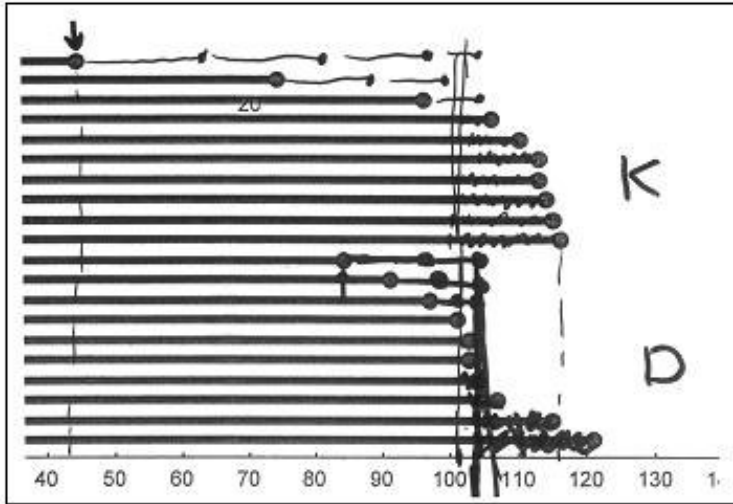


Figure 4. Scribbles on a transparency during class discussions after estimating means of both brands. The mean of brand D is slightly higher than that of K.

To assess students' understanding of distribution aspects and to establish a tighter relationship between informal statistical notions and graphs, we decided to "reverse" this battery task. In the last two teaching experiments, during the fourth lesson, we asked students to invent their own data according to certain characteristics such as "brand A is bad but reliable; brand B is good but unreliable; brand C has about the same spread as brand A, but it is the worst of all brands." Many students produced a graph similar to the one in Figure 5 (in this case, the variation of C is less than that of A). A sample response was:

Why is brand A better. Because it lives long. And it has little spread. Brand B is good but unreliable. Because it has much spread. But it lives long. Brand C has little spread but the life span is not very long.

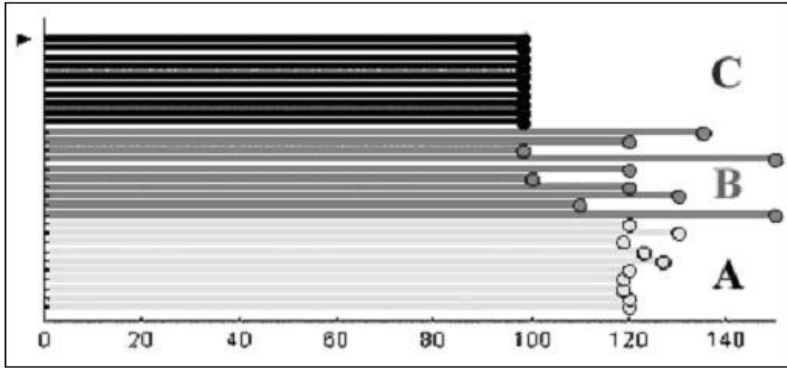


Figure 5. Invented data set according to certain features: Brand A is bad but reliable; brand B is good but unreliable; brand C has about the same spread as brand A, but it is the worst of all.

With hindsight, we have come to see this back-and-forth movement between interpreting graphs and constructing graphs according to statistical notions as an important heuristic for instructional design in data analysis, for a number of reasons:

- Students can express ideas with graphs that they cannot express in words (Lemke, 2003). If students invent their own data and graphs, teachers and researchers can better assess what students actually understand.
- If students think of characteristics such as “good but not reliable,” the lack of data prevents them from focusing on individual data, because it is cognitively impossible to imagine many individual data points. With this reverse activity, we create the need for a conceptual unity that helps in imagining a collection of data with a certain property. The notion of distribution serves that purpose (Section 3).
- In many schoolbooks, students mainly interpret ready-made graphs (Friel et al., 2001; Moritz, Chapter 10). And if students have to make graphs, the goal is too often just to learn how to produce a particular graph. De Lange, Burrill, Romberg, & van Reeuwijk (1993) and Meira (1995) strongly recommend letting students invent their own graphs. We may assume that students’ own graphs are meaningful and functional for them.
- The importance of the back-and-forth movement between data and graphs (or different graphs) is also indicated by the research on symbolizing. Steinbring (1997), for example, distinguishes *reference* systems and *symbol* systems. Students interpret a symbol system in the light of a better-known reference system. Reference systems are therefore relatively well known and symbol systems relatively unknown. In learning the relationship between a symbol system and a reference system, students must go back and forth between the two systems. A next step can then be that students use the symbol system they have just learned to reason with (Minitool 1, for example) as a reference system for a new symbol system (Minitool 2, for example), and so on.

From the examples of the first stage, it is clear that students informally reasoned about different aspects of distribution from the very start. They argued about the mean (how good the battery is), spread (reliability), chance for outliers or extreme values, and where the majority is (skewness). Without the bar representation the students would probably not have developed a compensating strategy for finding the mean. Their reasoning, however, was bound to one representation and two contexts.

