

REASONING ABOUT SHAPE AS A PATTERN IN VARIABILITY

ARTHUR BAKKER

Freudenthal Institute

Utrecht University

The Netherlands

a.bakker@ioe.ac.uk

SUMMARY

This paper examines ways in which coherent reasoning about key concepts such as variability, sampling, data, and distribution can be developed as part of statistics education. Instructional activities that could support such reasoning were developed through design research conducted with students in grades 7 and 8. Results are reported from a teaching experiment with grade 8 students that employed two instructional activities in order to learn more about their conceptual development. A “growing a sample” activity had students think about what happens to the graph when bigger samples are taken, followed by an activity requiring reasoning about shape of data. The results suggest that the instructional activities enable conceptual growth. Last, implications for teaching, assessment and research are discussed.

Keywords: Design research; Distribution; Instructional activities; Middle school level; Sampling

1. BACKGROUND OF THE RESEARCH

The first time I visited an American classroom I attended a statistics lesson in grade 5. When the teacher asked a question that sounded statistical but did not require a measure of center, one student, Malcolm, thoughtlessly muttered “meanmedianmode,” as if it were one word. My impression was that these students had been drilled to calculate mean, median, and mode, and to draw bar graphs, but did not use their common sense in answering statistical questions. This small incident exemplifies what a litany of research in statistics education reports on: too often, students learn statistics as a set of techniques that they do not apply sensibly. Even if they have learned to calculate mean, median, mode, and to draw histograms and box plots, they mostly do not understand that they can use a mean as a group descriptor when comparing two data sets—to give one example that is well documented (Konold & Higgins, 2003; McGatha, Cobb, & McClain, 2002; Mokros & Russell, 1995). This problem is not typically American; it also applies to the Dutch context, but to a lesser extent. The reason for this is probably that Dutch students mostly learn statistical concepts and graphs such as median, mode, histogram, and box plot about three years later than in the USA.

Despite differences between the curricula in different countries, the underlying problem remains the same: students generally lack the necessary conceptual understanding for analyzing data with the statistical techniques they have learned. The problem many statistics educators encounter is that students tend to perceive data just as a series of individual cases (a case-oriented view), and not as a whole that has characteristics that are not visible in any of the individual cases (an aggregate view). Hancock, Kaput, and Goldsmith (1992) note that students need to mentally construct such an aggregate before they can perceive a data set as a whole. Many researchers have encountered the same problem and experienced its persistency (e.g. Ben-Zvi & Arcavi, 2001; Wilensky, 1997).

The above implies that students need to develop a conceptual structure with which they can conceive data sets as aggregates. Konold and Pollatsek (2002) argue that students need to develop a conceptual understanding of signal and noise in order to understand what an average value is about in relation to the variation around that value. The present paper distinguishes two types of signals in noisy processes or patterns in variability. First, the signal can be a true value with error as noise around it. Such signals are apparent in repeated measurements of one item (Petrosino, Lehrer, & Schauble, 2003). The “center clump” is then an indication of where the true value probably is. Second, the signal might be a distribution, such as the shape of a smooth bell curve of the normal distribution, with which we model data. The noise in that case is the variation around that smooth curve. In either type of pattern, it is evident that students need good conceptual understanding before they can recognize signals in noisy processes. This paper focuses on the second type of signal in noisy processes or, other words, *shape as a pattern in variability*.

The concept of distribution is a structure with which students can conceive aggregate features of data sets (Cobb, 1999; Gravemeijer, 1999). Petrosino et al. (2003, p. 132) write: “Distribution could afford an organizing conceptual structure for thinking about variability located within a more general context of data modeling.” Of course, distribution is a very advanced concept that, in its full complexity, is far beyond the scope of middle school students (Wilensky, 1997). Nevertheless, it is possible to address the issue of how data are distributed from an informal situational level onwards by focusing on shape (Bakker & Gravemeijer, 2004; Cobb, 1999; Russell & Corwin, 1989). In the research of Cobb, McClain, and Gravemeijer (2003), students came to reason with “hills” to indicate “majorities” of data sets, which are informal terms to describe the hill shape of data sets and areas in graphs where most data points seemed to be. Bakker and Gravemeijer (2004) report that students reasoned how a “bump” would shift if older students were measured, and what would happen with the bump if the sample “grew”, i.e., began to include more and more cases. The seventh-grade students in their study came to see a pattern in the variability of different phenomena such as weight, height, and wingspan of birds.

In the first teaching experiment I conducted in grade 7 (Bakker, 2004), I focused on the concept of distribution, but this turned out to be too limited. Sampling, for instance, is also crucial to address in an early stage of statistical data analysis (see also Bakker & Gravemeijer, 2004). The research presented here focused on a broader set of key concepts that students, in my view, need to develop in order to analyze data in a meaningful way: variability, sampling, data, and distribution (cf. Garfield & Ben-Zvi, 2004). The main question of the overall research was, how can we promote coherent reasoning about variability, sampling, data, and distribution in a way that is meaningful for students with little statistical background?

The learning process aimed at in this research can be characterized as “guided reinvention” (Freudenthal, 1991). Students were stimulated to contribute their own ideas, strategies, and language in solving statistical problems (reinvention), but they were also provided with increasingly sophisticated ways to describe how data were distributed and to characterize data sets (planned guidance).

In this paper I report results from a teaching experiment that employed two instructional activities I developed. The paper analyzes students’ learning process in order to learn more about their development of key concepts underlying statistical data analysis, especially variability, sampling, data, and distribution. The two activities used seemed particularly promising for fostering coherent reasoning about these key concepts and were developed using a cyclic approach of designing instructional materials, testing them during classroom-based teaching experiments, analyzing students’ learning process, and revising the instructional materials. The first instructional activity, growing a sample or a data set, is an elaboration of an activity described in Bakker and Gravemeijer (2004). The second activity involved reasoning about shapes that students themselves had proposed. In grade 7, it turned out to be difficult for students to reason with shape, except for high achievers who reasoned with bumps. The teaching experiment reported here was therefore carried out in a higher grade (8th).

Below, I first elaborate on the methodological approach of design research employed in this eighth-grade teaching experiment, and then describe the subjects, data collection, and method of

analysis. Results are then presented regarding students' reasoning during two instructional activities, "growing a sample" and "reasoning about shapes". Finally, the results and their limitations are discussed, as well as implications for teaching, assessment, and research.

2. METHODOLOGY AND SUBJECTS

2.1. DESIGN RESEARCH

To answer the question of how coherent reasoning about variability, sampling, data, and distribution could be promoted, I needed to design instructional activities that could support such reasoning as well as to understand *how* those activities supported students' conceptual development. If the kind of education aimed at is not yet available, the required conditions first need to be created. Instructional design is therefore an important part of the research presented here. In general, if you want to change something you have to understand it and if you want to understand something you have to change it. In this approach, design and "research" are highly intertwined, and it will not surprise the reader that this type of research falls under the general heading of "design research." Design research typically involves cycles of three phases: a preparation and a design phase (of instructional materials for example), teaching experiments, and retrospective analyses (Cobb, Confrey, diSessa, Lehrer, & Schauble, 2003; Gravemeijer, 1994).

1. *Preparation and design phase.* In the research presented here, the preparation phase consisted of a literature survey, a historical study of the statistical concepts and graphs at issue (Bakker, 2003), and the first reformulation of a Hypothetical Learning Trajectory (HLT), which Simon (1995) defines as follows: "A hypothetical learning trajectory is made up of three components: the learning goal that defines the direction, the learning activities, and the hypothetical learning process – a prediction of how the students' thinking and understanding will evolve in the context of the learning activities" (p. 136). The hypothetical learning trajectory of the present study was to support students in reasoning about aspects of distribution and sampling using increasingly sophisticated concepts and graphs. Further details about hypothetical learning trajectories can be found in a special issue of *Mathematical Thinking and Learning* devoted to this topic: 6(2).

2. *Teaching experiment.* The HLT is tested and possibly revised during a teaching experiment. The anticipations formulated in the HLT give guidance to both teacher and researcher of what to focus on during instruction, interviewing, and observation. The teacher and researcher can adjust their original plans if new ideas seem to be better, so they need not wait till the end of the teaching experiment to change activities or even the end goal.

