DATA MODELLING IN THE BEGINNING SCHOOL YEARS

LYN ENGLISH

Queensland University of Technology

l.english@qut.edu.au

This paper argues for a renewed focus on statistical reasoning in the beginning school years, with opportunities for children to engage in data modelling. Some of the core components of data modelling are addressed. A selection of results from the first data modelling activity implemented during the second year (2010; second grade) of a current longitudinal study are reported. Data modelling involves investigations of meaningful phenomena, deciding what is worthy of attention (identifying complex attributes), and then progressing to organising, structuring, visualising, and representing data. Reported here are children's abilities to identify diverse and complex attributes, sort and classify data in different ways, and create and interpret models to represent their data.

Introduction

The need to understand and apply statistical reasoning is paramount across all walks of life. One only has to peruse the daily newspapers to see the variety of graphs, tables, diagrams, and other data representations that need to be interpreted. With unprecedented access to a vast array of numerical information, we can engage increasingly in democratic discourse and public decision making—that is, provided we have an appropriate understanding of statistics and statistical reasoning.

Young children are immersed in our data-driven society, with early access to computer technology and daily exposure to the mass media. With the rate of data proliferation has come increased calls for advancing children's statistical reasoning abilities, commencing with the earliest years of schooling (e.g., Franklin & Garfield, 2006; Langrall, Mooney, Nisbet, & Jones, 2008; Lehrer & Schauble, 2005; National Council of Teachers of Mathematics [NCTM], 2006; Shaughnessy, 2010).

We need to rethink the nature of young children's statistical experiences and consider how we can best develop the important mathematical and scientific ideas and processes that underlie statistical reasoning (Franklin & Garfield, 2006; Langrall et al., 2008; Leavy, 2007; Watson, 2006). There has been limited research, however, on developing young children's statistical reasoning. One approach in the beginning school years is through data modelling (English, 2010; Lehrer & Romberg, 1996; Lehrer & Schauble, 2007; Lehrer & Schauble, 2000).

In this paper, I first argue for the need to review young children's statistical experiences, with a focus on data modelling. I then address some findings from a data modelling activity implemented in second-grade classrooms during the second year of a three-year longitudinal study. In reporting some findings, I consider children's:

1. Recognition of diverse and complex attributes;
2. Identification of ways to sort and classify their data;
3. Models created in representing their data and their interpretations of their models.

Modelling

Data modelling is a developmental process, beginning with young children's inquiries and investigations of meaningful phenomena, progressing to identifying various attributes of the phenomena, and then moving towards organising, structuring, visualising, and representing data (Lehrer & Lesh, 2003). As one of the major thematic "big ideas" in mathematics and science (Lehrer & Schauble, 2000, 2005), data modelling should be a fundamental component of early childhood curricula. Limited research exists, however, on such modelling and how it can be fostered in the early school years. Indeed, the majority of the research on mathematical modelling has been confined to the secondary and tertiary levels, with the assumption that primary school children are not able to develop their own models and sense-making systems for dealing with complex situations (Greer, Verschaffel, & Mukhopadhyay, 2007).

Generating and selecting attributes

Early experiences with data modelling involve selecting attributes and classifying items according to these attributes (Lehrer & Schauble, 2000). As Lehrer and Schauble (2007) noted, it is not a simple matter to identify key attributes for addressing a question of interest—the selection of attributes necessitates "seeing things in a particular way, as a collection of qualities, rather than intact objects" (p. 154). Moreover, children have to decide what is worthy of attention (Hanner, James, & Rohlfing, 2002). Some aspects need to be selected and others ignored, the latter of which could be salient perceptually or in some other way. Frequently, however, young children are not given experiences in which they need to consider attributes in this way.

Classification activities presented in the early school years usually involve items with clearly defined and discernable features, such as green square shapes, blue triangular shapes etc. (Hanner et al., 2002). It is thus rather easy for children to classify items of this nature. In contrast, problems involving the consideration of more complex and varied attributes, which could define more than one classification group, present a greater challenge to young children.

Structuring and displaying data

Models are typically conveyed as systems of representation, where structuring and displaying data are fundamental—"Structure is constructed, not inherent" (Lehrer & Schauble, 2007, p. 157). However, as Lehrer and Schauble indicated, children frequently have difficulties in imposing structure consistently and often overlook important information that needs to be included in their displays or alternatively, they include redundant information. Providing opportunities for young children to structure and display data in ways they choose, and to analyse and assess their representations is important in addressing these early difficulties.