STAGE 2—DOTS REPLACE BARS

Our next aim was to let students reason about shapes of distributions in suitable representations and in different contexts. Additionally, we strove for quantification of informal notions such as frequency and the majority and to prepare students for using conventional aggregate plots such as histograms and box plots.

As mentioned in the previous section, Minitool 1 can be seen as a reference system for the new symbol system of Minitool 2. When solving problems with Minitool 1, the students reasoned with the endpoints of the bars. In Minitool 1, students could hide the bars, which they sometimes preferred, because “it is better organized.” The dot plot of Minitool 2 can be obtained by hiding the bars of Minitool 1 and imaginatively dropping the endpoints on the horizontal axis or on the other dots that prevent them from dropping further down (cf. Wilkinson, 1999). Note that the dots are stacked and do not move sideways to fill up white areas in the graph (Figure 6). The advantages of this dot plot representation are that it is easy to interpret, it comes closer to conventional representations of distributions than Minitool 1, and students can organize data in ways that come close to histogram and box plot, for instance.

Minitool 2 has more options to organize data than Minitool 1. Apart from sorting by size and by subgroup (color), students can also group data into their own groups, two equal groups (for the median), four equal groups (for a box plot, Figure 7a), equal interval width (for a histogram, Figure 7b), and fixed group size (Figure 6b). This last option turned out to be useful for stimulating reasoning about density.

A particular statistical problem that students solved with Minitool 2 was the one on jeans sizes. Students had to report to a factory the percentage of each size that should be made, based on a data set of the waist measurements (in inches) of 200 men. This activity, typically done during the ninth lesson, was meant to distract students’ attention away from the mean and toward the whole distribution. Furthermore, it could be an opportunity to let students reason about absolute and relative frequencies.

We expected that students would reason about several aspects of distribution when comparing different grouping options. The option of fixed group size (Figures 6b and 6c) typically evoked remarks such as “with the thin ones [the narrow bins] you know that there are many dots together.” We interpret such expressions as informal reasoning about density, which we see as a key aspect of distribution. Many students used the four equal groups option to support their conclusion that “you have to make a lot of jeans in sizes 34–36, and less of 44–46.” Generally, a

skeptical question was needed to provoke more exact answers: “If the factory hired you for \$1,000, do you think the factory would be satisfied with your answer?” Most students ended up with the fixed interval option and a table with percentages, that is, relative frequencies.

