DESIGN RESEARCH AND DESIGN HEURISTICS IN STATISTICS EDUCATION

Koeno Gravemeijer
Utrecht University, Netherlands
Arthur Bakker
University of London, United Kingdom
koeno@fi.uu.nl

Design research projects can be characterized as iterative and theory based attempts simultaneously to understand and improve educational processes. To contribute to a framework for design in statistics education, this paper draws on two design research projects, one carried out by Cobb, Gravemeijer and colleagues in the USA and one by Bakker and Gravemeijer in the Netherlands, both focusing on distribution as a core concept in the instructional design. Both projects were inspired by the theory of realistic mathematics education, which includes design heuristics such as guided reinvention, historical and didactical phenomenology, and emergent modeling. Each of these heuristics is briefly illustrated with examples from these two projects.

INTRODUCTION

When we look at the field of research in statistics education, we may discern a variety of research designs such as surveys, effect studies, studies into levels of understanding, and design research, each of those having its own strengths and weaknesses. Currently design research, also denoted as design experiments (Cobb, Confrey, *et al*., 2003; Gravemeijer, 1998), draws a lot of attention. In this paper we want to make the point that design research can make a valuable contribution to the constitution of a framework for shaping statistics education. We will elaborate this point by focusing on design research based on the theory for realistic mathematics education (RME). We believe that it is too early to construe a complete framework but we want to present some ingredients based on RME-inspired design experiments that have been carried out in the field of statistics education, at the middle school level. After a discussion of design research and RME theory, we will especially focus on a phenomenological analysis of statistical concepts, and the role of modeling. For our purpose it is not necessary to differentiate between 'mathematics' and 'statistics' in this paper even though they may be conceived as two different fields. For sake of simplicity, we will use the terms 'mathematics' and 'mathematical,' because RME has its origin in mathematics, and because statistics at middle school level is often taught as part of the mathematics curriculum.

DESIGN AND RESEARCH

Design research is aimed at creating innovative learning situations, the objective of which can be both understanding certain aspects of that learning situation, and understanding how such a learning situation can be brought about. Design research projects can be characterized as 'iterative, situated, and theory based attempts simultaneously to understand and improve educational processes.' (diSessa and Cobb, 2004, p. 80). The kind of design research we would like to discuss here aims at developing instructional sequences together with a rationale, or instruction theory, that underpins it. Within this type of research, we may discern various design cycles of preparing, designing, testing, and revision. The heart of such a cycle is formed by a teaching experiment, in which a preliminary instructional sequence is elaborated, tried out, and adjusted (Gravemeijer and Cobb, in press). The design and revision are based on anticipatory thought experiments, in which the researcher/designer tries to envision how the students might think and reason when they participate in the proposed instructional activities, and on the actual trial of these activities in the classroom. At the end of the teaching experiment, an improved version of the conjectured instruction theory is construed on the basis of a retrospective analysis. The retrospective analysis may also concern more encompassing issues, such as the pro-active role of the teacher or the role of modeling.

Even though currently much attention goes to the rigor of the *research* component of design research methods, we would argue that the quality of the *design* component is just as important. The design part should be well considered, and well founded. This design can be supported by a domain-specific instruction theory such as the theory for realistic mathematics education, RME, which itself is the result of a long history of design research in the Netherlands. With an eye on instructional design, we may cast RME theory in terms of three instructional design heuristics, guided reinvention, didactical phenomenology, and emergent modeling (Gravemeijer, 1999).

The first heuristic, 'guided reinvention,' builds on Freudenthal's (1973) idea that mathematics education should not take its point of departure in ready-made mathematics. Instead, students should be given the opportunity to experience a process similar to the process by which mathematics was invented. According to this heuristic, the designer has to try and formulate a potential learning route along which a process of collective reinvention (or progressive mathematization) might be supported.

