# An instructional design perspective on data-modelling[1] for learning statistics and modelling

**Jonas Bergman Ärlebäck, Per Blomberg, Per Nilsson**
Linköping University, Sweden
Linnaeus University, Sweden
Örebro University and Linnaeus University, Sweden

*This theoretical and methodological paper discusses the emerging theoretical framing and methodological considerations in our efforts to develop a theoretical approach supporting instructional design for teaching and learning statistics and mathematical modelling. From an instructional design point of view aligned with the goals in governing curricula documents and real classroom constraints, we argue for the integration of the models and modelling perspective on teaching and learning mathematics with a data-modelling approach to facilitate students' learning statistics and mathematical modelling. An application of the framework is given and future research discussed.*

## Background

Statistics has been described as the science of models and modelling through which we make sense of the world using theory-driven interpretations of data (Shaughnessy, 1992). Indeed, the essence of statistical thinking is argued to be centred around developing, testing, interpreting, and revising models in order to understand our world and the diverse phenomena in it (Horvath & Lehrer, 1998). With this view of statistics, there are many parallels with the general view and on-going discussion on the use and role of mathematical modelling in the teaching and learning of mathematics. Both statistics and mathematical modelling have been put forward as increasingly more important for students to learn in order to cope with and be productive in their everyday and professional lives (in the case of modelling see Blum, Galbraith and Niss (2007); and, in the case of statistics see for example Gal (2002)). This suggests to us, as also hinted by English and Sriraman (2010), that statistics potentially provides a rich and productive venue for learning modelling on the one hand, but on the other, it also suggests that statistics may advantageously be learned through modelling. This interrelating connection between statistics and modelling (in a more general sense) constitutes an important strand in our present thinking and on-going work presented in this paper.

Following the definition used by Radford (2008) and Wedege (2010), we in this paper seek to develop a focused *theoretical approach*, that is, a framework "based on a system of basic theoretical principles combined with a methodology" (Wedege, 2010, p. 65). The sought theoretical approach should guide and support teachers and researchers in designing instructional sequences for the teaching and learning of statistics and mathematical modelling in everyday mathematics classrooms focusing on pre-defined learning objectives as prescribed in governing curricula documents. In addition to draw on the suggested benefitting symbiotic effects from intertwiningly learning statistics and modelling, we find it equally important for the theoretical approach we seek to develop to seriously acknowledge and respect the constraints affecting everyday teaching practices in schools. Two palpable constrains on teachers' practices are time constraints and content constrains, with the latter regulated in governing curricula documents specifying what it is that students should learn in the different grades and courses. Especially when the content and concepts becomes more advanced and abstract as the students become older and progress in their mathematics courses, the amount of time the teacher has available to spend on a given topic becomes considerably more limited. Ideally therefore, we want our theoretical approach to explicitly respect these constraints in that it should facilitate instructional design of productive learning situations for students that are focused from the point view of content as well as efficient.

To this end, we want to use the fundamental ideas underpinning the data-modelling approach for learning statistics described by Lehrer and colleagues (Horvath & Lehrer, 1998; Lehrer & Schauble, 2000; 2004) and to put these in the larger framework of model-development sequences (Lesh, Cramer, Doerr, Post, & Zawojewski, 2003) within the models and modelling perspective on teaching and learning (Lesh & Doerr, 2003b). By adapting this integrating networking strategy (e.g. Prediger, Bikner-Ahsbahs, & Arzarello, 2008), we argue that the theoretical approach that emerges (1) provides a potentially highly productive symbiotic approach for learning statistics and mathematical modelling; (2) integrates prescribed curricula learning objective in a natural and time efficient way at all educational levels; and (3) results in instructional designs that provide promising learning possibilities for students and teachers as well as rich and productive research settings to further both the instructional designs themselves as well as the field of mathematics education research.

## A modelling framework for instructional design of statistics learning

We now proceed to briefly discuss a models and modelling perspective on teaching and learning mathematics, before focusing on the ideas of Lehrer and colleagues' data-modelling approach. We end this section by presenting what should be considered the contribution of this paper: our emerging theoretical approach for instructional design of symbiotic teaching and learning of statistics and mathematical modelling.