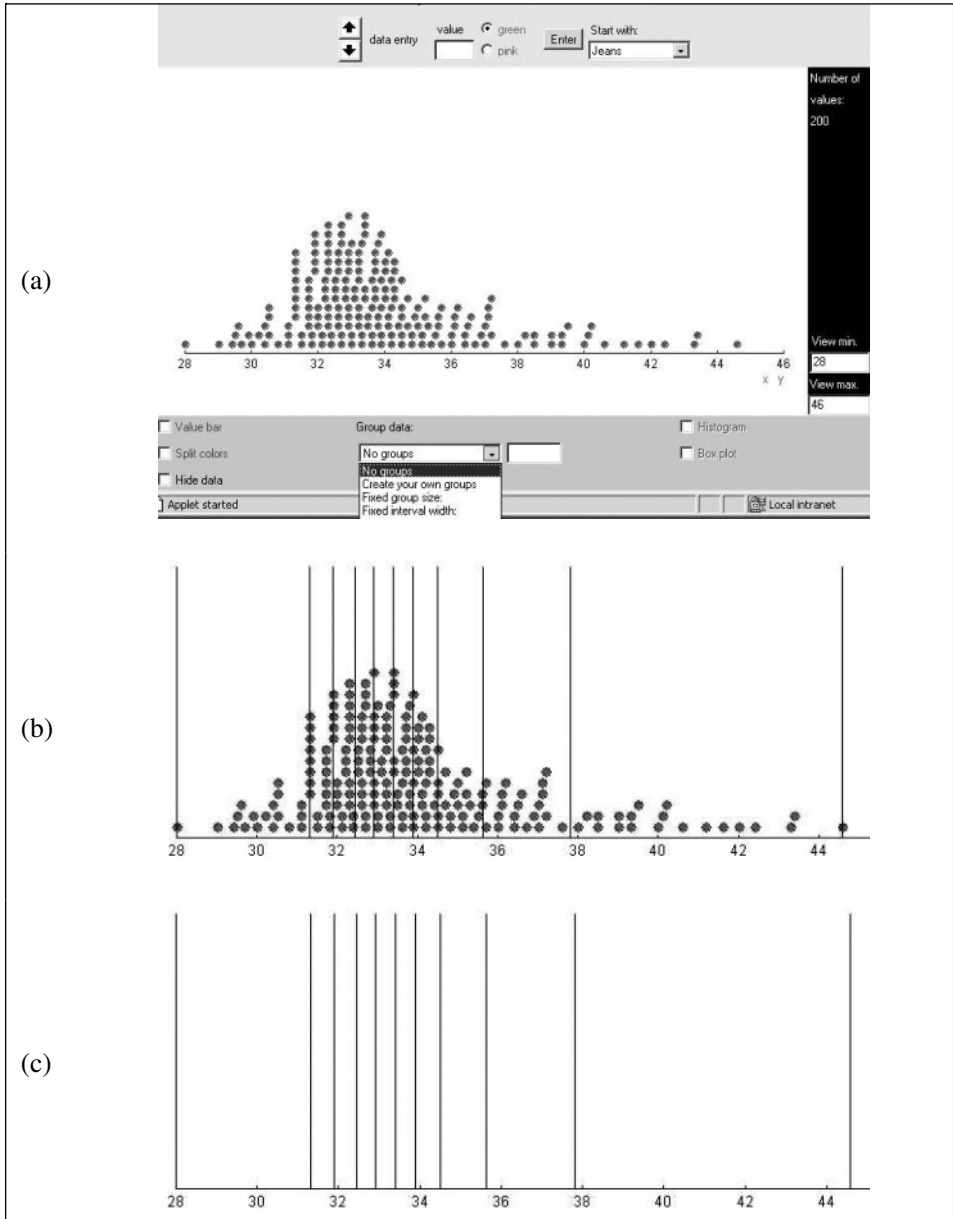


Figure 6. (a) Minitool 2 with jeans data set (waist size in inches, $n = 200$). (b) Fixed group size with 20 data points per group. (c) Minitool 2 with “hide data” function.

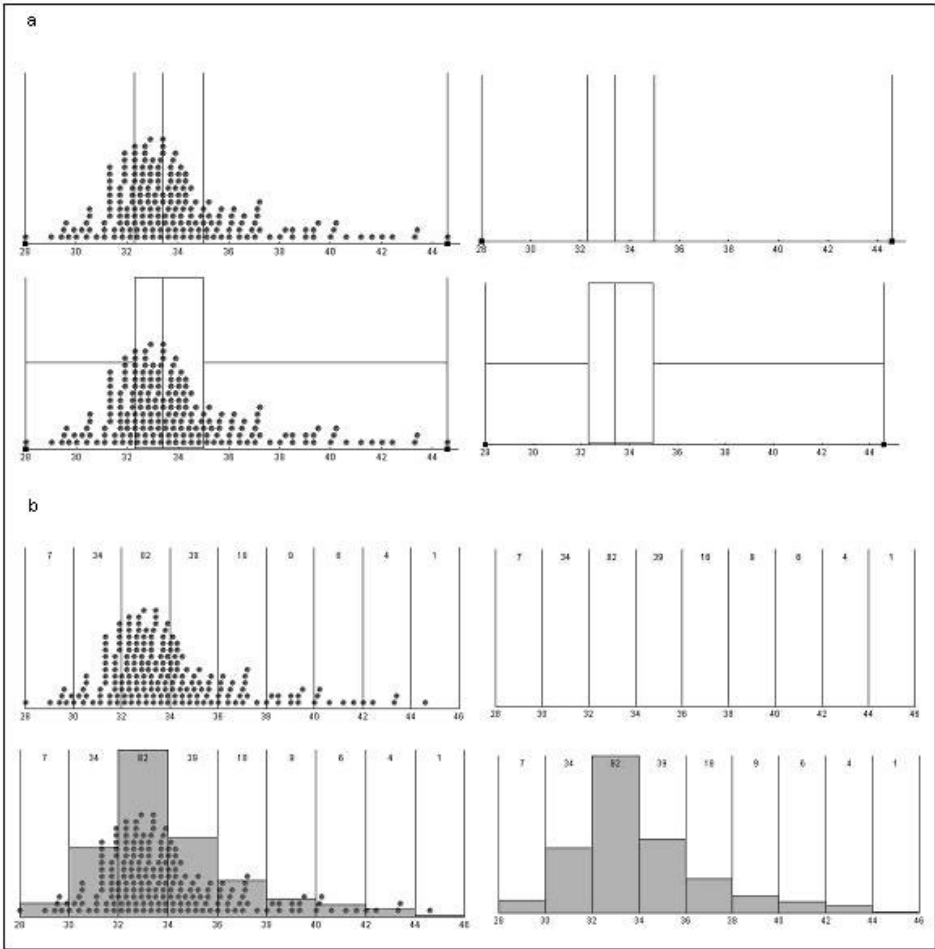


Figure 7. (a) Four equal group option with and without data. Box plot overlay was added after these seventh-grade teaching experiments. (b) Fixed interval width option with and without data. Histogram overlay was added after these seventh-grade teaching experiments.

An instructional idea that emerged during the last teaching experiment was that of “growing samples.” Discussing and predicting what would happen if we added more data appeared to lead to reasoning about several aspects of distribution in a coherent way. For the background to this activity, we have to go back to a problem from the beginning of the instructional unit:

In a certain hot air balloon basket, eight adults are allowed [in addition to the driver]. Assume you are going to take a ride with a group of seventh-graders. How many seventh-graders could safely go into that balloon basket if you only consider weight?