The second heuristic concerns the phenomenology of mathematics, and asks for a didactical phenomenological analysis. Freudenthal (1983) makes a distinction between a mathematical phenomenology, an historical phenomenology, and a didactical phenomenology. The mathematical phenomenology analyses how a mathematical thought object organizes mathematical phenomena, the historical phenomenology analyses how, and by what organizing activity, various concepts, procedures and tools are constituted over time, and the didactical phenomenology looks at both from a didactical perspective. Freudenthal speaks of mathematical phenomenology as analyzing how a mathematical 'thought object' (a concept, procedure or tool) organizes certain phenomena. If we presume that mathematical thought objects are initially invented to solve particular problems, these problem situations come to the fore as phenomena that are organized by certain mathematical objects. In this manner, we may find phenomena that 'beg to be organised' as Freudenthal (1983, p. 32) calls it. Such problems, in other words, might function as situations where students feel the need to invent a concept, tool, or procedure. Didactical phenomenological analysis, therefore, is closely connected with the idea of guided reinvention: it informs the researcher/designer about a possible reinvention route.

The third design heuristic, emergent modeling, is closely connected to *developing* more formal mathematical knowledge. In this respect, it differs from mathematical modeling, which more has the character of *applying* formal mathematical knowledge. The latter requires problem solvers to use their formal mathematical knowledge to construct a mathematical model that fits the problem situation. Emergent modeling denotes a long-term process that encompasses more informal forms of modeling, and which is meant to generate the kind of mathematical knowledge that problem solvers need to construct mathematical models.

In emergent modeling the objective is to help the students to model their own informal mathematical activity. 'Emergent' in this context means that the model and the situation are mutually constituted in the course of modeling activity: the model and the situation modeled co-evolve. It is only at the end of this process that the model and the situation modeled are constituted as two separate entities, and so emergent modeling eventually leads to mathematical modeling. In relation to this we may discern two forms of abstraction, or abstracting: one that concerns the activity of solving a given problem (which corresponds with mathematical modeling), and another that concerns the long-term process of developing more abstract mathematical knowledge (which corresponds with emergent modeling). What is aimed for, in emergent modeling, is that the model with which the students initially model their own informal mathematical activity gradually develops into a model for more formal mathematical reasoning. In this manner, emergent modeling can play an important role in the instruction that has to lay a basis for mathematical modeling. In short, the *model of* informal mathematical activity develops into a *model for* more formal mathematical reasoning.

PHENOMENOLOGICAL ANALYSIS

In the following, we will take the aforementioned design heuristics and the yield of RME-inspired design research as guidelines for distinguishing some components of a framework for statistics education. In doing so we will merge the guided reinvention and the didactical phenomenology heuristics, firstly because of the interrelatedness of the two, and secondly because the historical phenomenology that we will elaborate below touches upon both.

We will structure our overview by discerning two distinct ways of dealing with variation. The first concerns the use of statistical measures such as mean and median to eliminate variation by reducing the data to one number. The second tries to account for the variation by describing the distribution of the data—in part by using the same statistical measures. Design research carried out by Bakker (2004) shows that historically the reduction to one number mainly concerned the mean.

MEAN

The mean seems to have various historical sources. One of them is the implicit use of an estimated mean as a means to calculate the total of a series of unequal sets, which reportedly was used in India well before the fourth century AD (the examples in this section are elaborated in Bakker and Gravemeijer, in press). The strategy to find the total number of leaves on a tree, for instance, would be to try to pick—what we would call in modern terminology—an 'average branch.' That is to say, one has to look for the branch that will generate the best estimate for total number of leaves, when its number of leaves is multiplied with the number of branches. Here, the anticipated calculation is the inverse of the common way of calculating the mean, while, at the same time, the chosen average branch can be seen as representative for all other branches in respect tot the total number of leaves. Mark that the construal of the 'average branch' fits Freudenthal's idea of construing a mathematical 'thought object' as a means for organizing a certain phenomenon. Here the average branch is developed to get a handle on a set of varying quantities. Moreover, the idea of evening out the differences between the whole set of branches is very similar to reallocation, which refers to the conceptual understanding that underlies the algorithm for calculating the mean. Furthermore, a design experiment that was inspired by this ancient history showed that 7th-grade students spontaneously invented a similar strategy when asked to estimate the number of elephants in a photo.