**A models and modelling perspective of teaching and learning**

In a models and modelling perspective on teaching and learning (Lesh & Doerr, 2003b), a model is an externally representable conceptual system (consisting of objects, operations, relations, and interaction-governing rules) used to describe, explain, predict, or understand some other system (Lesh & Doerr, 2003a). In the case of mathematics, graphs, tables, algebraic expressions, computer animations, enacted actions, and spoke and written language are all examples of external representational systems.

From a models and modelling perspective learning is model development (ibid). Hence, we consider learning statistics to be for students to develop models for statistical reasoning. Here, the verb develop stresses, and refers to, the dynamic aspect of this process, where students' models repeatedly are developed, modified, extended and revised through "multiple cycles of interpretations, descriptions, conjectures, explanations and justifications that are iteratively refined an reconstructed by the learner" (Doerr & English, 2003, p. 112).

To support and facilitate students' development of models, sequences of structurally similar activities called model-development sequences have been introduced as tools for structuring teaching (Doerr & English, 2003; Lesh et al., 2003). A model-development sequence always begins with a model-eliciting activity with the purpose of putting the students in a meaningful situation where they are confronted with a need to develop or recall a model (c.f. (Freudenthal, 1983)). Other purposes of the model-eliciting activity are to make the students' previous experiences and models visible (to themselves, their peers and teachers) as well as explicit articulated objects that can be reflected upon and discussed.

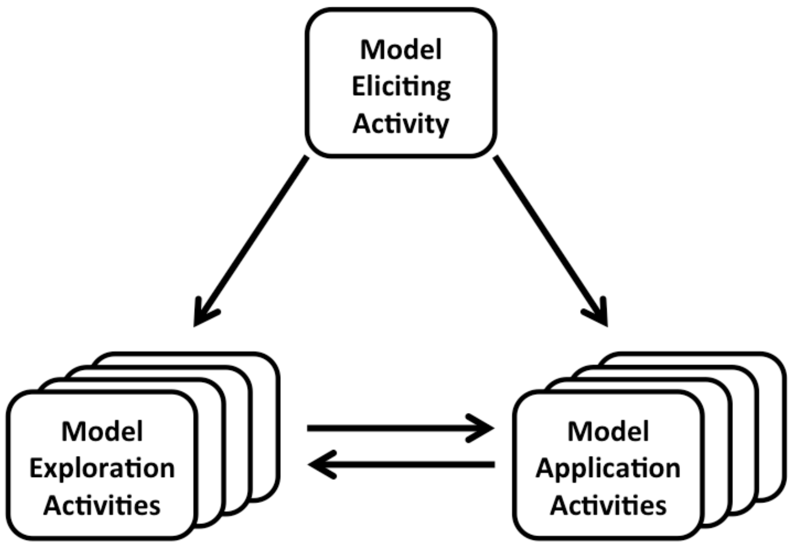


Figure 1. The general structure of a model-development sequence

In a model-development sequence, the model-eliciting activity is followed by one or more structurally related model-explorations activities and model-application activities (see Figure 1). The model-exploration activities focus on exposing and exploring the underlying mathematical structure of the elicited model. An important part here is for the students to work with different representations of the model as well as to develop an understanding for how to use different representations productively. The model-application activities allow the students to apply their models in other situations and contexts. Throughout working with the different activities within a model-development sequence students are constantly subjecting their (evolving) models to testing, modifying, revising and development.

**Data-modelling**

Lehrer and Schauble (2000) consider data-modelling to be a nested approach to students' classroom inquires with inherent processes facilitating development of students' ideas and models of big ideas and key concept in statistic. The idea is to put students in a, for them, realistic and meaningful "data-modelling context" (p. 636) where the starting point is student generated questions. Building on their understandings of the situation and problem at hand, the students develop and investigate feasible solutions (models) to their questions by engaging in a cyclic-like inquiry process illustrated in Figure 2. In other words, based on their contextual situated questions, the students have to identify and decide on what attributes are influencing the situation and are relevant for answering their questions; to collect data for these attributes (or select from given data); to choose ways of representing, organising, and displaying the data; and finally, to analyse the data and try to answer their questions by making inferences, often of an informal nature (English & Sriraman, 2010).