This question was meant to let students think about variation of weight, sampling, and representativeness of the average. A common solution in all classes was that students estimated an average or a typical weight for both adults and children. Some used the ratio of those numbers to estimate the number of children allowed, but most students calculated the total weight allowed and divided that by the average student weight. The student answers varied from 10 to 16.

This activity formed the basis for a class discussion on the reliability of the estimated weights, during which we asked for a method of finding more reliable numbers. A student suggested weighing two boys and two girls. The outcome of the discussion was that the students decided to collect weight data from the whole class. (In the second teaching experiment, they also collected height data.)

In the next lesson, we first showed the sample of four weight data in Minitool 2 (Figure 8a) and asked what students expected if we added the rest of the data. Students thought that the mean would be more precise. Because we did not want to focus on the mean, we asked about the shape and the range. Some students then conjectured that the range would be larger, and others thought the graph would grow higher. After showing the data for the whole class (Figure 8b), we asked what would happen if we added the data for two more classes (Figure 8c). In this way, extreme values, spread, and shape became topics of discussion. The graphs that students made to predict the shape if sample size were doubled tended to be smoother than the graphs students had seen in Minitool 2 (Figure 8d). In our interpretation, students started to see a pattern in the data—or in Konold and Pollatsek's words, a "signal in the noise" (Chapter 8). We concluded that stimulating reasoning about distribution by "growing samples" is another useful heuristic for instructional design in statistics education.

A conjecture about students' evolving notion of distribution that was confirmed in the retrospective analyses was that students tend to divide unimodal distributions into three groups of low, "average," and high values. We saw this conceptual grouping into three groups for the first time in the second teaching experiment when we asked what kind of graph students expected when they collected height data. Daniel did three trials (Figure 9). During his second trial, he said: "You have smaller ones, taller ones, and about average." After the third trial he commented: "There are more around the average." Especially in the third trial, we clearly see his conceptual organization into three groups, which is a step away from focusing on individual data points.

One step further is when students think of small, average, tall, and "in between." When in the final test students had to sketch their class when ordered according to height, Christa drew Figure 10 and wrote: "There are 3 smaller ones, about 10 average, 3 to 4 taller, and of course in between."

The "average" group, the majority in the middle, seems to be more meaningful to students than the single value of the mean. Konold and colleagues (2002) call these ranges in the middle of distributions *modal clumps*. Our research supports their view that these modal clumps may be suitable starting points for informal reasoning about center, spread, and skewness. When growing samples, students might even learn to see such aspects of distribution as stable features of variable processes.

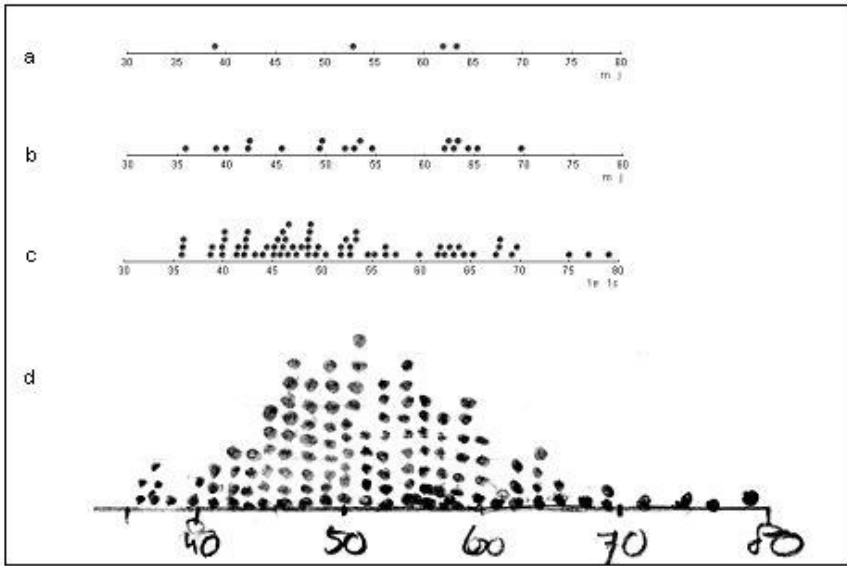


Figure 8. Growing samples (weight data in kg): (a) Four students; (b) one class; (c) three classes; (d) a student's smoother prediction graph of larger sample.