Constructing and displaying their data models involves children in creating their own forms of inscription. By the first grade, children have already developed a wide range of inscriptions, including common drawings, letters, numerical symbols, and other referents. As children create and use their own inscriptions they also develop an "emerging meta-knowledge about inscriptions" (Lehrer & Lesh, 2003), which diSessa, Hammer, Sherin, and Kolpakowski (1991) termed, metarepresentational knowledge. These developing inscriptional capacities provide a basis for children's mathematical activity. Indeed, inscriptions are mediators of mathematical learning and reasoning; they not only communicate children's mathematical thinking but they also shape it (Lehrer & Lesh, 2003; Olson, 1994). Developing a repertoire of inscriptions, appreciating and assessing their qualities and use, and using their inscriptions to explain or persuade others, are essential for data modelling. Yet children are often taught traditional representational systems as isolated topics at a specified point in the curriculum, without really understanding when and why these systems are used.

Role of context

The nature of task design is a key feature of data modelling activities. Stillman, Brown, and Galbraith's (2008) notion of "modelling as vehicle" (p. 143) is applicable here. Such modelling involves choosing contexts in which stimuli for the desired mathematics learning are embedded. Genuine problem situations are used as vehicles for students to construct significant mathematical ideas and processes rather than simply apply previously taught procedures. Furthermore, the mathematics that students engage with in solving such modelling problems usually differs from what they are taught traditionally in the curriculum for their grade level (English, 2003a; 2008; Lesh & Zawojewski, 2007).

When solving data modelling problems children need to appreciate that data are numbers in context (Langrall, Nisbet, Mooney, & Jansem, 2011; Moore, 1990), while at the same time abstract the data from the context (Konold & Higgins, 2003). Research has shown that both the data presentation and the context of a task itself have a bearing on the ways students approach problem solution—presentation and context can create both obstacles and supports in developing students' statistical reasoning, emphasising the need to consider carefully task design (e.g., Cooper & Dunne, 2000).

Methodology

The participants were from an inner-city Australian school. In the first year of the study, three classes of first-grade children (2009, mean age of 6 years 8 months) and their teachers participated. The classes continued into the second year of the study, the focus of this paper (2010, mean age of 7 years 10 months, n=68).

A teaching experiment involving multilevel collaboration (English, 2003b; Lesh & Kelly, 2000) was adopted here. This approach focuses on the developing knowledge of participants at different levels of learning (student, teacher, researcher) and is concerned with the design and implementation of experiences that maximise learning at each level. The teachers' involvement in the research was vital; hence regular professional development meetings were conducted. This paper addresses aspects of the student level.

Activity: Baxter Brown's Shop

The initial activity implemented in the second year of the longitudinal study continued the story context (purposely created) from the first year of activities. The context involved the adventures of Baxter Brown (a "westipoo"—West Highlander X toy poodle). The children requested more stories about Baxter Brown in the second year of the study; hence the Baxter Brown's Shop was created. The Baxter Brown stories, presented as picture books, were read to the children in a whole class setting.

The Baxter Brown's Shop story told of the mischievous supermarket expeditions Baxter took with his owners, Mr and Mrs Brown. The dog created various forms of mayhem as he raced down the supermarket aisles. Following the story, the children were shown a simple table of data indicating the different types of mayhem he had created. As a whole class, the children were to determine whether Baxter Brown was becoming more mischievous as his week in the supermarket progressed. In the second component of the activity, the focus here, it was explained that Baxter Brown was subsequently banned from the supermarket and thus ended up creating his very own shop in his bedroom. The children were given an A3 chart comprising illustrations of 16 supermarket items that displayed diverse attributes (the items were a carton of milk, a frozen pizza, apples, coco pops, pasta, a tin of sliced pineapple, fresh carrots, a packet of cheese, a packet of bread rolls, a packet of biscuits, a container of apple juice, a carton of eggs, a tin of dog food, a packet of fish, packaged chicken, and a packet of Cheezels). Working in groups, the children responded to a number of written questions, including: (a) What are some things you notice about the shopping items? Make a list of these. (b) There are lots of things you have noticed. To help Baxter Brown here, what are some ways in which you might sort and classify your data? (c) Which way do you like best? (d) To make it easier for Baxter Brown, how might you represent your data? What are some different ways? (e) Which way do you like best? Why? (f) Now represent your data on your sheet of paper, and (g) What are some things that your representation tells Baxter Brown? On completion of the activity, the groups presented class reports. Children's responses to questions (a), (b), (f), and (g) are reported here.

Data collection and analysis

In each of the second-grade classrooms, two focus groups (of mixed achievement levels and chosen by the teachers) were videotaped and audiotaped. There were 17 groups of children, five in one class and six in each of the remaining classes. The range of data collected was analysed using iterative refinement cycles for analysis of children's learning (Lesh & Lehrer, 2000), together with constant comparative strategies (Strauss & Corbin, 1990), where data were coded and examined for patterns and trends. For questions (a), (b), and (g), some groups gave more responses than others. For question (f), one group created two models while all other groups created just one model. The analysis of data addressed the total number of responses for each question across all groups.