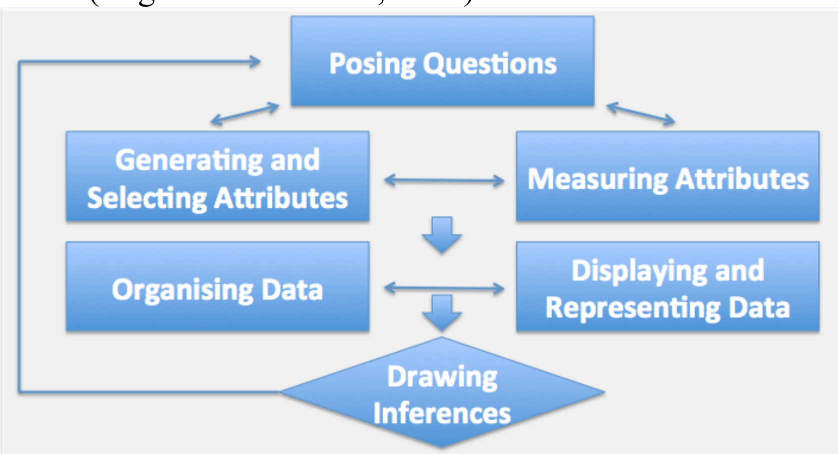


Figure 2. Components of data-modelling (Lehrer & Schauble, 2004) as adapted and presented by English and Sriraman (2010, p. 280).

Used with its open onset, drawing on "students-as-designer contexts for their fruitfulness in provoking and sustaining student engagement with data" (Lehrer & Romberg, 1996, p. 71), data-modelling in classroom settings normally spans over multiple lessons. From a learning point of view a data-modelling approach facilitates students' development of central and big ideas in statistics as well as giving them first hand experiences and a holistic view of the different components and aspects involved in statistical analysis (Lehrer & Schauble, 2000). From a research point of view on the other hand, studies involving implementing a data-modelling approach in everyday classrooms have proved to provide rich research sites for more focused studies, such as on student's development of understanding chance and uncertainty (Horvath & Lehrer, 1998) and students' development of understanding variation (Lehrer & Schauble, 2004).

**A theoretical approach for instructional design for the teaching and learning of statistics and modelling – an extended data-modelling approach**

We now turn to propose an expansion of the data-modelling approach discussed above. We do this by situating data-modelling as a model-exploration activity within a model-development sequence focusing on a particular learning objective by adapting an integrating networking strategy (e.g. Prediger et al., 2008). In doing this, we gain a structured way to focus on both learning and exploring specific statistics curricular content and mathematical modelling within a confined and limited number of activities.