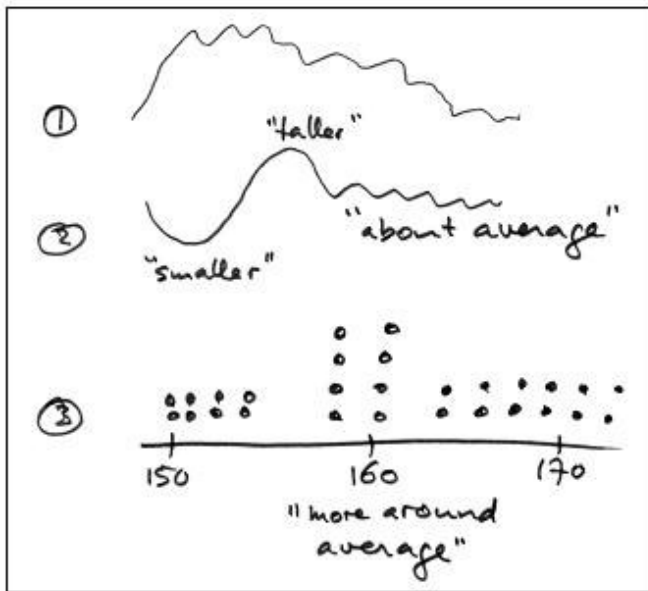


Figure 9. Three prediction trials of height data; the second and third show three groups.

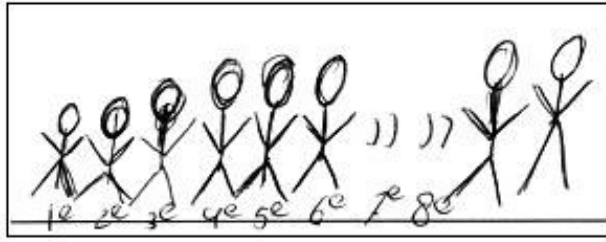


Figure 10. Class ordered by height. Christa’s explanation: “There are 3 smaller ones, about 10 average, 3 to 4 taller, and of course in between.”

STAGE 3—SYMBOLIZING DATA AS A “BUMP”

Though students in the first two teaching experiments started to reason with majorities and modal clumps in the second stage, they did not explicitly reason with shape. We had hoped that they would reason with “hills,” as was the case in the teaching experiment of Cobb, Gravemeijer, and McClain (Cobb, 1999), but they did not. A possible reason is that their teaching experiment lasted 34 lessons, whereas ours lasted only 12 or 15 lessons. In the second teaching experiment, we decided to try something else. In line with the reasons to let students invent their own data (Section 5), we asked students to invent their own graphs of their own data. As a follow-up of the balloon activity mentioned earlier, the students had to make a graph for the balloon rider, which she could use in deciding how many students she could safely take on board.

The students of the second teaching experiment drew various graphs. The teacher focused the discussion on two graphs, namely, Michiel’s and Elleke’s (Figure 11).

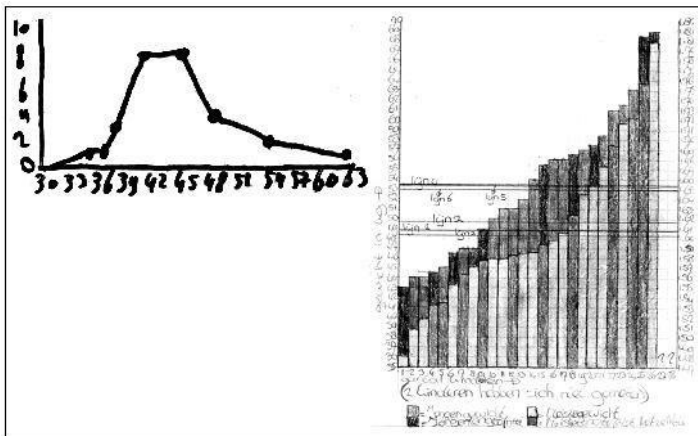


Figure 11. Michiel’s graph (left) and Elleke’s graph.

The shorter bars represent students' weights; the lightest bars signify girls' weights. Though all students used the same data set, Michiel's graph on a transparency does not exactly match the values in Elleke's graph on paper. Michiel's graph is more like a rough sketch.

Michiel's graph is especially interesting, since it offered the opportunity to talk about shape. Michiel explained how he got the dots as follows. (Please note that a translation of ungrammatical spoken Dutch into written English does not sound very authentic.)

Michiel: Look, you have roughly, averagely speaking, how many students had that weight and there I have put a dot. And then I have left [y-axis] the number of students. There is one student who weighs about 35 [kg], and there is one who weighs 36, and two who weigh 38 roughly.

And so on: the dot at 48, for example, signifies about four students with weights around 48. After some other graphs had been discussed, including that of Elleke, the teacher asked the following question.

Teacher: What can you easily see in this graph [by Michiel]?

Laila: Well, that the average, that most students in the class, uhm, well, are between 39 and, well, 48.

Teacher: Yes, here you can see at once which weight most students in this class roughly have, what is about the biggest group. Just because you see this bump here. We lost the bump in Elleke's graph.

It was the teacher who used the term bump for the first time. Although she had tried to talk about shapes earlier, this was the first time the students picked it up. As Laila's answer indicates, Michiel's graph helped her to see the majority of the data—between 39 and 48 kg. This “average” or group of “most students” is an instance of what Konold and colleagues (2002) call a modal clump. Teachers and curriculum designers can use students' informal reasoning with clumps as preparation for using the average as a representative value for the whole group, for example.