3. *Retrospective analysis.* The retrospective analysis is meant to find out if the anticipations of the HLT were right, to find patterns in students' learning processes and to understand the role of the instructional materials (activities, software). New insights mostly lead to the revision of the instructional materials, the end goals, or the route to be taken next time and a revised HLT that can guide the next teaching experiment.

Overall, the idea behind developing an HLT is not to design the perfect instructional sequence, which in my view does not exist, but to provide empirically grounded results that others can adjust to their local circumstances. The HLT remains hypothetical because each situation, each teacher, and each class is different. Yet patterns can be found in students' learning that are similar across different teaching experiments (Bakker, 2004). Those patterns and the insights of how particular instructional activities support students in particular kinds of reasoning can be the basis for a more general instructional theory of how a particular domain can be taught.

2.2. SETUP, SUBJECTS, DATA COLLECTION, AND ANALYSIS

Setup. This paper focuses on the fourth and the sixth lessons of a series of ten lessons, each 50 minutes long. In these specific lessons, on which the retrospective analyses also centered, students reasoned about larger and larger samples and about the shape of distributions. Half of the lessons

were carried out in a computer lab and as part of them students used two software minitools (Cobb, Gravemeijer, Bowers, & Doorman, 1997), simple Java applets with which they analyzed data sets on, for instance, battery life span, car colours, and salaries (Figure 1). As a researcher I was responsible for the instructional materials and the teacher was responsible for the teaching, though we discussed in advance on a weekly basis both the materials and appropriate teaching style. Three pre-service teachers served as assistants and helped with videotaping and interviewing students and with analyzing the data.

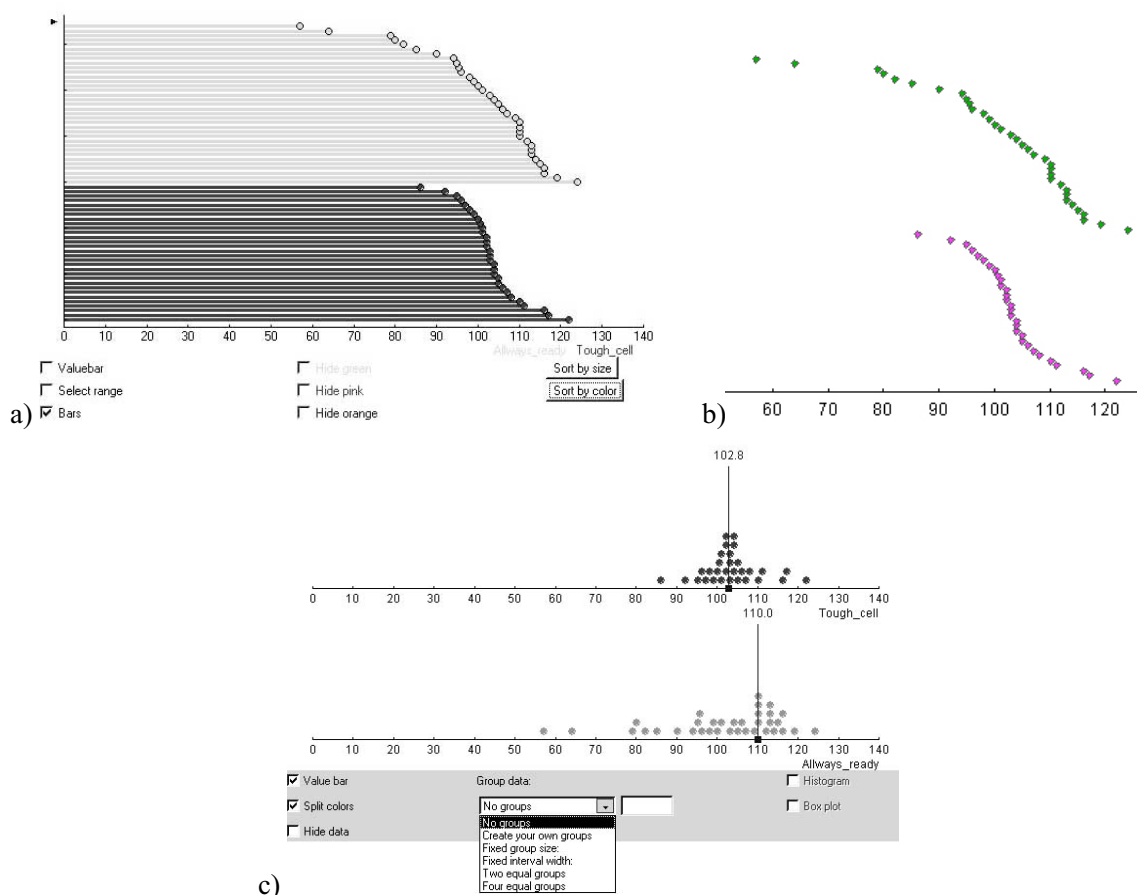


Figure 1. a) Minitool 1 showing a value-bar graph of battery life spans in hours of two brands. b) Minitool 1, but with bars hidden. c) Minitool 2 showing a dot plot of the same data sets.

Subjects. The teaching experiment was carried out in an eighth-grade class with 30 students in a public school in the center of the Dutch city of Utrecht in the fall of 2001. The students in this study were being prepared for pre-university (*vwo*) or higher vocational education (*havo*). The top 35-40% of Dutch students attend these types of education. The remaining 60-65% of students are prepared for other types of vocational education (*vmbo*). Other relevant background information is that school textbooks play a central role in the practice of Dutch mathematics education. Students are expected to be able to work through the tasks by themselves, with the teacher available to help them if necessary. As a consequence, tasks are broken down into very small steps and real problem solving is rare. Students' answers tend to be superficial, in part because they have to deal with about eight different problem contexts per lesson (Van den Boer, 2003). The students in the class reported on here were not used to whole-class discussions, but rather to be "taken by the hand" as the teacher called it; they were characterized by the three assistants as "passive but willing to cooperate." These eighth-grade students had no prior instruction in statistics; they were acquainted with bar and line graphs, but not with dot plots, histograms, or box plots. Students already knew the mean from calculating their report grades, but mode and median were not introduced until the second half of the instructional sequence after variability, data, sampling, and shape had been topics of discussion.

Data collection. The collected data on which the results presented in this paper are based (regarding the fourth and sixth lessons) include student work, field notes, and the audio and video recordings of class activities that the three assistants and I made in the classroom. An essential part of the data corpus was the set of mini-interviews we held during the lessons; they varied from about twenty seconds to four minutes, and were meant to find out what concepts and graphs meant for students, or how the mini-tools were used. These mini-interviews influenced students' learning because they often stimulated reflection. However, I think that the validity of the research was not put in danger by this, since the aim was to find out how students learned to reason with shape or distribution, not whether teaching the sequence in other eighth-grade classes would lead to the same results in the same number of lessons. Furthermore, the interview questions were planned in advance, and discussed with the assistants.

Analysis. For the analysis, I read the transcripts, watched the videotapes, and formulated conjectures on students' learning on the basis of episodes identified in the transcripts and video. The generated conjectures were tested against 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, a method resembling Glaser and Strauss's constant comparative method (Glaser & Strauss, 1967). About one quarter of the episodes, including those discussed in this paper, and the conjectures belonging to these episodes were judged by the three assistants who attended the teaching experiment. The amount of agreement among judges was very high, over 95%. Only the conjectures that all of us agreed upon were kept. An example of a conjecture that was confirmed was that students tended to group data sets, real or imagined, into three groups of low, "average", and high values.

For analyzing students' reasoning with diagrams I used the semiotics of Peirce (1976), in particular his concepts of diagrammatic reasoning and hypostatic abstraction. *Diagrammatic reasoning* involves three steps: constructing a diagram, experimenting with it, and reflecting upon the results. An important part of the reflection step is to describe what is seen in diagrams (bars, dots, relationships, shapes). The process of describing qualities of those objects can be called *predication*. Van Oers (2000) uses the following definition: "Predication is the process of attaching extra quality to an object of common attention (such as a situation, topic or theme) and, by doing so, making it distinct from others" (p. 150). Next, *hypostatic abstraction* is one of the forms of abstraction that Peirce distinguished: a predicate becomes an object in itself that can have characteristics. This is linguistically reflected in the transition from a predicate (e.g. most, lying out) to a noun (majority, outlier). Note that this paper focuses only on predication or hypostatic abstraction in instances of diagrammatic reasoning.