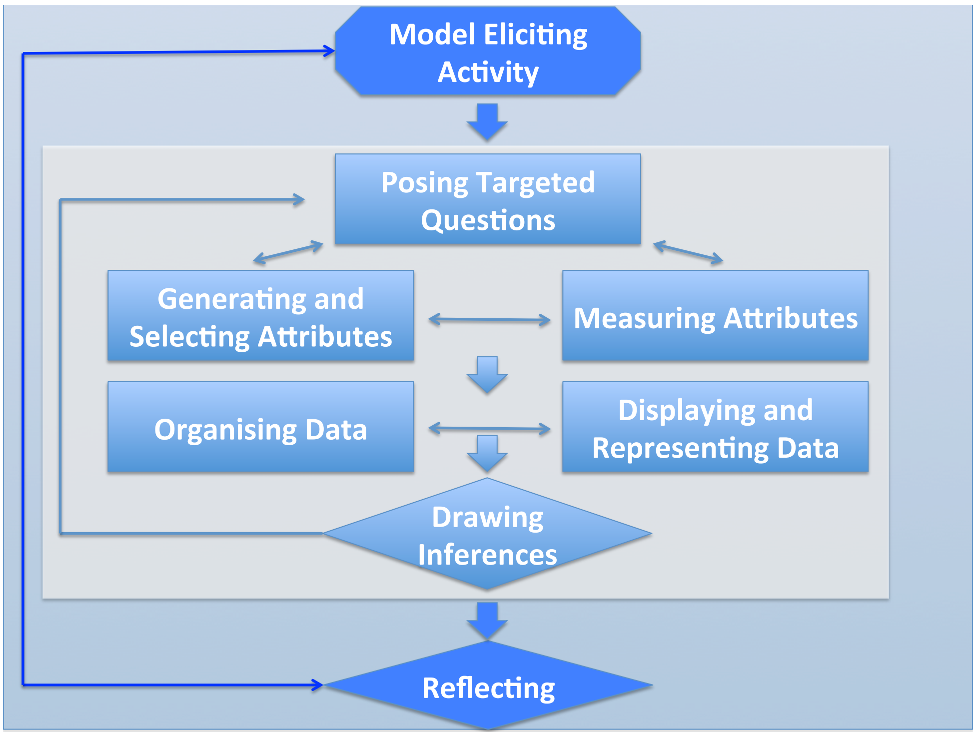


Figure 3. Components of the extended data-modelling approach

Situating data-modelling as a model-exploration activity within a model-development sequence means that in our extended data-modelling approach, an initiating model-eliciting activity must be added, that elicits the students' ideas and models with respect to the given learning objective (see Figure 3). Besides providing a situation were the students need to develop and express their thinking (models) about the targeted learning objective, this initiating modelling-eliciting activity also helps focuses the students in posing questions when engaged in the model-exploration activity (that is, in the original data-modelling approach). In other words, the model-eliciting activity funnel the students' thinking towards the particular learning objective in question, in that their experiences from the model-eliciting activity narrow down the viable questions for the sequential model-exploration activity and make the posing of questions more targeted. This is a shift from the original data-modelling approach represented by Figure 2, where the students' interests and own preferences to a large extent determine what questions to investigate and how to go about in trying to answer their posed questions, resulting in a more unpredictable activity in terms learning outcomes.

In our emerging theoretical approach (Figure 3) the initiating model-eliciting activity is followed by a traditional data-modelling activity (c.f. Figure 2) where students' content-wise focused elicited models are tested, challenged, adjusted, revised and developed as they investigate and analyse real data. It is the combination of these two first activities that lay the foundation for the students' symbiotic learning of statistics and mathematical modelling. In the theoretical aproach in Figure 3 there is a dialectic relationship between the goal to learn a specified statistical content and the goal to learn mathematical modelling. During the model-eliciting activity and the first model-exploration activity, the statistics' leaning objective is in the foreground whereas learning mathematical modelling is in background; one can say that the students learn statistical content using a mathematical modelling methodology, that mathematical modelling is the method and vehicle supporting the students' learning of statistics.

In addition to add an initial model-eliciting activity to the data-modelling approach, we also explicitly add an activity ending the model-exploration activity connecting back to the students' initially elicited models. By explicitly connecting back to the point of departure the students are offered an opportunity to reflect on both their own development as well as on the working methods and processes leading to it. This reflecting activity also reverses the emphasis in that the mathematical modelling process comes to the foreground and the focused statistical content in the background. To explicitly reflect and connect back to the starting point, or applied real world situation and original questions, is a key characteristics of all conceptualisations of modelling (Blum et al., 2007).

The extended data-modelling approach illustrated in Figure 3, can be seen as an extended model-eliciting activity in the original conceptualization of model development sequences by Lesh et al. (2003). That is, the models the students elicit and develop through engaging in this extended model-eliciting activity can then be further developed, explored and applied though carefully sequenced model-exploration activities and model-application activates.

By virtue of design, the theoretical approach in Figure 3 offers and allows an opportunity to design instruction that simultaneously integrates teaching and learning of focused content learning alongside an additional more general learning object or more general abilities such as problem solving, getting a holistic view of mathematical modelling, and for the students to experience the uses, benefits and power of mathematics.

## An example an application of the theoretical approach

We now continue to give an example illustrating our emerging theoretical approach in action. Due to the limited space available we will only focus on some selected design decisions of the actually designs informed by the theoretical approach developed in this paper, and not present any data, analysis or results. The example is from on-going work investigating how students develop their inferential reasoning.

### Developing students' informal statistical inference reasoning

Recent research in statistics education has focused on different aspects of the role and function of students' informal statistical reasoning in their everyday lives and for learning formal statistical inference (Biehler & Pratt, 2012; Makar & Rubin, 2009). In this context, the theoretical approach presented above was used to design a teaching experiment to support students developing their informal statistical inference reasoning. Of more specific interest was students' informal statistical inferences drawn from samples to larger populations (statistics) and what attribute students consider to be relevant for making such an inference legitimate (modelling).