Here, the teacher used the term bump to draw students' attention to the shape of the data. By saying that “we lost the bump in Elleke's graph,” she invited the students to think about an explanation for this observation. Nadia reacted as follows.

Nadia: The difference between ... they stand from small to tall, so the bump, that is where the things, where the bars [from Elleke's graph] are closest to one another.

Teacher: What do you mean, where the bars are closest?

Nadia: The difference, the endpoints [of the bars], do not differ so much with the next one.

Eva added to Nadia's remarks:

- Eva:* If you look well, then you see that almost in the middle, there it is straight almost and uh, yeah that [teacher points at the horizontal part in Elleke's graph].
- Teacher:* And that is what you [Nadia] also said, uh, they are close together and here they are bunched up, as far as [...] weight is concerned.
- Eva:* And that is also that bump.

These episodes demonstrate that, for the students, the bump was not merely a visual characteristic of a certain graph. It signified a relatively large number of data points with about the same value—both in a hill-type graph and in a value-bar graph. For the students, the term bump signified a range where there was a relatively high density of data points. The bump even became a tool for reasoning, as the next episode shows, when students revisited the battery task as one of the final tasks.

- Laila:* But then you see the bump here, let's say [Figure 3].
- Ilona:* This is the bump [pointing at the straight vertical part of the lower 10 bars].
- Researcher:* Where is that bump? Is it where you put that red line [the vertical value bar]?
- Laila:* Yes, we used that value bar for it [...] to indicate it, indicate the bump. If you look at green [the upper ten], then you see that it lies further, the bump. So we think that green is better, because the bump is further.

The examples show that some students started to reason about density and shape in the way intended. However, they still focused on the majority, the modal clump, instead of the whole distribution. This seemed to change in the 13th lesson of the second teaching experiment

In that lesson, we discovered that asking students to predict and reason without available data was helpful in fostering a more global view of data. A first example of such a prediction question is what a graph of the weights of eighth-graders would look like, as opposed to one of seventh-graders. We hoped that students would shift the whole shape instead of just the individual dots or the majority.

- Teacher:* What would a graph of the weights of eighth-graders look like?
- Luuk:* I think about the same, but another size, other numbers.
- Guyonne:* The bump would be more to the right.
- Teacher:* What would it mean for the box plots?
- Michiel:* Also moves to the right. That bump in the middle is in fact just the box plot, which moves more to the right.

It could well be that Luuk reasoned with individual numbers, but he thought that the global shape would look the same. Instead of talking about individual data points, Guyonne talked about a bump, in singular, shifted to the right. Michiel related to the box plot as well, though he just referred to the box of the box plot.

Another prediction question also led to reasoning about the whole shape, this time in relation to other statistical notions such as outliers and sample size. Note that

students used the term *outliers* for extreme values, not for values that are questionable.

Researcher: If you would measure all seventh-graders in the city instead of just your class, how would the graph change, or wouldn't it change?

Elleke: Then there would come a little more to the left and a little more to the right. Then the bump would become a little wider, I think. [She explained this using the term outliers.]

Researcher: Is there anybody who does not agree?

Michiel: Yes, if there are more children, than the average, so the most, that also becomes more. So the bump stays just the same.

Albertine: I think that the number of children becomes more and that the bump stays the same.

In this episode, Elleke relates shape to outliers; she thinks that the bump grows wider if the sample grows. Michiel argues that the group in the middle also grows higher, which for him implies that the bump keeps the same shape. Albertine's answer is interesting in that she seems to think of relative frequency: for her the shape of the distribution seems to be independent of the sample size. If she thought of absolute frequency she would have thought that the bump would be much higher. Apparently, the notion of a bump helped these students to reason about the shape of the distribution in hypothetical situations. In this way, they overcame the problem of seeing only individual data points and developed the notion of a bump, which served as a conceptual unity.

There are several reasons why predictions about shape in such hypothetical situations can help to foster understanding of shape or distribution. First, if students predict a graph without having data, they have to reason more *globally* with a property in their mind. Konold and Higgins (2002) write that with the individuals as the foci, it's difficult to see the forest for the trees. Our conclusion is that we should ask questions about the forest, or predict properties of other forests—which we consider another heuristic for statistics education. This heuristic relates to the cognitive limitations mentioned in Section 5: If there are no available data and students have to predict something on the basis of some conceptual characteristic, it is impossible to imagine many individual data points.

A second reason has to do with the *smoothness* of graphs. Cobb, McClain, and Gravemeijer (2003) assume that students can more easily reason about hills if the hills are smooth enough. We found evidence that the graphs students predict tend to be smoother than the graphs of real data, and we conjecture that reasoning with such smoother graphs helps students to see the shape of a distribution through the variation or, in other words, the signal through the noise (Konold & Pollatsek, Chapter 8). If they do so, they can model data with a notion of distribution, which is the downward perspective we aimed for (Section 3).

