EMERGENT MODELING AS THE BASIS FOR AN INSTRUCTIONAL SEQUENCE ON DATA ANALYSIS

Koeno Gravemeijer
Utrecht University
The Netherlands

This paper discusses an instructional design heuristic called "emergent modeling", with an instructional sequence on data analysis as an example. The emergent modeling approach is presented as an alternative for instructional approaches that focus on teaching ready-made representations. In relation to this, a distinction is made between modeling as "translation" and modeling as "organizing". Emergent modeling fits the latter. Within this perspective, the model and the situation modeled are mutually constituted in the course of modeling activity. This gives the label "emergent" a dual meaning. It refers to both the process by which models emerge, and the process by which these models support the emergence of more formal mathematical knowledge. This is reflected in the exemplary instructional sequence, in which the model co-evolves with the notion of distribution as an enty.

INTRODUCTION

In recent years, the role of symbols and models has become a central topic of attention in mathematics education. We might, in fact, speak of a renewed interest in models. Models have played a key role in mathematics education that is now labeled a "transmission model of teaching". Tacit and visual models were seen as powerful means to support learning for understanding. By acting with well-designed concrete models, students were expected to discover the mathematics that was embedded in the models. Gradually, however, this type of approach lost its credibility. Apart from problems with student understanding and proficiency, doubts emerged about its theoretical validity.

The core of the critique is that external representations do not come with intrinsic meaning, but that the meaning of external representations is dependent on the knowledge and understanding of the interpreter. This creates a problem that is known as "the learning paradox" (Bereiter, 1985), which can be captured with the question: *How is it possible to learn the symbolizations, you need to come to grips with new mathematics, if you have to have mastered this new mathematics to be able to understand those very symbolizations?*

We may ascribe the learning paradox to a view in which learning is seen as a process of acquiring mathematics as a ready-made system. In relation to this we may refer to Nemirovsky's (1994) notion of a symbol system. A symbol system is typically thought as an object that exists independent from any agent. A Cartesian graph, for instance, can be viewed as an example of a symbol system in and of itself. In this view, the mathematics is separated from the learner, and the learning paradox arises from this separation. The learning paradox dissolves when we adopt a more dynamic view of learning, which allows the students to reinvent mathematics. Within a reinvention approach, mathematical symbols and models could be developed in a bottom-up manner. In relation to this, we may quote Meira (1995), who proposes that "activity-oriented view that takes cultural conventions, such as notational systems, to shape in fundamental ways the very activities from which they emerge, at the same time that their meanings are continuously transformed as learners produce and reproduce them in activity" (p. 270).

Following Meira (1995), we may envision a dynamic process in which symbolizations and meaning co-evolve, and in which the ways that symbols are used and the meanings they come to have are seen to be mutually constitutive. An instructional design heuristic that takes this dialectic relation into account is the notion of "emergent modeling" that forms the theme of this paper.

EMERGENT MODELING

Before launching into an exposition on emergent models, however, it may be helpful to explain that this view of modeling differs from the more common notion of modeling as mathematical modeling. We may clarify the difference by distinguishing two forms of modeling, one that can be typified as "translation" and one that is better characterized as "organizing".

According to the mathematical modeling view, students have to translate problem situations into mathematical expressions that can function as models. From this perspective, it is important that the students are aware of the distinction between the model and the situation, and learn to assess whether the model is more or less adequate against the backdrop of contextual factors such as the goals of the modeler (Greer, 1997). In the alternative view, a model is the result of an organizing activity. It is in the process of structuring a problem situation that the model emerges. Within this perspective, the model and the situation modeled co-evolve and are mutually constituted in the course of modeling activity. Thus, when we characterize modeling as a process of mathematization by which the situation is being structured in terms of mathematical relationships, the distinction between the model and the situation modeled dissolves.

The idea of modeling as organizing is rooted in Freudenthal's notion of mathematics as a human activity. For Freudenthal (1973, 1991), mathematical activity mainly consists of organizing, or mathematizing. Over the years, Freudenthal's ideas have been worked out in the domain-specific instruction theory for "realistic mathematics education" (RME) (Gravemeijer, 1994; Treffers, 1987). In RME, modeling as an activity is further elaborated in a didactical sense. The idea is that informal ways of modeling emerge when students are reorganizing their activity while solving contextual problems. Later, these ways of modeling may serve as a basis for developing more formal mathematical knowledge. At first a model is constituted as a context-specific *model of* acting in a situation, then the model is generalized over situations. The model changes character, it becomes an entity of its own, and as such it can function as a *model for* more formal mathematical reasoning.