Selection of findings

Children's identification of attributes

The children's responses to question (a), asking for things that they noticed about the given items, were analysed iteratively with two main categories of responses identified, namely, attributes that were primarily qualitative in nature and those that comprised a quantitative element. Qualitative attributes were favoured over quantitative ones. Of the 41 group responses to this question, 28 were of the former type and 13 of the latter. Examples of the qualitative attributes included "dinner foods, brekky foods, dessert," "foods to cook, foods not to cook," "healthy/non-healthy food," and "cans, bottle, plastic, bags, boxes." Quantitative attributes included, "Only one dog food," "all items are under $10 apart from the chicken," "there are only two tins," and "there are more foods than drinks."

Children's identification of ways to sort and classify data

Half of the 24 responses to question (b) ("What are some of the ways you might sort and classify your data?") referred to complementary categories such as "drinks and food," healthy and not healthy," "cereals and not cereals," and "fruit and non-fruit." Two groups offered three or more categories, such as "Dinner, snack, lunch, and breakfast." A further two groups suggested putting like items together such as "drinks together, cans together, things that are in boxes together." One other group recommended sorting by cost ("highest price to lowest,") and another suggested sorting by shape and size ("big food, small food;" "thickest and thinnest").

Children's model creations

Three main models were evident in the children's recorded representations of their sorting and classifying of data. These were models that comprised (a) lists of items in labelled columns, (b) sets of items enclosed in a curve, and (c) items grouped in two divisions (horizontal or vertical) on the A3 sheet provided.

Items in columns

Five of the 17 groups developed models that displayed two to five labelled columns, with the names of the items recorded one under the other in the respective columns. One of these models used the shop context to define the categories, namely, "fridge" and "cupboard," while another model listed healthy and unhealthy items with prices attached. The model that listed items in three columns displayed interesting categories, namely, "dry," "wet," and "dry and wet." In their class report, one group member explained, "I said how about things that are dry and things that are wet so we decided to put that down, and then I thought about pasta and I said if pasta, if you just had pasta by itself it would be dry but if you cooked it, then it would be wet, so then it would be both, so it depends on if you cooked it or if you kept it raw." Another interesting model that comprised five columns displayed items listed under the categories of shapes, namely, "rectangular prism," "sphere," "cone," "cylinder," and "pyramid." The children explained that there were no items that were a pyramid shape and so that column was left blank.

Items enclosed in a curve

Four groups created a model with item names enclosed within a curve. One of these models comprised a large oval, divided into four, with the divisions labelled “healthy,” “not healthy,” “both,” and “dog food.”

Horizontal/vertical division

The most common models were those that comprised either a vertical or horizontal division of the A3 sheet of paper (either with or without actual dividing lines) and displayed item names or illustrations or both. Nine groups developed models of this nature, with one of the groups creating two models, the other being labelled items within a curve. Interesting classifications were evident in these models, such as “combination [of foods]” and “things that taste good by itself.” The influence of task context was also visible in these models, such as in the last group who drew two tables (one at the top of the paper and the other at the bottom) with illustrations of items lined up across the tables and prices attached. Another group used a shop context of “fridge” and “cupboard.” Their model displayed a drawing of a fridge (labelled “fridge” on the top left-hand corner) with illustrations of cold items stacked on shelves and a cupboard (labelled “cupboard” on the top right-hand corner) with non-cold items illustrated on shelves. Another group incorporated a food pyramid within their model.

Children’s interpretation of their models

Due mainly to lack of time, not all groups provided responses for question (g), where the children were asked what their representation tells Baxter Brown. Of the 18 responses, 13 focused primarily on a contextual interpretation (with a common focus on healthy eating) and a literal reading of their model, that is, there was limited interpretation of the data (cf. Curcio, 2010, first level of graph comprehension). Examples of such responses included, “There’s a healthy aisle and an unhealthy aisle,” “there’s dry food and wet food and both,” and “It tells him what to put in the fridge and what to put in the cupboard; if he doesn’t put it in the right place he might get sick.” Only five responses referred to any relational observations (cf. Curcio’s second level of graph comprehension [“reading between the data”]). Such responses included: “There are more healthy things than unhealthy things, “there is more food than drinks and less drinks than food,” and “Our grid tells Baxter Brown what aisles the shape foods are in. It tells him there are no pyramid shape items like the camera stands-tripods. He would have to go to another shop.”

Discussion and concluding points

This paper has addressed the need for a renewed focus on statistical reasoning in the beginning school years, with a focus on data modelling. In such activities, children interpret and investigate meaningful phenomena involving the identification of diverse and complex attributes, in contrast to the standard attributes they usually experience in early mathematics curricula. Building their data models engages children in organising, structuring, visualising, and representing their data in ways that they, themselves, choose.