Another implicit use of the mean—or maybe better, the use of a total as a substitute of the mean—can be found in examples such as the Trial of the Pyx, which took place at the Royal Mint of Great Britain from the twelfth to the nineteenth century. Every day one of the coins was put in a box, and after some time the box was weighed and a sample of coins was melted to investigate the pureness of the coins. The underlying notion of both strategies seems to be the realization that although individual cases may vary, one may expect (the total of) a group to be rather constant.

Eighth-grade students reasoned the same way when they were asked to estimate the median reaction time of an arbitrary 8th-grader that might walk into the room (Cobb, McClain, and Gravemeijer, 2003). They had just measured their own reaction times and had aggregated the individual median reaction times of the whole class into a class median reaction time. However, they did not see the median reaction time of their classroom as an indicator of the probable reaction time of an arbitrary 8th-grader. Instead, they referred to the large variation in their individual measurements to argue that you could not predict anything about that arbitrary 8th-grader. However, when asked about another 8th-grade classroom that might do the same experiment, they were unanimous that that classroom would yield similar results as they had obtained—with the same variation.

History shows that translating the predictability of the group into using the mean (or the median for that matter) as a representative value is a big step. We may, however, conjecture an instructional sequence that would begin with developing the insight that the total results of a group or a sample will be more or less stable. The next step would be made when the students start using the mean instead of the group total. Then one may expect that what the mean and the individual data signify for the students gradually changes, in that initially the mean summarizes the original data, while later the individual data are seen as variation around the mean value. We may take 'average travel time' as an example to illustrate this shift. We tend to think of the time we need to get from home to work as a given amount of time, with some variation. This average travel time, however, is a construct that has taken on a life of its own. Based on our experience, we have construed the average travel time, and instead of thinking of averaging over individual trips, we now tend to see individual trips as deviating from a fixed travel time.

As a third historical example we want to mention the practice of calculating the mean in astronomy as a way of dealing with measurement errors. Here one reasoned that the positive errors and the negative errors would roughly cancel each other out. Adding all measurement values and dividing the total by the number of measures would then result in a good estimate of the true value. Initially the actual distribution of the errors was of no concern. In the eighteenth century however, scholars started to speculate about the shape of that distribution.

DISTRIBUTION

The emerging interest for the distribution of measurement errors in physics and astronomy preceded the use of statistics in other areas. In the 19th century, historical figures such as Quetelet and Galton started to apply a similar kind of reasoning on phenomena of which the variation that was not caused by some measurement error, such as in demographic studies. Here, it was immediately clear that one value—whether the mean or the median—would not suffice to describe natural phenomena such as the heights of men. One would also have to describe the *distribution* of the individual measures. Again, we may observe the construal of a mathematical thought object, 'distribution,' as a means for organizing phenomena. The variety of distributions in different fields of interest then created the need for new descriptors, such as median and quartiles, and eventually new graphical representations, such as stem and leaf plots and box plots.

According to the design research by Bakker (2004), 7th-graders already have some notion of distribution. When asked to describe or predict data sets of weight of children, for example, they typically characterized such natural phenomena as consisting of a large group that has more or less the normal value, a smaller group with smaller values and a similar group with bigger values. Bakker also designed a lesson in an 8th-grade classroom, where the students were challenged to elaborate their implicit ideas on distribution. They were asked to criticize various possible shapes of distributions that they predicted and that were also discussed in history by various scholars; this included a semi-circle and a triangle. In that discussion, students used common-sense reasoning to argue such shapes would be unlikely; instead they preferred a bell curve, which some students came up with themselves. They even came to realize that the distribution of the weights of all 8th-graders in the country would be skewed, due to the fact that there are more children with overweight than with underweight.