The shift from model-of to model-for concurs with a shift in the students' thinking, from thinking about the modeled situation, to a focus on mathematical relations. In doing so, the students gradually build a framework of mathematical relations. Then, the model begins to derive its meaning from this emerging framework of mathematical relations, and the model becomes more important as a base for mathematical reasoning than as a way to symbolize mathematical activity in a particular setting. In this sense, the role of the model gradually changes as it takes on a life of its own. As a consequence the model can become a referential base for more formal mathematical reasoning.

Within the above developmental progression, we can discern four types of activity, which we may denote as levels even though they do not involve a strictly ordered hierarchy (Gravemeijer, 1994; Gravemeijer, Cobb, Bowers, and Whithenack, 2000, Gravemeijer, 1999):

- activity in the *task setting*, in which interpretations and (situation-specific) solutions depend on understanding of how to act in that setting
- *referential* activity, in which models-of refer to activity in the setting described in instructional tasks
- *general* activity, in which an orientation on mathematical relations and strategies make it possible to act and reason independently of situation-specific imagery
- more *formal* mathematical reasoning, which is no longer dependent on the support of models-for mathematical activity

To conclude this discussion of the emergent models heuristic, we want to make two observations. Firstly, we want to emphasize that in RME, formal mathematics is not seen as something "out there" that the student has to connect with. Instead, formal mathematics is seen as something that grows out of the students' activity. The students are not expected to experience formal mathematics differently from informal mathematics. The distinction informal-formal is primarily a relative distinction that is relevant from an instructional designer's perspective. We expect the students to expand their mathematical reality while participating in instructional activities. Then the goal of an instructional sequence may be called formal in that this goal refers to some mathematical reality that a certain group of students still has to construe. We will speak of "more formal mathematical reasoning", in the context of that instructional sequence, when the students build on arguments that are located in that new mathematical reality.

Secondly, we want to note that model in the model-of/model-for metaphor has to be understood as a global overarching concept. In the concrete elaboration in the sequence, the overarching model takes on various manifestations. The idea is that the students will use various symbolizations as tools, and that each activity with a newer symbolization or tool is experienced as a natural extension of the activity with the earlier symbolization. Consequently, the formal mathematical symbols that will eventually be used, will be rooted in concrete activities of the students. The dynamic character of this process justifies the term emergent models. However, the meaning of the label is broader. It refers both to the process by which models emerge, and to the process by which these models support the emergence of more formal mathematical knowledge.

In summary, we may observe that there are three interrelated processes. Firstly, there is the overarching model, which first emerges as a model of informal mathematical activity, and then gradually develops into a model for more formal mathematical reasoning. Secondly, the model-of/model-for transition involves the constitution of some new mathematical reality--which can be called formal in relation to the original starting points of the students. Thirdly, in the concrete elaboration of the instructional sequence, there is not one model, but the model is actually shaped as a series of symbolizations. In the following section of the paper, we want to take these three processes as a point of departure to further elucidate the emergent models heuristic.

DATA ANALYSIS AS AN EXAMPLE

We will take a teaching experiment on data analysis carried out by Cobb, Gravemeijer, McClain and Konold in a 7th-grade classroom in Nashville (USA) as an example. The general goal of this sequence is that the students come to view data sets as distributions, of which one can discern characteristics that are relevant, when resolving issues concerning the situation where the measurements have been taken. The students develop this notion by participating in the activity of exploratory data analysis, and by reflecting on that activity. The starting points for the instructional activities are realistic problems that provide a reason for analyzing data. Having a reason is essential in our view, for this reason guides the very process of data analysis.

Looking at the three processes mentioned above, we may conclude that the instructional designer will have to consider the choice of “the model”, the new mathematical reality the students are to construe, and the series of symbolizations that will instantiate the model in the concrete instructional activities. We will start the discussion of the instructional sequence that we used in the teaching experiment on data analysis by considering the goals of the instructional sequence. Firstly, we ask ourselves: What constitutes the new mathematical reality we want the students to construe, and what are the mathematical relations involved? Secondly, we ask ourselves: What is the overarching model, and what do the underlying inscriptions consist of?

New mathematical reality. In answer to the first question, we can report that what is to be construed as new mathematical reality by the students may be denoted as distribution-as-an-entity. We want the students to come to view data sets as entities that are distributed within a space of possible values rather than a plurality of values (Hancock, Kaput, & Goldsmith, 1992; Wilensky, 1997). Another argument for choosing distribution as a central statistical idea is that in conventional statistics courses, statistical measures like mean, mode, median, spread, quartiles, (relative) frequency, regression, and correlation are taught as a set of independent definitions. In contrast, these statistical measures come to the fore as characteristics of distributions in the conjectured learning trajectory. Likewise, conventional representations like histogram, and box plot come to the fore as means to characterize distributions.