3. RESULTS

This section presents the analysis of students' reasoning during the two instructional activities on which this paper is focused. The first, carried out in the fourth lesson, is reasoning about larger and larger samples, or larger and larger data sets. I call this activity "growing a sample" (following Konold & Pollatsek, 2002), though some readers may prefer to call it "growing a data set." The term "sample" is preferred here because one intention of the activity was to let students think about samples versus populations. The second activity, carried out in the sixth lesson, is reasoning about shapes of these data sets. For each activity, I first summarize the hypothetical learning trajectory (HLT) of that lesson, and then present the analysis.

3.1. GROWING A SAMPLE

The overall goal of the growing samples activity as formulated in the hypothetical learning trajectory for this fourth lesson was to let students reason about shape in relation to sampling and distribution aspects in the context of weight. The idea was to start with students' own ideas and guide them toward more conventional notions and representations.

Before getting to the growing samples activity, in a previous (third) lesson students had answered the question of how many eighth graders could go into a hot air balloon if normally eight adults (apart from the balloon pilot) were allowed. They also had made a prediction of a weight graph of eighth graders, without having any data available yet. It was apparent that students tended to choose small group sizes such as 10 or 15. The students had also had two lessons of experience with two the minitools. The first minitool supplies a value-bar graph in which each bar has a length corresponding to the data value it represents (Figure 1a); the second minitool provides a dot plot (Figure 1c). In the first minitool, students can organize data, for instance by sorting or hiding subsets of data, and by sorting the data by size. They can also hide the bars, so that they only see the endpoints of the bars (Figure 1b). In the second minitool, these endpoints have been collapsed onto the axis. Students can organize data with different options, for instance making their own groups, two equal groups (precursor to the median), four equal groups (precursor to box plot), and fixed interval width (precursor to histogram).

The activity of growing a sample built on the balloon activity. It consisted of three cycles of making sketches of a hypothetical situation and comparing those sketches with graphs displaying real data sets. In the first cycle students had to make a graph of their own choice of a predicted weight data set with sample size 10. The results were discussed by the teacher to challenge this small sample size, and in the subsequent cycles students had to predict larger data sets (one class, three classes, all students in the province). Three such cycles took place as described below. The teacher and I tried to strike a balance between engaging students in statistical reasoning and allowing their own terminology on the one hand, and guiding them in using conventional and more precise notions and graphical representations on the other.

First cycle of growing a sample

The text of the student activity sheet for the fourth lesson started as follows:

Last week you made graphs of predicted data for a balloon pilot. During this lesson you will get to see real weight data of students from another school. We are going to investigate the influence of the sample size on the shape of the graph.

a. Predict a graph of ten data values, for example with the dots of minitool 2.

The sample size of ten was chosen because the students had found that size reasonable after the first lesson in the context of testing the life span of batteries. Figure 2a shows examples for three different types of diagrams the students made to show their predictions: there were three value-bar graphs (such as in minitool 1, e.g., Ruud's diagram), eight with only the endpoints (such as with the option of minitool 1 to "hide bars", e.g., Chris's diagram) and the remaining nineteen plots were dot plots (such as in minitool 2, e.g., Sandra's diagram). This means that their graphs were heavily influenced by their experiences with the minitools.

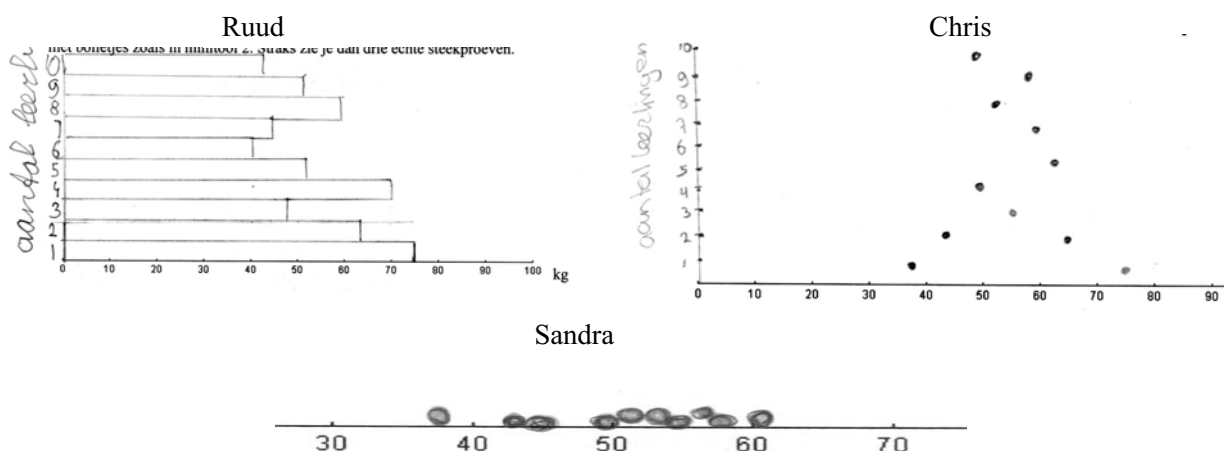


Figure 2a. Student predictions (Ruud, Chris, and Sandra) for ten data points (weight in kg)

For the remainder of this section, the figures and written explanations of these three students are demonstrated, because their work gives an impression of the variety of the whole class. The learning abilities of these students varied considerably: Ruud and Chris's report grades were in the bottom third of the class whereas Sandra had the best overall report score of the class across all subjects. I have chosen those three students because their diagrams represent all types of diagrams made in this class, also for other cycles of growing a sample.

To stimulate the reflection on the graphs, the teacher showed three samples of ten data points on the blackboard and students had to compare their own graphs (Figure 2a) with the graphs of the real data sets (Figure 2b).

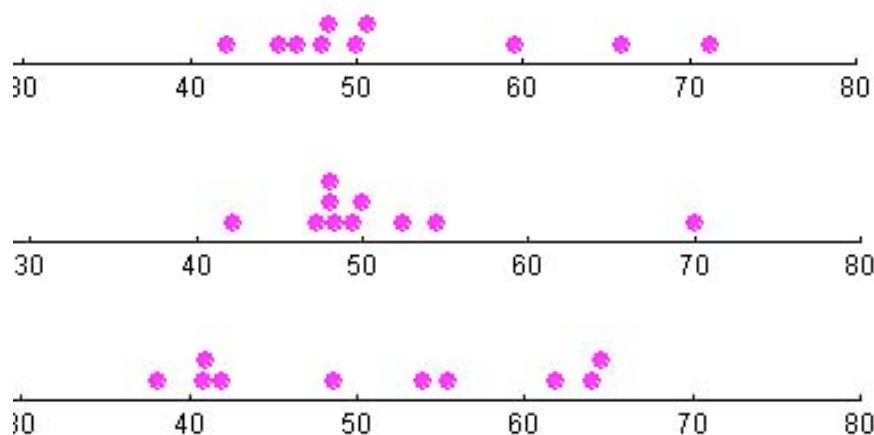


Figure 2b. Three real data sets in minitool 2

b. You get to see three different samples of size 10. Are they different than your own prediction? Describe the differences.

The reason for showing three small samples was to show the variation among these samples. There were no clear indications, though, that students conceived this variation as a sign that the sample size was too small for drawing conclusions, but they generally agreed that larger samples were more reliable. There was a short class discussion about the graphs with real data before students worked for themselves again. The point relevant to the analysis is that students started using predicates to describe aggregate features of the graphs. Please note that a grammatical translation into English of ungrammatical spoken Dutch does not always sound very authentic.

Teacher:	We're going to look at these three different ones [samples in Figure 2b]. Can anyone say something yet? Give it a try.
Jacob:	In the middle [graph], there are more together.
Teacher:	Here [pointing to the middle graph of Figure 2b] there are many more together, clumped or something like that. Who can mention other differences?
Jacob:	Well, uh, the lowest, I think it's all the furthest apart.
Teacher:	Those are all the furthest apart. Here [in the middle graph] they are in one clump. Are there any other things you notice, Gigi?
Gigi:	Yes, the middle one has just one at 70. [This is a case-oriented view.]
Teacher:	There's only one at 70 and the rest are at 60 or lower? Yes?
	Can you say something about the mean perhaps?
Rick:	The mean is usually somewhere around 50.