The teaching experiment was situated in the context around the growing concern of obesity and health issues due to peoples sedentary and immobility, and the initial model-eliciting activity the student worked on in small groups was:

> Research suggest that young people should walk at least 10 000 steps a day to stay physically fit. What is the probability that a person of your age completes more that 10 000 steps a day? Your task is to write a letter to the teacher where you present assumptions, reasoning and calculation you have done to solve the task.

The students drew on their previous knowledge and experiences to identify relevant attributes for setting up a model as well as to estimate numerical parameters in order to come up with a probability. By engaging in this model-eliciting activity the students' ideas (models) became explicit objects of thoughts, and open for discussion. The activity also allows the students the opportunity to make visible what mathematical and statistical knowledge and constructs the considered applicable and relevant for the situation.

In the data-modelling task that followed, the students used pedometers to collect empirical data of their own physical activities. Based on the data collected by the members of their group, the students then revisited the question from the initial model-eliciting activity. The variation of number of steps in the collected data supported the students in verifying the relevance of their prior identified attributes or in identifying new ones.

The reflecting activity that ended the teaching experiment was two-folded; partly it consisted of the students writing a report, and partly is consisted of a teacher-lead whole class discussion. Both forms of reflections focused on the informal statistical inferences drawn by the students as well as the modelling processes leading up to their conclusions. In addition, during the whole class discussion, all the students' samples from their different groups were aggregated, which was brought issues of sample size and numbers of samples on the table. Here, these ideas about the role of sample size and the numbers of sample in making inferences suggest directions for further model development and sequential model-exploration activities as well as model-application activities. This latter point however, is subject for future research.

## Conclusions, implications and future research

In this paper we have presented an emerging theoretical approach for supporting instructional design for teaching and learning statistics and mathematical modelling symbiotically. We drew on fundamental ideas underpinning the data-modelling approach for learning statistics described by Lehrer and colleagues (Horvath & Lehrer, 1998; Lehrer & Schauble, 2000; 2004) and integrated (c.f. Prediger et al., 2008),. these into the larger framework of model-development sequences (Lesh et al., 2003) within the models and modelling perspective on teaching and learning (Lesh & Doerr, 2003b). In addition, we sought to integrate sensitivity to constraints imposed on everyday teaching practices from the limited amount of time available and prescribed student learning outcomes in governing curricula documents.

By drawing on well establish research methodologies and approaches, we argue that our emerging theoretical approach in a productive way pulls together fundamental theoretical ideas and respects important real classroom constraints resulting in a supportive model for thinking about, designing and developing teaching and learning of statistics and mathematical modelling symbiotically. In addition, since the definition of models within the models and model perspective is broad and general, the emphasise on prescribed curricula learning objectives facilitate a natural as well as time efficient and flexible approach for instructional design that arguably could be applied at all educational levels.

We acknowledge that that there are challenges in adapting and implementing teaching based on the emerging theoretical approach presented in this paper. However, as well as potentially resulting in instructional designs that provide promising learning possibilities for students and teachers, we plan to use a design based research paradigm (Cobb, Confrey, diSessa, Lehrer, & Schauble, 2003) to continue to develop our theoretical approach. In this context our work presented here provides us with rich and productive research settings to further both the instructional designs themselves as well as the field of mathematics education research.

A line of research that we find specifically fruitful and appealing, is to further draw on the related literature discussing the different components of our theoretical approach combined with empirical investigations to develop a set of design principles specific for teaching and leaning of statistics and mathematical modelling symbiotically. In the case of model-eliciting there are six well established design principles (Lesh, Hoover, Hole, Kelly, & Post, 2000) which provide a natural starting point for this endeavour. In addition, we intend to build and continue to further the work on the emerging design principles for model-exploration activities and model-application activities initiated in Ärlebäck and Doerr (submitted).

## Notes

1. We acknowledge that much of the referred literature in this paper is American and hence uses the America spelling *modeling*. Also note that at times the notions *data modeling* and *data-modeling* is used interchangeably in the literature. We however, for consistency reasons, consequently use the English spelling *modelling* and the notation *data-modelling*.