To clarify what we mean by distribution as a learning goal, we may start by observing that for us, distribution is intimately connected with the idea of “shape”. Like when we colloquially speak of the “bell shape” of a normal distribution. Although, as indicated at the beginning of this paper, the imagery of a bell-shaped distribution, entails more than a mere figural inscription. In relation to this, we may ask ourselves, what does the curve of Figure 1 represent?

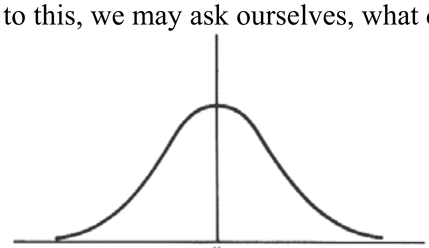

Figure 1. Bell Curve

At first sight, the height of a point of the curve seems to signify the number of cases that have a measure equal to the corresponding measurement value. However, such a point does not have any width, thus we are working with an endless precision. And there will probably not be a single case with exactly that measure. Consequently, the graph can be seen as an idealization, or as the limit of a series of (relative frequency) histograms, with their interval widths approaching zero. We believe that approaching distribution from this angle would be far to demanding for 7th-grade students. However, another way to think about such a graph is as a density function. We believe that this offers a way into a qualitative understanding of distribution. In this conceptualization, the height of a point on the graph does not signify a number of cases, but the density of data points around that value. From this perspective, distribution can be thought of in terms of shape and density. Shape and density in turn can be seen as means to organize collections of data points in a space of possible data values. In relation to this, we may mark that Hancock, Kaput, and Goldsmith (1992) found that students tend to see data as attributes of individuals, which implies that students will have to reorganize their thinking to be able to see data as possible values of a variable.

In summary, we may conclude that important mathematical relations concern: shape, density, variable, and data points in a space of possible values. These are the relations that will have to constitute the network of mathematical relations that will be instrumental in the transition from the level of referential activity to the level of general activity. This implies that in order to support this transition, those mathematical relations have to become a topic for discussion in the classroom. To this we may add the issue of multiplicative reasoning, firstly since that is implied by the notions of shape and density, and secondly since it will come to play when comparing data sets of unequal size.

Emergent model and symbolizations. In answer to the second question, we may describe the overarching model as “a graphical representation of the shape of a distribution”. Given the tight connection between distribution and shape, it seems self-explanatory that the overarching model is tied to shape. With “a graphical representation of the shape of a distribution” we, of course, do not mean just the figural inscription itself, but also what we hope it will signify for the student. The most common graph of the shape of a distribution is the graph of a density function we discussed earlier. However, we may also think of histograms, box plots, or stem-and-leaf diagrams. To find the graph to start the sequence with, we have looked for a graph that would most closely match an intuitive image of a set of measures. It should be a graph that the students could, in principle, invent themselves. In relation to this, the notion of a scale line came to mind. Especially measures of a linear type, like “length”, and “time” are often represented by scale lines. These considerations let to the choice of a graph that consists of value bars, which each value bar signifying a single measure (Figure 2). Next we looked for a type of graph that might function as a transition stage between the magnitude-value-bar graph and the graph of a density function, such as a dot plot (Figure 3).

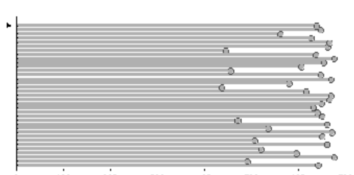

Figure 2. Value-Bar Graph

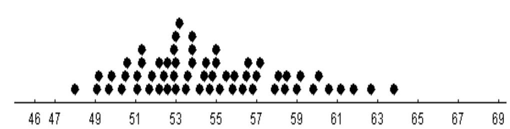

Figure 3. Dot Plot

Within a dot plot, the density of the data points in a given region translates itself in the way the dots are stacked. Consequently, the height of the stacked dots at a given position can be interpreted as a measure for the density at that position. In this sense, the visual shape of the dot plot can be seen as a qualitative precursor to the graph of a density function. On the other hand, the dot plot can be seen as a more condense form of a line plot that leaves out the value bars and only keeps the end points.

The aforementioned graphs are embedded in software (mini)tools that can be used for exploratory data analysis on an elementary level. Point of departure, is a bottom-up approach in which the minitools are perceived by the students as tools that are compatible with their conception of analyzing data, and are experienced as sensible tools in that regard. So for the students the primary function of the minitools is to help them structure and describe data sets in order to make a decision or judgment.