The written answers of the three students were the following:

Ruud:	Mine looks very much like what is on the blackboard.
Chris:	The middle-most [diagram on the blackboard] best resembles mine because the weights are close together and that is also the case in my graph. It lies between 35 and 75 [kg].
Sandra:	The other [real data] are more weights together and mine are further apart.

Ruud's answer is not very specific, like most of the written answers in the first cycle of growing samples. Chris used the predicate "close together" and added numbers to indicate the range, probably as an indication of spread. Sandra used such terms as "together" and "further apart," which address spread. The students in the class used common predicates such as "together," "spread out" and "further apart" to describe features of the data set or the graph. For the analysis it is important to note that the students used predicates (together, apart) and no nouns (spread, average) in this first cycle of growing samples. Spread can only become an object-like concept, something that can be talked about and reasoned with, if it is a noun. In the semiotic theory of Peirce (1976), such transitions from the predicate "the dots are spread out" to "the spread is large" are important steps in the formation of concepts.

Second cycle of growing a sample

The students generally understood that larger samples would be more reliable. With the feedback students had received after discussing the samples of ten data points in dot plots, students had to predict the weight graph of a whole class of 27 students and of three classes with 67 students (27 and 67 were the sample sizes of the real data sets of eighth graders of another school).

c. We will now have a look how the graph changes with larger samples. Predict a sample of 27 students (one class) and of 67 students (three classes).

d. You now get to see real samples of those sizes. Describe the differences. You can use words such as majority, outliers, spread, average.

During this second cycle, all of the students made dot plots, probably because the teacher had shown dot plots on the blackboard, and because dot plots are less laborious to draw than value bars (only one student started with a value-bar graph for the sample of 27, but switched to a dot plot for the sample of 67). The hint on statistical terms was added to make sure that students' answers would not be too superficial, as often happened before, and to stimulate them to use such notions in their reasoning. It was also important for the research to know what these terms meant for them. When the teacher showed the two graphs with real data, there was once again a short class discussion in which the teacher capitalized on the question of why most student prediction now looked pretty much like what was on the blackboard, whereas with the earlier predictions there was much more variation. No student had a reasonable explanation, which indicates that this was an advanced question. When comparing their own graphs (Figure 3a) with real data (Figure 3b), the same three students wrote:

Ruud:	My spread is different.
Chris:	Mine resembles the sample, but I have more people around a certain weight and I do not really have outliers, because I have 10 about the 70 and 80 and the real sample has only 6 around the 70 and 80.
Sandra:	With the 27 there are outliers and there is spread; with the 67 there are more together and more around the average.

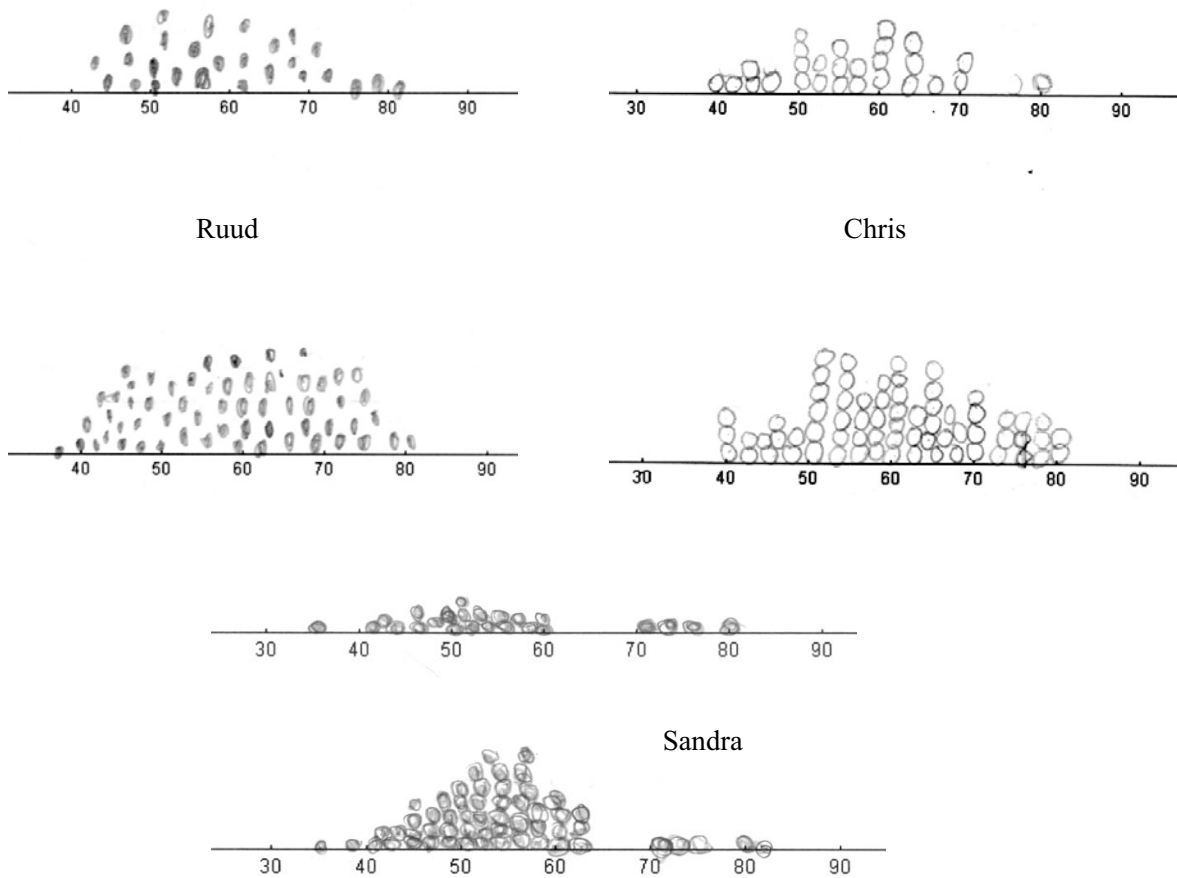


Figure 3a. Predicted graphs for one and for three classes by Ruud, Chris, and Sandra

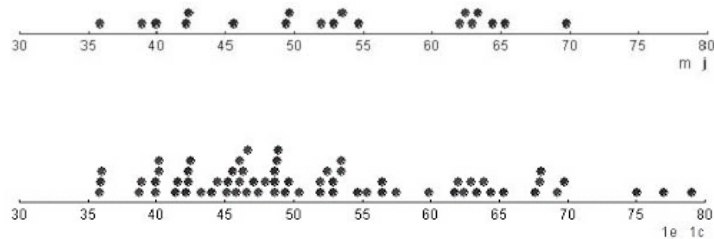


Figure 3b. Real data sets of size 27 and 67 of students from another school

Here, Ruud addressed the issue of spread. Chris was more explicit about a particular area in her graph, the category of high values. She also correctly used the term “sample,” which was newly introduced in the second lesson. Sandra used the term “outliers” at this stage, by which students meant “extreme values” (not necessarily exceptional or suspect values). She also seemed to locate the average somewhere and to understand that many students are about average. These examples illustrate

that students used statistical notions for describing properties of the data and diagrams. From a statistical point of view, these terms were not very precise. With “mean” students generally meant “about average” or “the middle typical group”; with “spread” they meant “how far the data lie apart”. And with “sample” they seemed to mean just a bunch of people, not necessarily the data as being representative for a population (cf. Schwartz et al., 1998).

In contrast to the first cycle of growing a sample, students used nouns instead of just predicates for comparing the diagrams. Ruud (like others) used the noun “spread,” whereas students earlier used only predicates such as “spread out.” Of course, this does not always imply that if students use these nouns that they are thinking of the right concept. Statistically, however, it makes a difference whether we say, “the dots are spread out” or “the spread is large.” In the latter case, spread is an object-like entity that can have particular aggregate characteristics that can be measured (for instance by the range, the interquartile range, or the standard deviation). Other notions, outliers, sample, and average, are now used as nouns, that is as conceptual objects that can be talked about and reasoned with.