In statistics courses, statistical measures like mean, mode, median, spread, quartiles are often taught as a set of independent definitions. We may argue, however, that these statistical measures actually are means to characterize distributions, and that the same holds for conventional representations such as histogram and box plot. Based on this line of reasoning, *distribution* was chosen as a central statistical idea that serves as a potential end point of an instructional sequence in the design experiments described by Cobb, McClain, and Gravemeijer (2003; see also Gravemeijer and Cobb, in press). Here we may again apply the didactical phenomenology design heuristic, which would mean that we have to start by analyzing the notion of distribution as an object. This then may lead to the conclusion that we may think of distribution in terms of shape and density. Following this line of thought, we may conclude that shape and density are the phenomena that are organized by distribution as an object. Density and shape in turn organize collections of data points in a space of possible data values. In practice, this may be visualized as a dot plot, showing data points on an axis. These data points on an axis can be seen as a means of organizing data values. In relation to this, we may note that Hancock, Kaput, and Goldsmith (1992) and many other researchers found that middle school students had difficulty with viewing data as a plurality of values of one variable. They tended to see data as attributes of individuals. For example, 'five foot' would not be seen as a value of the variable length but as a characteristic of Anna. A first step therefore would involve supporting a shift in attention from attributes to the notion of values of a variable.

The above phenomenological analysis offers guidelines for the design of a reinvention route along which students may develop the notion of distribution as an object via a process of progressively organizing phenomena. A key point of departure of the design experiment, as it was carried out by Cobb, Gravemeijer and colleagues, was that the students had to analyze data sets for a reason. The students had to organize and compare data sets in answer to a question or a problem. Along the way, students were to develop statistical notions such as median, quartiles, shape, and density. The first form of organizing consisted of measuring, which usually involved creating some sort of dimension against which the phenomenon can be measured. The students, for instance, had to develop the notion of braking distance, and think of a way of measuring that, in order to compare the safety of cars of different brands. The actual data sets were then supplied by the teacher. The first step in organizing those data sets was in inscribing the data values in a magnitude-value-bar plot—in which each data point is visualized with a bar, with its length proportional to its data value. As a next step, the end points of those bars are plotted in the form of a dot plot. Then it became possible to talk about the shape of the distribution, and new organizing thought objects such as ‘shape,’ and ‘density’ could be construed—although the students used words like ‘the shape of the hill’ and ‘majority.’ As a final step, students came to view data sets as entities that are distributed within a space of possible values rather than a plurality of values. In other words, the students started to develop a notion of a distribution as an object with certain characteristics, such as spreadoutness, and symmetry or skewness.

MODELING

The above description suggests a strong intertwinement of didactical phenomenology, guided reinvention, and emergent modeling. Apart from new concepts, also new tools have to be developed in the subsequent organizing activities which clearly relate to emergent modeling. To clarify the emergent modeling design heuristic, it may be helpful to point to the three processes that are involved. One is that of the model/of-model/for shift, which in practice will be spread out over a series of sub-models. The second process involves the constitution of a chain of signification, in which each new sub-model supersedes the earlier sub-model. The third process concerns a shift in the level of activity, from more informal mathematical reasoning to more formal mathematical reasoning, which is tied to the construction of some new mathematical reality. All three components offer guidelines for instructional design. The first offers an alternative for the common use of ready-made tactile or visual models in mathematics education. Instead of offering students ready-made models that supposedly convey meaning, students are given the opportunity to construe both models and meaning for themselves. The second points to the idea of a cumulative sequence, in which each (sub-)model derives its meaning for the students from their experience with an earlier (sub-)model—so there is always a history to build on for the students. The third relates to the fact that we reserve the model/of-model/for terminology for transitions that encompass the constitution of some new mathematical reality. This requires the researcher/designer to consider and explicate what this new mathematical reality consists of. In this context, some new mathematical reality is to be understood as a framework of mathematical relations, which the students construct, together with the mathematical objects that derive their meaning from that framework.

We will elucidate these three points with the aforementioned design experiment. We start with the model-of/model-for shift. The overarching model may be described as a visual model of a data set. We argue that a ready-made visual model, such as a curve of a normal distribution, initially was not transparent to the students (see, e.g., Bakker, 2004). By starting with the more informal model of magnitude-value-bars to visualize a data set, students were given the opportunity to build on their earlier experiences with scale lines and simple graphs, and to develop the more sophisticated model and its meaning with help of the teacher. In practice this big model-of/model-for shift was paved with smaller transitions between sub-models.