## References

Ärlebäck, J. B., & Doerr, H. M. (submitted). Moving beyond a single modelling activity. Book chapter submitted to the Proceedings of ICTMA 16.

Biehler, R., & Pratt, D. (2012). Research on the reasoning, teaching and learning of probability and uncertainty. *Zdm*, *44*(7), 819–823. doi:10.1007/s11858-012-0468-0

Blum, W., Galbraith, P., & Niss, M. (2007). Introduction. In W. Blum, P. Galbraith, H.-W. Henn, & M. Niss (Eds.), *Modelling and applications in mathematics education. The 14th ICMI study.* (pp. 3–32). New York: Springer.

Cobb, P., Confrey, J., diSessa, A. A., Lehrer, R., & Schauble, L. (2003). Design Experiments in Educational Research. *Educational Researcher*, *32*(1), 9–13.

Doerr, H. M., & English, L. D. (2003). A modeling perspective on students' mathematical reasoning about data. *Journal for Research in Mathematics Education*, 110–136.

English, L. D., & Sriraman, B. (2010). Problem solving for the 21st century. In B.

Sriraman & L. D. English (Eds.), *Theories of mathematics education* (pp. 263–290). Springer.

Freudenthal, H. (1983). *Didactical phenomenology of mathematical structures*. Boston: Kluwer.

Gal, I. (2002). Adults' statistical literacy: Meanings, components, responsibilities. *International Statistical Review*, *70*(1), 1–25.

Horvath, J. K., & Lehrer, R. (1998). A model-based perspective on the development of children's understanding of chance and uncertainty. In S. P. Lajoie (Ed.), *Reflections on statistics: Learning, teaching, and assessment in grades K-12* (pp. 121–148). Mahwah, NJ: Lawrence Erlbaum Associates Publisher.

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

Lehrer, R., & Schauble, L. (2000). Modeling in mathematics and science. In R. Glaser (Ed.), *Advances in instructional psychology. Volume 5* (Vol. 5, pp. 101–159). Mahwah, NJ: Lawrence Erlbaum Associates Publisher.

Lehrer, R., & Schauble, L. (2004). Modeling Natural Variation Through Distribution. *American Educational Research Journal*, *41*(3), 635–679.

Lesh, R. A., & Doerr, H. M. (2003a). Foundations of a Models and Modeling Perspective on Mathematics Teaching, Learning, and Problem Solving. In R. A. Lesh & H. M. Doerr (Eds.), *Beyond constructivism: Models and modeling perspectives on mathematics problem solving, learning, and teaching* (pp. 3–33). Mahwah, NJ: Lawrence Erlbaum Associates.

Lesh, R. A., & Doerr, H. M. (Eds.). (2003b). *Beyond constructivism: Models and modeling perspectives on mathematics problem solving, learning, and teaching*. Mahwah, NJ: Lawrence Erlbaum Associates.

Lesh, R. A., Cramer, K., Doerr, H. M., Post, T., & Zawojewski, J. S. (2003). Model development sequences. In R. A. Lesh & H. M. Doerr (Eds.), *Beyond constructivism: Models and modeling perspectives on mathematics problem solving, learning, and teaching* (pp. 35–58). Mahwah, NJ: Lawrence Erlbaum Associates.

Lesh, R. A., Hoover, M., Hole, B., Kelly, A. E., & Post, T. (2000). Principles for developing thought-revealing activities for students and teachers. In A. E. Kelly & R. A. Lesh (Eds.), *Handbook of research design in mathematics and science education* (pp. 591–645). Mahwah, NJ; London: L. Erlbaum.

Makar, K., & Rubin, A. (2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal*, *8*(1), 82–105.

Prediger, S., Bikner-Ahsbahs, A., & Arzarello, F. (2008). Networking strategies and methods for connecting theoretical approaches: first steps towards a conceptual framework. *Zdm*, *40*(2), 165–178.

Radford, L. (2008). Connecting theories in mathematics education: challenges and possibilities. *Zdm*, *40*(2), 317–327.

Shaughnessy, M. J. (1992). Research in probability and statistics: Reflections and directions. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 465–494). New York: Macmillan.

Wedege, T. (2010). Connecting theories in mathematics education: From bricolage og professionalism. *Nordic Studies in Mathematics Education*, *15*(1), 59–78.