Third cycle of growing a sample

The aim of the hypothetical learning trajectory was that students would come to draw continuous shapes and reason about them using statistical terms. During teaching experiments in the seventh-grade experiments (Bakker & Gravemeijer, 2004), experiments in two American sixth-grade classes, and a visit to an American group of ninth graders, reasoning with continuous shapes turned out to be difficult to accomplish, even if it was asked for. It often seemed impossible to nudge students toward drawing the general, continuous shape of data sets represented in dot plots. At best, students drew spiky lines just above the dots. This underlines that students have to construct something new (a notion of signal, shape, or distribution) with which they can look differently at the data or the variable phenomenon.

In this last cycle of growing the sample, the task was to make a graph showing data of all students in the city, not necessarily with dots. The intention of asking this was to stimulate students to use continuous shapes and dynamically relate samples to populations, without making this distinction between sample and population explicit yet. The conjecture was that this transition from a discrete plurality of data values to a continuous entity of a distribution is important to foster a notion of distribution as an object-like entity with which students could model data and describe aggregate properties of data sets. The task proceeded as follows:

e. Make a weight graph of a sample of all eighth graders in Utrecht. You need not draw dots. It is the shape of the graph that is important.

f. Describe the shape of your graph and explain why you have drawn that shape.

The figure of the same three students are presented in Figure 4 and their written explanations were:

Ruud:	Because the average [values are] roughly between 50 and 60 kg.
Chris:	I think it is a pyramid shape. I have drawn my graph like that because I found it easy to make and easy to read.
Sandra:	Because most are around the average and there are outliers at 30 and 80 [kg].

Ruud’s answer focused on the average group, or “modal clump” as Konold and colleagues (2002) call such groups in the center. During an interview after the fourth lesson, Ruud literally called his graph a “bell shape,” though he had probably not encountered that term in a school situation before (three other students also described their graphs as bell shapes). This is probably a case of reinvention. Chris’s graph was probably inspired by line graphs that the students made during mathematics lessons. She introduced the vertical axis with frequency, though such graphs had not been used before in the statistics course. Sandra probably started with the dots and then drew the continuous shape.

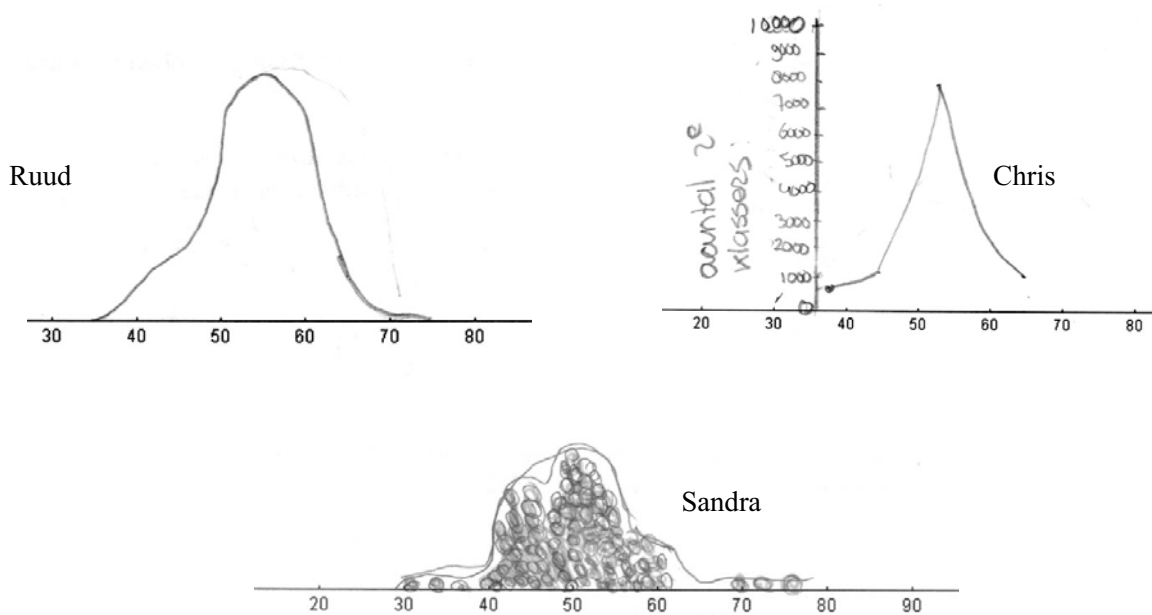


Figure 4. Predicted graphs for all students in the city by Ruud, Chris, and Sandra

In this third cycle of growing a sample, 23 students drew a bump shape. The words they used for the shapes were pyramid (three students), semicircle (one), and bell shape (four). Although many students draw continuous shapes, I did not exactly know what these shapes meant for them. Therefore, in the next section, I analyze students' reasoning with such shapes in the sixth lesson, which built on the fourth lesson. Furthermore, almost all student graphs looked roughly symmetrical, which is not surprising when the history of distribution is taken into account (Steinbring, 1980). In real life, however, the phenomenon of weight shows distributions that are skewed to the right. The skewness of weight data is caused by a "left wall effect" (two students had in fact drawn a left wall in the fourth lesson). By a left wall I mean that the lower limit (say about 30 kg) is relatively close to the average (53 kg) and the upper limit is relatively far away from the average (for example, sumo wrestlers can weigh 350 kg). The lower limit of 35 kg serves as a left wall, because adults can hardly live if they are lighter than 30 kg. This left wall in combination with no clear right wall causes the distribution to be skewed to the right. So far we had focused on spread and center as the core aspects of distribution, but skewness is another important characteristic of a distribution. Once there are different shapes to talk about, for example symmetrical or skewed, students can characterize shapes with different predicates. According to the hypothetical learning trajectory, skewness therefore had to become a topic of discussion as well in the following lessons. The next section shows how this was accomplished in the sixth lesson.

3.2. REASONING ABOUT SHAPES

In collaboration with the teacher, the following activity was designed with the purpose to make shape and in particular skewness a topic of discussion. To focus the students' attention on shape and skewness, the five shapes depicted in Figure 5 were drawn on the blackboard. They included three shapes mentioned by the students (a semicircle, a pyramid, a bell shape) and two skewed shapes (one unimodal distribution skewed to the right, and one skewed to the left). Students had to explain which shapes could not match the general distribution of people's weights based on their knowledge. The teacher expected that it would be easier for students to engage in the discussion if they could argue which shapes were not correct, instead of defending the shape they had chosen.

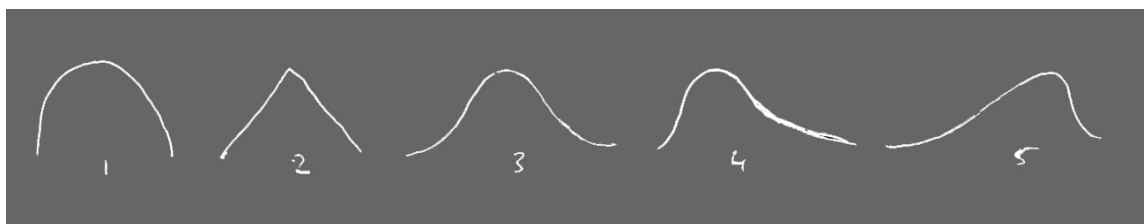


Figure 5. Five shapes as drawn on the blackboard: (1) semicircle, (2) pyramid, (3) normal distribution, (4) distribution skewed to the right, (5) distribution skewed to the left

The teacher chose students from the groups who thought that a particular shape on the blackboard could not be right. For all shapes except the normal shape, many students raised their hands. Apparently, most students expected a “normal” shape (number 3 in Figure 5).

1. First, Gigi explained why the semicircle (1) could not be right.

Gigi:	Well, I thought that it was a strange shape (...) For example, I thought that the average was about here [a little to the right of the middle] and I thought this one [top of the hill] was a little too high. It has to be lower. And I thought that here, that it was about 80, 90 [kg], and I don't think that so many people weigh that much or something [points at the height of the graph at the part of the graph with higher values].
Teacher:	(...) Does everybody agree with what Gigi says?
Tom:	Yes, but I also had something else. That there are no outliers. That it is straight and not that [he makes a gesture with two hands that looks like the tails of a normal distribution]. I would expect it to slope more if it goes more to the outside [makes the same gesture].

These students used statistical notions such as “outliers”, although in an unconventional way, and height to explain shape issues, especially frequency. Furthermore, they used their knowledge of the context to reason about shape.