We noted that the emergent modeling heuristic requires that each new sub-model builds on experience with acting with an earlier sub-model. In the actual instructional sequence, a computer applet was used to create the magnitude-value-bar graphs. This applet has various options that enable the students to sort the data, to separate two or more data sets, and to structure data sets on the basis of the lengths of the bars. The latter focused the attention of the students on the position of the end points in respect to the *x*-axis, when they were comparing data sets. This both fostered the notion of data point as possible values on a variable (signified by the *x*-axis), and prepared the students for working with the dot plots that could be created with a second applet. Thanks to this earlier experience, the dots of the dot plot signified the end points of value bars to students. Working with dot plots, they had to construe meaning for the height of the assembled dots, and the started to speak of ‘the hill,’ which they could interpret in the context of the problem. Subsequently, the use of computer tool options, such as splitting the data into two or three equal groups, helped the students to characterize distributions in terms of shape and density.

With uni-modal distributions, the smallest interquartile range became equated with what they called the 'majority,' and the median was construed as an indicator for where the 'hill' would be. This then enabled them to see 'the shape of the hill' in a box-plot type representation (see Cobb, McClain, and Gravemeijer, 2003).

Finally, the shift in the level of activity asks for a clarification of the framework of mathematical relations that is being developed by the students. In this example, these relations involve notions as shape, density, spread, skewness, but also median, quartiles, and extremes. The students did not fully develop the formal statistical language; rather these notions formed the background against which they started to form the notion of distribution as an object.

CONCLUSION

We had to leave various interesting results of design research in statistics undiscussed. We hardly discussed the idea of data creation, and did not speak of mean as a measure, nor of 'cultivating statistical interest,' or developing notions of sampling. Still we hope to have created an image of what a framework for statistics education might look like. And we have shown how design research and the design heuristics of RME can contribute to the development of such a framework. As a final remark, we like to add although we have tried to make a case for the value of design research, one should not infer that we want to disregard the value of other forms of research.

REFERENCES

Bakker, A. (2004). *Design Research in Statistics Education: On Symbolizing and Computer Tools*. Utrecht, the Netherlands: CD Beta Press.

Bakker, A. and Gravemeijer, K. P. E. (in press). A historical phenomenology of mean and median. *Educational Studies in Mathematics,* 63(1).

Cobb, P., Confrey, J., diSessa, A. A., Lehrer, R., and Schauble, L. (2003). Design experiments in educational research. *Educational Researcher,* 32, 9-13.

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

diSessa, A. A., and Cobb, P. (2004). Ontological innovation and the role of theory in design experiments. *Journal of the Learning Sciences*, 13, 77-103.

Freudenthal, H. (1973). *Mathematics as an Educational Task*. Dordrecht, the Netherlands: Reidel.

Freudenthal, H. (1983). *Didactical Phenomenology of Mathematical Structures*. Dordrecht, the Netherlands: Reidel.

Gravemeijer, K. P. E. (1998). Developmental research as a research method. In J. Kilpatrick and A. Sierpinska (Eds.), *Mathematics Education as a Research Domain: A Search for Identity*, Vol. 2, (pp. 277-295). Dordrecht: Kluwer Academic Publishers.

Gravemeijer, K. P. E. (1999). How emergent models may foster the constitution of formal mathematics. *Mathematical Thinking and Learning*, 1(2), 155-177.

Gravemeijer, K. and Cobb, P. (in press). Design research from the learning design perspective, In J. van den Akker, K. Gravemeijer, S. McKenney and N. Nieveen (Eds.), *Educational Design Research.* London: Routledge.

Hancock, C., Kaput, J. J., and Goldsmith, L. T. (1992). Authentic inquiry with data: Critical barriers to classroom implementation. *Educational Psychologist*, 27, 377-364.