A last example illustrates how several students came to reason about distributions. These two girls were not disturbed by the fact that distributions did not look like hills in Minitool 1. The question they dealt with was whether the distributions of the battery brands looked normal or skewed, where *normal* was informally defined as "symmetrical, with the median in the middle and the majority

close to the median.” The interesting point is that they used the term *hill* to indicate the majority (see Figure 3), although it looked straight in the case-value bar graph. This indicates that the hill was not a visual tool; it had become a conceptual tool in reasoning about distributions.

Albertine: Oh, that one [battery brand D in Figure 3] is normal [...].

Nadia: That hill.

Albertine: And skewed if like here [battery brand K] the hill [the straight part] is here.

DISCUSSION

The central question of this chapter was how seventh-grade students could learn to reason about distributions in informal ways. In three stages, we showed how certain instructional activities, supported by computer tool use and the invention of graphs, stimulated students to reason about aspects of distributions. After a summary of the results we discuss limitations of this study and implications for future research.

When solving statistical problems with Minitool 1, students used informal words such as *majority*, *outliers*, *reliability*, and *spread out*. The examples show that students reasoned about aspects of distribution from the very start of the experiment. The students invented data sets in Minitool 1 that matched certain characteristics of battery brands such as “good but not reliable.” We argued that letting students invent their own data sets could stimulate them to think of a data set as a whole instead of individual data points (heuristic 1). The bar representation of Minitool 1 stimulated a visual compensation strategy of finding the mean, whereas many students found it easier to see the spread of the data in Minitool 2.

When working with Minitool 2, students developed qualitative notions of more advanced aspects of distribution such as frequency, classes, spread, quartiles, median, and density. The dot plot representation in combination with the options to structure data into two equal groups, four equal groups, fixed group size, and fixed interval width supported the development of an understanding of the median, box plot, density, and histogram respectively. Like Konold and colleagues (2002), we expect that modal clumps are useful to help students reason with center and other distribution aspects. Growing samples is a promising instructional activity to let students reason with stable features of variable processes (heuristic 2). The big ideas of sampling and distribution can thus be developed coherently, but how this could be done is a topic of future research.

In the third stage, students started to reason with bumps in relation to statistical notions such as majority, outliers, and sample size in hypothetical situations and in relation to different graphs. We argued that predictions about the shape and location of distributions in hypothetical situations are useful to foster a more global view and to let students see the signal in the noise (heuristic 3).

IMPLICATIONS

The results of this research study suggest that it is important to provide opportunities for students to contribute their own ideas to the learning process, which requires much discussion and interaction during class. We believe that formal measures such as median and quartiles should be postponed until intuitive notions about distribution have first been developed. We also encourage teachers to allow students to use less than precise statistical definitions as students develop their reasoning, and then make a transition to more specific definitions as students are able to comprehend these details. We are convinced that teachers should try to learn about how students are reasoning about distribution by listening and observing as well as by gathering assessment data. A type of assessment that we found useful asked students to create a graph representing statistical information. One such task that was very effective asked students to make graphs that were compatible with a short story with both informal and statistical notions related to running practice. There were no restrictions on the type of graph students could use. We had deliberately incorporated characteristics in the story that ranged from easy (the fastest runner needed 28 minutes) to difficult (the spread of the running times at the end was much smaller than in the beginning but the range was still pretty big). This is the item we used:

A seventh grade is going to train for running 5 km. To track their improvement they want to make three graphs. One before training starts, one halfway through, and one after ten training sessions. Draw the graphs that belong to the following story:

- Before training started some students were slow and some were already very fast. The fastest ran the 5 km in 28 minutes. The spread between the other students was large. Most of them were on the slow side.
- Halfway through, the majority of the students ran faster, but the fastest had improved his time only a little bit, as had the slowest.
- After the training sessions had finished, the spread of the running times was much smaller than in the beginning, but the range was still pretty big. The majority of the students had improved their times by about 5 minutes. There were still a few slow ones, but most of the students had a time that was closer to the fastest runner than in the beginning.

We found that students were able to represent many elements in their graphs and we learned more about their thinking and reasoning by examining their constructions.

Although we conclude that it is at least possible for seventh-graders to develop the kind of reasoning about distribution that is shown in this chapter, it should be stressed that the students in these experiments had above-average learning abilities and had been stimulated to reflect during mini-interviews. Other students probably need more time or need to be older before they can reason about distribution in a similar way.

Another limitation of this study is that the examples of the third stage were to a certain extent unique for the second teaching experiment. What would have happened if Michiel had not made his “bump” graph? This research does not completely answer that question (there was some reasoning with bumps in the third and fourth teaching experiment), but it shows what the important issues are and which heuristics might be useful for instructional activities.

In addition, we noticed that making predictions graphs without having data is not a statistical practice that automatically emerges from doing an instructional sequence such as the one described here. We concluded this from observations during the two subsequent school years, when two novice teachers used the materials in 12 other seventh-grade classes. When we asked prediction questions, the students seemed confused because they were not used to such questions. An implication for teaching is that establishing certain socio-mathematical norms and certain practices (Cobb & McClain, Chapter 16) are as important as suitable computer tools, carefully planned instructional activities, and skills of the teacher to orchestrate class discussions.