2. Because all of the students seemed to agree that the semicircle was not the right shape, the teacher wiped it off the blackboard and turned to the pyramid shape (2). This discussion involved “outliers” (the extreme values) and the mean in relation to shape.

Mourad:	Well, I didn't think this was the one, because, yeah, I don't think that a graph can be that rectangular.
Teacher:	The graph is not so rectangular? [inviting him to say more]
Mourad:	No, there are no outliers or stuff.
Alex:	It does have outliers; right at the end of both it does have outliers.
Student:	That is just the bottom [of the graph].
Alex:	At the end of the slanting line, there is an outlier, isn't it? (...)
Anna:	But the middle is the mean and everything else is outlier. [Other students say they do not agree, e.g. Fleur]
Fleur:	Who says that the middle is the mean?
Anna:	Yes, yes, roughly then.
Teacher:	Tom, you want to react.
Tom:	Look, if you have an outlier, then it has to go straight a bit [makes a horizontal movement with his hands]; otherwise it would not be

	an outlier (...) but that is not what I wanted to say. I wanted to react, that it [this graph] could not be the right one, because the peak is too sharp and then the mean would be too many of exactly the same.
Mike:	He just means that of one weight exactly all these kids have the same weight, so if the tip is at I-don't-know-how-many kilos, maybe 60 kilos, that all these kids are exactly 60 kilos.

This transcript shows that students started to react to each other. Before this lesson they mainly reacted to questions from the teacher, a type of interaction that is very common unfortunately (Van den Boer, 2003). In other words, the activity stimulated students to participate and their passive attitude started to change. Because the students agreed that the pyramid was not the right shape, the teacher wiped this shape off the blackboard also.

3. Next, Sofie was asked to explain why the bell shape (3) could not be the right shape. Before the discussion, almost all of the students thought this was the right shape (one girl admitted she did not know).

Sofie:	I had it that this was not the one, because there are also kids who are overweight. Therefore, I thought that it should go a bit like this [draws the right part a little more to the right, thus indicating a distribution skewed to the right, like Figure 5.4]. (...)
Rick:	That means that there are more kids much heavier, but there are also kids much less, so the other side should also go like that [this would imply a symmetrical graph].
Tom:	Guys, this is the right graph!

Because there was no agreement, the teacher did not wipe the graph off the board.

4. Next, Mike had to explain why he thought that the fourth, skewed graph could not be right.

	I thought that this was not it because... if the average is perhaps, if this is the highest point, then this [part on the left] would be a little longer; then it would have a curve like there [left half of the third graph]. I think that this cannot be right at all, and I also find it strange that there are so many high outliers. Then you would maybe come to 120 kilos or so. [Note that there were no numbers in the graphs.]
--	---

5. Last, Ellen spoke about the fifth graph, which was skewed to the left:

	Well, I think this one is also wrong because there are more heavy people than light people. And I think that eighth graders are more around 50 kilos. That's it.
--	--

Tom then objected, "it says 50 nowhere," and a lively discussion between the two evolved.

Thus, as intended, skewness became a topic of discussion, even in relation to center and "outliers". Next time we would certainly want to pay more attention to what students mean by such terms. Some students argued that the mean need not be the value in the middle. Still students seemed to make no clear distinctions between midrange, mean, and mode. Because the mode is not a measure that is often used in statistics, it was not the intention to address the mode unless students were

already reasoning with it. Since students at this point argued about the mean versus the value that occurred the most, the teacher and I decided to introduce a name for the mode, which these students had not learned before.

Researcher:	The value that occurs the most often has a name; it is called the mode [pointing at the value where the distribution has its peak]. (...) Who can explain in this graph [skewed to the right] whether the mean is higher or lower than the mode? (...)
Rick:	There are just more heavy people than light people, and therefore the mean is higher. (Note: Rick's remark makes sense if we interpret it to mean that heavy people are those to the right of the mode and light people those left of the mode).

In this way, there were opportunities to introduce statistical terms and relate them to each other, because students were already talking about the corresponding concepts or informal precursors to them. Traditionally, the mode is just introduced as the value that occurs the most, but here it was introduced as a characteristic of a distribution, albeit informally. The median was introduced later, in the ninth lesson, as the value that yields two equal groups (as can be done in *minitool 2*).

4. DISCUSSION

The main question of the overall research was: how can we promote coherent reasoning about variability, sampling, data, and distribution in a way that is meaningful for students with little statistical background? By carrying out several teaching experiments, some instructional activities turned out to be more effective than others. The activities analyzed in this paper, concerning growing a sample and reasoning about shapes, appeared particularly engaging and useful. The purpose of this paper was therefore to analyze students' learning process as exemplified in these two instructional activities and learn more about their development of concepts underlying statistical data analysis. I used Peirce's semiotics as an instrument of analysis in order to detect what the crucial elements of those activities were and what kind of learning these activities supported. The next section (4.1) discusses those key elements for each activity and speculates about what can be learned from the analysis presented here. The last two sections address the limitations (4.2) and implications (4.3) of the study.

4.1. THE INSTRUCTIONAL ACTIVITIES

Growing a sample (fourth lesson). The activity of growing a sample involved short cycles of constructing diagrams of new hypothetical situations, and comparing these with other diagrams of a real sample of the same size. The activity has a broader empirical basis than just the teaching experiment reported in this paper, because it emerged from a previous teaching experiment (Bakker & Gravemeijer, 2004) as a way to address shape as a pattern in variability and also resembles the growing samples activity described by Konold and Pollatsek (2002) in a fifth-grade classroom.

The activity was also based on design heuristics that were defined during previous teaching experiments (Bakker, 2004). One of those heuristics is to sometimes stay away from data so as to avoid students adopting a case-oriented view. Also, by asking students to compare their own diagrams with those representing real data, we invite them to "compare forests instead of trees"—compare data sets instead of individual data points. Moreover, by letting students predict a situation, the need is created to use conceptual tools for predicting that situation. By the cyclic approach taken such design heuristics for statistics education were validated by the research.

In the design of the activity several other issues also played a role. First, the teacher and I have wondered if the context of weight was suitable for this age group (13 years old). Many teachers and

textbooks avoid this context because it is so sensitive, but we found it striking how well students knew this context and how their predictions resembled the actual samples in many respects. The delicacy of this subject might explain part of their engagement during class discussions.

Another important pedagogical issue is the length of class discussions. In earlier lessons the teacher and I had noticed that these students found it hard to concentrate during class discussions for longer than about ten minutes. A cycle of producing a diagram for a sample of a specific size, and comparing it with a real sample requires only short periods of concentration. Providing real data in between their inventions demanded short periods of reflection and feedback. We also promoted more individual work than in a previous experiment so as to give all students the opportunity to predict and reflect themselves as opposed to listen to other students during one long class discussion.

As a way to generalize the results, I analyzed students' reasoning as an instance of diagrammatic reasoning, which typically involves constructing diagrams, experimenting with them, and reflecting on the results of the previous two steps. More generally, Bakker and Hoffmann (in press) argue that diagrammatic reasoning forms an opportunity for concept development. In this growing samples activity, the quick alternation between prediction and reflection during diagrammatic reasoning appears to create ample opportunities for hypostatic abstraction, for instance of the notion of spread.

In the first cycle involving predicting a small data set, students noted that the data were more spread out, but in subsequent cycles, students wrote or said that the spread was large. From the terms used in this fourth lesson, I conclude that many statistical concepts such as center (average, majority), spread (range and range of subsets of data), and shape had become topics of discussion (hypostatic abstractions) during the growing samples activity. Some of these words were used in a rather unconventional way, which implies that students needed more guidance at this point. Shape became a topic of discussion as students predicted that the shape of the graph would be a semicircle, a pyramid, or a bell shape, and this was exactly what the hypothetical learning trajectory (HLT) aimed at. Given the students' minimal background in statistics and the fact that this was only the fourth lesson of the sequence, the results were quite promising. Note, however, that such activities cannot simply be repeated in other contexts; they always need to be adjusted to local circumstances if they are to be applied in other situations.

Reasoning about shapes (sixth lesson). The aim of reasoning about shapes was that students would learn to reason about skewed shapes, and they did so in terms of the context (e.g. using "heavy" and "light"). The satisfactory outcome of this activity was that the students came to reason with statistical notions in a way they had not demonstrated before and were more engaged in the discussion than we had observed before, and this included students with low grades for mathematics.