The first minitool displays individual data values as value bars as shown in Figure 2. With this minitool, one can display two or three data sets. The minitool has various options, like sorting and partitioning data, which the students can use to decribe and compare data sets. The second minitool displays individual data values as dots in a dot plot. This minitool can display two data sets at a time, and various tool options are available, to help the student structure the distribution of data points in a dot plot. These options include: making your own groups, making four equal groups, making groups of a certain size, and making equal intervals.

To conclude, we want to use some example from the Nashville teaching experiment to explicate what the model-of /model-for transition entail in this sequence. One of the first tasks of the sequence concerns the comparison the life spans of to brands of batteries, Though Cell and Always Ready. The students do not actually measure life spans, instead the teacher and students talk through the process of data creation. First they discuss what important characteristics of batteries are. Then, when they have decided on life span as an important characteristic, they discuss how the life span of a battery could be measured. And finally the results of such a measurement for ten Though Cell batteries and ten Always Ready batteries are presented as value bars in the first minitool (Figure 4).

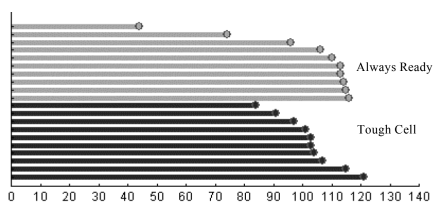

Figure 4. The Life Span of Two Brands of Batteries

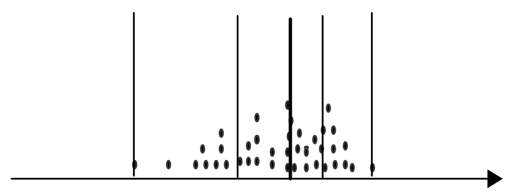

Figure 5. Dot Plot Structured into Four Equal Groups

The students introduced the term “consistency” to argue that they “would rather have a consistent battery (...) than one that you just have to try to guess”. We may interpret this argument as refering to the shape of the distribution, which is visible in the way the endpoints of the value bars are distributed in regard to the axis. In relation to this, we can speak of *a graphical representation of the distribution* as a *model of* a set of measures. Eventually the students used the four-equal-groups display of the second minitool to reason about shape and density (Figure 5).

The distance between two vertical bars that mark a quartile can were interpreted as indicating how much the data are “bunched up”. Moreover, the median started to function as an indicator of “where the hill is”. Finally, the students started to treat distributions as entities. In this regard, we may describe the four-equal groups display as *a graphical representation of the distribution* that started to function as a *model of* a model for reasoning about distributions.

ACKNOWLEDGEMENT

The analyses reported in this paper was supported by the National Science Foundation under grant No. REC 9604982 and by the Office of Educational Research and Improvement under grant No. R305A60007. The opinions expressed do not necessarily reflect the view of either the Foundation or OERI.

REFERENCES

Bereiter, C. (1985). Towards a solution of the learning paradox. *Review of Educational Research, 55*, 201-226.

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

Freudenthal, H. (1991). *Revisiting mathematics education.* Dordrecht: Kluwer Academic Publishers.

Gravemeijer, K.P.E. (1994). *Developing realistic mathematics education*. Utrecht: Cdß Press.

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

Gravemeijer, K., Cobb, P., Bowers, J., & Whitenack, J., (2000). Symbolizing, Modeling, and Instructional Design. In P. Cobb, E. Yackel and K. McClain (Eds.), *Communicating and symbolizing in mathematics: Perspectives on discourse, tools, and instructional design.* Mahwah, NJ: Lawrence Erlbaum Associates, 225 - 273.

Greer, B. (1997). Modelling reality in the mathematics classroom: The case of word problems. *Learning and Instruction, 7*, 293-307.

Hancock, C., Kaput, J.J., & Goldsmith, L.T. (1992). Authentic inquiry with data: Critical barriers to classroom implementation. *Educational Psychologist, 27*, 337-364.

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

Nemirovsky, R.C. (1994). On ways of symbolizing: The case of Laura and the velocity sign. *Journal of Mathematical Behavior, 13*, 389-422.

Treffers, A. (1987). *Three dimensions: A model of goal and theory description in mathematics instruction—The Wiskobas Project*. Dordrecht, The Netherlands: Reidel.

Wilensky, U. (1997). What is normal anyway? Therapy for epistemological anxiety. *Educational Studies on Mathematics, 33*, 171-202.