These teachers also reported that some of the statistical problems we had used or designed were too difficult and not close enough to the students’ world of experience. The teachers also needed much more time than we used in the first year, and they found it difficult to orchestrate the class discussions. We acknowledge that the activities continually need to be adjusted to local contingencies, that the mini-interviews probably had a learning effect, and that the teachers needed more guidance for teaching such a new topic. Hence, another question for future research is what kind of guidance and skills teachers need to teach these topics successfully.

NOTE

We thank the teachers Mieke Abels, Maarten Jasper, and Mirjam Jansens for joining this project. The research was funded by the Netherlands Organization for Scientific Research under number 575-36-003B. The opinions expressed in this chapter do not necessarily reflect the views of the Organization.

REFERENCES

- Bakker, A. (2003). The early history of average values and implications for education. *Journal of Statistics Education*, 11(1). Online: <http://www.amstat.org/publications/jse/v11n1/bakker.html>
- Biehler, R (2001). Developing and assessing students’ reasoning in comparing statistical distributions in computer-supported statistics courses. In C. Reading (Ed.), *Proceedings of the Second International Research Forum on Statistical Reasoning, Thinking, and Literacy (SRTL-2)*. Armidale, Australia: University of New England.
- Ben-Zvi, D., & Arcavi, A. (2001). Junior high school students’ construction of global views of data and data representations. *Educational Studies in Mathematics*, 45(1), 35–65.
- 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. P. E. (2003). Learning about statistical covariation. *Cognition and Instruction* 21(1), 1–78.
- Cobb, P., & Whitenack, J. W. (1996). A method for conducting longitudinal analyses of classroom videorecordings and transcripts. *Educational Studies in Mathematics*, 30(3), 213–228.
- de Lange, J., Burrill, G., Romberg, T., & van Reeuwijk, M. (1993). *Learning and testing mathematics in context: The case—data visualization*. Madison, WI: University of Wisconsin, National Center for Research in Mathematical Sciences Education.
- Edelson, D. C. (2002). Design research: What we learn when we engage in design. *Journal of the Learning Sciences*, 11(1), 105–121.
- Freudenthal, H. (1991). *Revisiting mathematics education: China lectures*. Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Friel, S. N., Curcio, F. R., & Bright, G. W. (2001). Making sense of graphs: Critical factors influencing comprehension and instructional implications. *Journal for Research in Mathematics Education*, 32(2), 124–158.
- Gravemeijer, K. P. E. (1994). *Developing realistic mathematics education*. Utrecht, The Netherlands: CD Bèta Press.
- Hancock, C., Kaput, J. J., & Goldsmith, L. T. (1992). Authentic enquiry with data: Critical barriers to classroom implementation. *Educational Psychologist*, 27(3), 337–364.
- Konold, C., & Higgins, T. (2002). Highlights of related research. In S. J. Russell & D. Schifter & V. Bastable (Eds.), *Developing mathematical ideas: Working with data* (pp. 165–201). Parsippany, NJ: Seymour.
- Konold, C., Robinson, A., Khalil, K., Pollatsek, A., Well, A. D., & Wing, R., et al. (2002). Students' use of modal clumps to summarize data. In B. Phillips (Ed.), *Developing a Statistically Literate Society: Proceedings of the International Conference on Teaching Statistics [CD-ROM]*, Cape Town, South Africa, July 7–12, 2002.
- Lemke, J. L. (2003). Mathematics in the middle: Measure, picture, gesture, sign, and word. In M. Anderson, A. Sáenz-Ludlow, S. Zellweger, & V. V. Cifarelli (Eds.), *Educational perspectives on mathematics as semiosis: From thinking to interpreting to knowing* (pp. 215–234). Ottawa, Ontario: Legas Publishing.
- Meira, L. (1995). Microevolution of mathematical representations in children's activity. *Cognition and Instruction*, 13(2), 269–313.
- Mokros, J., & Russell, S. J. (1995). Children's concepts of average and representativeness. *Journal for Research in Mathematics Education*, 26(1), 20–39.
- National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: NCTM.
- Petrosino, A. J., 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.
- Sfard, A. (2000). Steering (dis)course between metaphors and rigor: Using focal analysis to investigate an emergence of mathematical objects. *Journal for Research in Mathematics Education*, 31(3), 296–327.
- Steinbring, H. (1997). Epistemological investigation of classroom interaction in elementary mathematics teaching. *Educational Studies in Mathematics*, 32(1), 49–92.
- Strauss, A. & Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Thousand Oaks, CA: Sage.
- Wilkinson, L. (1999). Dot plots. *American Statistician*, 53(3), 276–281.
- Zawojewski, J. S., & Shaughnessy, J. M. (2000). Mean and median: Are they really so easy? *Mathematics Teaching in the Middle School*, 5(7), 436–440.