To learn from the results, I speculate on the crucial features of this activity. First, the lack of formal rules and definitions probably makes it easier for low-achieving students to participate in the discussion. I furthermore conjecture that the lack of data, the game-like character and students' knowledge about the context were important factors, but also the fact that they had to argue *against* certain shapes. Such reasoning is safer than defending the shape they think is right. Note that the way in which mode, average, and other statistical notions were discussed contrasts drastically with what is common in statistics curricula because average values and spread were discussed in relation to shape, not just as computational operations on data values.

As mentioned in the beginning of the paper, I strove for a process of guided reinvention. This notion hints at the challenge of striking the balance between giving guidance to students on the one hand and giving them the freedom to reason using their own terminology on the other. This issue can be illustrated with a metaphor that Frege wrote to Hilbert in 1895 (Frege was one of the first modern logicians and philosophers of language, and Hilbert was a formalist mathematician). The topic was using and making symbols in mathematical discourse.

I would like to compare this with lignification [transformation into wood]. Where the tree lives and grows, it must be soft and sappy. If, however, the soft substance does not lignify, the tree cannot grow higher. If, on the contrary, all the green of the tree transforms into wood, the growing stops. (Frege, 1895/1976, p. 59; translation from German)

On the one hand, if statistical concepts are defined before students even have an intuitive idea of what these concepts are for (such as mean, median, and mode), then the tree transforms into wood and

students' conceptual development can be hindered (as discussed in the beginning of this paper). On the other hand, if teachers and instructional materials do not guide students well in a process of reinvention, the tree stays weak and cannot grow higher. It is evident that the notions of average, outliers, distribution, and sample in the present research needed to be developed into more precise notions, but at least students developed a language that was meaningful to them, an image that could be sharpened later on or, staying with the metaphor, a sappy part of the tree that can be lignified later. With reference to the lignification metaphor, the teacher and I had been reasonably successful in getting students to participate in reasoning about these shapes. However, they often used terms (in particular "outliers") in unconventional or vague ways, which is not surprising given the small number of lessons (ten in total) of the teaching experiment.

In terms of diagrammatic reasoning, this lesson was mainly devoted to reflection on shapes, but there were also examples of mental experimentation (what would the shape look like if...). Skewness was addressed within the weight context, but had not been predicated yet in terms of "left-skewed" or "right-skewed." Students mainly used two distribution aspects in their reasoning, average and the tails (what they called "outliers"). These notions are hypostatic abstractions that have become reasoning tools. From the analysis I concluded that students probably had the following understanding of distribution: there are many values around average (high rounded part in the sketch) and few low and high values, which is evidenced by the horizontal tails of the shape. This was indeed aimed at in the hypothetical learning trajectory.

4.2. LIMITATIONS

The purpose of analyses such as the one presented in this paper is that researchers and teachers can adjust such instructional activities to their own circumstances. A hypothetical learning trajectory always remains hypothetical, but others may learn from it, provided the conditions in which the design research has been conducted are clear to the audience. According to Freudenthal (1991, p. 161),

[design] research means: experiencing the cyclic process of development and research so consciously, and reporting on it so candidly that it justifies itself, and that this experience can be transmitted to others to become like their own experience.

It is therefore necessary to highlight the conditions and limitations of this study, which we do in this section.

Relevant information to judge the results in this paper is first that the teacher was experienced (11 years of teaching) and was preparing her dissertation in mathematics education. The other adults in the classroom were myself and assistants, who interviewed and observed students, and avoided teaching. Yet the mini-interviews probably had a learning effect, which means the interview questions we asked should be considered part of the HLT. The growing samples activity was successful in one seventh and one eighth grade class, but the activity of reasoning about shapes was only carried out in a single eighth grade class. Unlike with the growing samples activity we do not have more empirical support for the value of the reasoning about shapes activity than from this one class of 30 students. The school was not exceptional for a *havo-vwo*-school in the center of a large Dutch city, which implies that students probably belonged to the top 40% of Dutch students.

Researchers who would like to repeat such activities also need to take into account that we asked students about sample size from the first lesson onwards and that we tried to foster a classroom culture in which students were willing to discuss. This was not at all easy, because they were not used to whole-class discussion, but like Dutch students, used to self-reliant working on small tasks of a computational nature. We were therefore pleasantly surprised that a student, near the end of the teaching experiment, characterized statistics as "a little arithmetic and a lot of thinking."

As mentioned in the analysis of the growing samples activity, students' diagrams were strongly influenced by the two minitools they had used, but they also used line graphs taught in mathematics lessons. It is hard to decide whether the use of the minitools limited students' own diagrams to those provided by the software or whether it inspired them to make diagrams they would not have made without prior experience with the software. An argument for the former is that students did not make

any other type of graph than they had used with the minitools or in mathematics lessons. An argument for the latter is that the minitool representations were apparently meaningful to them despite the short exposure to them (by the third lesson they had only seen the second minitool, but had no hands-on experience with it). Other research designs such as comparative studies may be needed to decide such issues.

Unlike with the instructional materials developed for grade 7, the statistics unit for grade 8 has not been implemented in a school. Based on the experiences with two novice teachers who used the materials for grade 7, I expect that other teachers than the experienced teacher I worked with would need more time to reach similar results and it is possible that their students would not reach the same quality of reasoning in a first attempt to teach the unit without researchers and assistants interviewing in the classroom. The results presented in this paper therefore need to be interpreted as being possible to recreate given favorable conditions of sufficient time and support being available.

4.3. IMPLICATIONS

As a springboard to implications for teaching, assessment, and research, I raise the following question: why do almost all school textbooks follow the same routes and introduce mean, median, and mode as a trinity, and provide students with graphical tools such as histogram and box plot long before students have the conceptual understanding to use such tools sensibly? G. Cobb (1993, parag. 53) compared the situation with a night picture of a city: “if one could superimpose maps of the routes taken by all elementary books, the resulting picture would look much like a time-lapse night photograph of car taillights all moving along the same busy highway”. Apart from the phenomenon of copying what others do, one important reason for this phenomenon could be that mean, median, mode, and graphs seem so easy to teach and, even more importantly, to assess. As argued in the beginning of this paper, however, this view easily leads to superficial understanding if students are not provided with ample opportunities to develop conceptual understanding of these statistical notions and graphs. The instructional activities presented in this paper are attempts to give students such opportunities.

Teaching. For several reasons, the approach taken in this paper is challenging for the teacher. The teacher plays an important role in steering the topic of discussion towards statistically important issues such as center and spread. This requires establishing a classroom culture in which students are willing to engage in discussions, which can be hard if they are used to working self-reliantly.

As argued before, the episodes analyzed in this paper can be framed as instances of diagrammatic reasoning, the key steps of which are making a diagram, experimenting with it, and reflecting on the results. During this diagrammatic reasoning, hypostatic abstractions such as majority, average, and shape can become objects and reasoning tools in the discourse. The analyses suggest the following recommendations that are not tied to the particular instructional activities considered here.

First, it is clear students need to diagrammatize—make their own diagrams that make sense to them, but also learn to make powerful conventional types of diagrams that are likely to become meaningful to them. The results show that the minitools software had a large influence on the diagrams students made themselves (see Section 4.2 for a discussion of this issue).

Second, students need to experiment with diagrams. Educational software such as the minitools can be useful in this stage of diagrammatic reasoning. The software should offer diagrams that students understand, but it should also offer opportunities for learning more advanced, culturally accepted diagrams. Apart from physical experimentation, mental experimentation is important, for instance when answering questions about hypothetical situations (Bakker & Gravemeijer, 2004).

Third, reflection should be stimulated. Throughout the research we noticed that the best reasoning occurred during teacher-directed class discussions that were not in the computer lab. One of the core issues of the reflection step is that students learn to describe (“predicate”) and predict aggregate features of data sets, because that is an essential characteristic of statistical data analysis. Predicates should become topics of discussion so that they can be taken as entities in themselves. For example, talking about “most” data can lead to talking about the “majority”; describing how dots are “spread

out” can lead to saying that “the spread is large.” These are examples of what Peirce called “hypostatic abstraction.”

It is striking that these steps of diagrammatic reasoning, though they appear to be crucial to learning statistics, are so underexposed in most school textbooks. If we accept that diagrammatic reasoning is a basis for concept development (cf. Bakker & Hoffmann, in press), the above options are worth considering. A possible sequence that teachers could follow is (1) let students make their own diagrams but also offer types of diagrams that are likely to become meaningful, (2) enable students to experiment with diagrams both physically (e.g. using software) and mentally (e.g. by asking what-if questions), and (3) involve students in a reflection step in which they describe precisely what they see (clumps, majorities, shapes) and where they see it in a diagram.

Assessment. The learning that results from an approach of the form taken here may be harder to assess than whether students have learned to calculate average values or draw a histogram. In the approach taken here, the teacher has to accept that students’ notions stay informal for a while, provided enough effort is taken to make informal notions more precise. In countries and states with high-stake accountability for the assessment of students’ progress, this may be difficult to accomplish (cf. Makar & Confrey, 2004). We therefore need assessment items that assess what we find important and that might be used on large-scale tests.

Research. More research is needed into the question of how students can develop their own informal notions, such as center clumps, spread, and shapes, into conventional measures of center, variation, and other distribution aspects, and how teachers can support this development. The semiotic analysis suggests that one key issue is that the topic of a group discussion should be clear, and the teacher plays an important role in directing students’ attention to that topic. Research is needed into the question of how teachers can be supported to help students in this regard.

ACKNOWLEDGMENTS

I thank Koeno Gravemeijer for his advice during the research, Corine van den Boer for teaching the eighth-grade class, and Carolien de Zwart, Sofie Goemans, and Yan-Wei Zhou for assisting in multiple ways. I also thank Nathalie Kuijpers for translating the transcripts and correcting the English, Jantien Smit, Anneleen Post, Katie Makar, and Phillip Kent for their editing help, and the editor Iddo Gal and anonymous reviewers for their helpful suggestions. The research was funded by the Netherlands Organization for Scientific Research under grant number 575-36-003B.

REFERENCES

- Bakker, A. (2003). The early history of statistics and implications for education. *Journal of Statistics Education, 11*(1). [Online: www.amstat.org/publications/jse/v11n1/bakker.html]
- Bakker, A. (2004). *Design research in statistics education: On symbolizing and computer tools*. Utrecht, The Netherlands: CD-Beta Press. [Online: <http://www.library.uu.nl/digiarchief/dip/diss/2004-0513-153943/inhoud.htm>]
- Bakker, A., & Gravemeijer, K. P. E. (2004). Learning to reason about distribution. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 147–167). Dordrecht, the Netherlands: Kluwer Academic Publishers.
- Bakker, A., & Hoffmann, M.H.G. (in press). Diagrammatic reasoning as the basis for developing concepts: A semiotic analysis of students’ learning about statistical distribution. *Educational Studies in Mathematics*.
- 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.
- Cobb, G. (1993). Considering statistics education: A National Science Foundation Conference. *Journal of Statistics Education, 1*(1). [Online: www.amstat.org/publications/jse/v1n1/cobb.html]

- Cobb, P. (1999). Individual and collective mathematical development: The case of statistical data analysis. *Mathematical Thinking and Learning*, 1(1), 5–43.
- Cobb, P., Confrey, J., diSessa, A. A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher*, 32(1), 9–13.
- Cobb, P., Gravemeijer, K. P. E., Bowers, J., & Doorman, M. (1997). *Statistical Minitools* [applets and applications]. Nashville, TN & Utrecht: Vanderbilt University & Freudenthal Institute, Utrecht University. [Translated and revised in 2001 by Bakker, A.] [Online: www.wisweb.nl/en]
- Cobb, P., McClain, K., & Gravemeijer, K. P. E. (2003). Learning about statistical covariation. *Cognition and Instruction* 21, 1–78.
- Frege, F. L. G. (1976). *Wissenschaftlicher Briefwechsel [Scientific correspondence]* (Gabriel, G., Ed.) (First ed. Vol. 2). Hamburg, Germany: Meiner.
- Freudenthal, H. (1991). *Revisiting mathematics education: China lectures*. Dordrecht, the Netherlands: Kluwer Academic Publishers.
- Garfield, J., & Ben-Zvi, D. (2004). Research on statistical literacy, reasoning, and thinking: Issues, challenges, and implications. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 397–409). Dordrecht, the Netherlands: Kluwer Academic Publishers.
- Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory; Strategies for qualitative research*. Chicago: Aldine Publishing Company.
- Gravemeijer, K. P. E. (1994). Educational development and developmental research. *Journal for Research in Mathematics Education*, 25, 443–471.
- Gravemeijer, K. P. E. (1999, April). An instructional sequence of analysing univariate data sets. Paper presented at the Annual Meeting of the American Education Research Association, Montréal, Canada.
- 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. (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., & Pollatsek, A. (2002). Data analysis as the search for signals in noisy processes. *Journal for Research in Mathematics Education*, 33, 259–289.
- Konold, C., Robinson, A., Khalil, K., Pollatsek, A., Well, A. D., Wing, R., & Mayr, S. (2002). Students' use of modal clumps to summarize data. In B. Phillips (Ed.), *Proceedings of the International Conference on Teaching Statistics, Cape Town, South Africa*. [CD-ROM] Voorburg, The Netherlands: International Statistics Institute.
- Makar, K., & Confrey, J. (2004). Secondary teachers' reasoning about comparing two groups. In D. Ben-Zvi & J. Garfield (Eds.), *The challenges of developing statistical literacy, reasoning, and thinking* (pp. 353–373). Dordrecht, the Netherlands: Kluwer Academic Publishers.
- McGatha, M., Cobb, P., & McClain, K. (2002). An analysis of students' initial statistical understanding: Developing a conjectured learning trajectory. *Journal of Mathematical Behavior*, 21, 339–355.
- Mokros, J., & Russell, S. J. (1995). Children's concepts of average and representativeness. *Journal for Research in Mathematics Education*, 26, 20–39.
- Peirce, C. S. (1976). *The new elements of mathematics* (Eisele, C., Ed.) (Vol. I-IV). The Hague-Paris/Atlantic Highlands, N.J.: Mouton/Humanities Press.
- 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.
- Russell, S. J., & Corwin, R. B. (1989). Statistics: The shape of the data. *Used numbers: Real data in the classroom. Grades 4-6*. Washington, DC: National Science Foundation.

- Schwartz, D. L., Goldman, S. R., Vye, N. J., Barron, B. J., & Cognition and Technology Group at Vanderbilt (1998). Aligning everyday and mathematical reasoning: The case of sampling assumptions. In S. P. Lajoie (Ed.), *Reflections on statistics: Learning, teaching, and assessment in grades K-12* (First ed., pp. 233–273). Mahwah, NJ: Lawrence Erlbaum Associates.
- Simon, M. A. (1995). Reconstructing mathematics pedagogy from a constructivist perspective. *Journal for Research in Mathematics Education*, 26, 114–145.
- Steinbring, H. (1980). *Zur Entwicklung des Wahrscheinlichkeitsbegriffs - Das Anwendungsproblem in der Wahrscheinlichkeitstheorie aus didaktischer sicht*. [On the development of the probability concept - The applicability problem in probability theory from a didactical perspective]. Bielefeld: Institut für Didaktik der Mathematik der Universität Bielefeld.
- Van den Boer, C. (2003). *Als je begrijpt wat ik bedoel. Een zoektocht naar verklaringen voor achterblijvende prestaties van allochtone leerlingen in het wiskundeonderwijs* [If you know what I mean. A search for an explanation of lagging results of mathematics education among ethnic minority students]. [CDROM] Utrecht, The Netherlands: Beta Press.
- Van Oers, B. (2000). The appropriation of mathematics symbols: A psychosemiotic approach to mathematics learning. In P. Cobb, E. Yackel & K. McClain (Eds.), *Symbolizing and communicating in mathematics classrooms; Perspectives on discourse, tools, and instructional design* (pp. 133–176). Mahwah, NJ: Lawrence Erlbaum Associates.
- Wilensky, U. (1997). What is normal anyway? Therapy for epistemological anxiety. *Educational Studies in Mathematics*, 33, 171–202.

ARTHUR BAKKER
Since September 1, 2004:
Institute of Education
University of London
23-29 Emerald Street
London WC1N 3QS
United Kingdom