Evident in the children’s responses was the influence of task context, which appeared to present both support and obstacles in the children’s reasoning. For example, the children's familiarity with the task context appeared to enable them to identify a diverse range of attributes, some quite unexpected, such as a consideration of food combinations, items for different meal times, and items that were identified as "dry," "wet,' and "dry and wet." On the other hand, the children were very aware of healthy and non-healthy foods (from their health lessons) and this could have overshadowed a possible broader range of attributes being identified. The shop context also influenced some of the groups' model creations where fridges, cupboards, and tables were used to represent data. The impact of task context was also evident in the children's interpretations of their models, where there was a focus on shop aisles and also food storage.

The nature of the task items appeared to have a further impact on the children's identification of attributes and ways to sort and classify the data. Qualitative rather than quantitative features were considered. Nevertheless, the children did identify a wide range of qualitative features despite not making many numerical comparisons. The children's models were not as varied as those created in the first year of the study, (e.g., English, 2010), where the story context focused on Baxter Brown cleaning up his very messy room. Children were given various sets of multiple cut-out items to work with and generated a range of representational models including graphical formats. Perhaps the use of multiple manipulative items might have broadened the children's responses in the present activity. Nevertheless, the children did display an awareness of representational structure in developing their models, with their use of inscriptions enabling clear interpretation and communication of their models. The role of task context in young children's mathematical learning has been highlighted over the years (e.g., Watson, 2006) but this study suggests that further consideration is needed.

Young children need rich opportunities to develop their statistical reasoning abilities; data modelling activities provide such opportunities. However, despite the increased calls for renewed attention to statistical learning in the early school years, research examining young children's statistical developments remains in its infancy.

Acknowledgement

This project reported is supported by a three-year Australian Research Council (ARC) Discovery Grant DP0984178 (2009-2011). Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the author and do not necessarily reflect the views of the ARC. I wish to acknowledge the excellent support provided by the senior research assistant, Jo Macri.

References

Cooper, B. & Dunne, M. (2000). *Assessing children's mathematical knowledge: Social class, sex and problem solving*. Buckingham, UK: Open University Press.

Curcio, F. (2010). *Developing data graph comprehension* (3rd edn.). Reston, VA: National Council of Teachers of Mathematics.

diSessa, A., Hammer, D., Sherin, B., & Kolpakowski, T. (1991). Inventing graphing: Metarepresentational expertise in children. *Journal of Mathematical Behavior, 10*(1), 117 –160.

English, L. D. (2003a). Mathematical modelling with young learners. In S. J. Lamon, W. A. Parker, & S. K. Houston (Eds.), *Mathematical modelling: A way of life* (pp. 3 –18). Chichester, UK: Horwood.

English, L. D. (2003b). Reconciling theory, research, and practice: A models and modelling perspective. *Educational Studies in Mathematics, 54*(2/3), 225 –248.

English, L. D. (2008). Introducing complex systems into the mathematics curriculum. *Teaching Children Mathematics, 15*(1), 38 –47.

English, L.D. (2010). Young children's early modelling with data. *Mathematics Education Research Journal*, *22*(2), 24 –47.

Franklin, C. A., & Garfield, J. (2006). The GAISE project: Developing statistics education guidelines for grades pre-K-12 and college courses. In G. Burrill & P. Elliott (Eds.), *Thinking and reasoning with data and chance* (68th Yearbook, pp. 345 –376). Reston, VA: National Council of Teachers of Mathematics.

Greer, B., Verschaffel, L., & Mukhopadhyay, S. (2007). Modelling for life: Mathematics and children's experience. In W. Blum, W. Henn, & M. Niss (Eds.), *Applications and modelling in mathematics education: ICMI Study 14* (pp. 89 –98). Dordrecht: Kluwer.

Hanner, S., James. E., & Rohlfing, M. (2002). Classification models across grades. In R. Lehrer, & L. Schauble (Eds.), *Investigating real data in the classroom* (pp. 99 –117). New York: Teachers College Press.

Konold, C., & Higgins, T. L. (2003). Reasoning about data. In Kilpatrick, J., Martin, W. G., & Schifter, D. (Eds.). (2003). *A research companion to Principles and Standards for School Mathematics*. Reston, VA: National Council of Teachers of Mathematics.

Langrall, C., Nisbet, S., Mooney, E., & Jansem, S. (2011). The role of context expertise when comparing data. *Mathematical Thinking and Learning Journal*, *13*(1 –2), 47 –67.

Langrall, C., Nisbet, S., Mooney, E., & Jones, G. (2008). Elementary students' access to powerful mathematical ideas. In L. D. English (Ed.), *Handbook of international research in mathematics education* (2nd ed.) (pp. 109–135). New York: Routledge.

Leavy, A. (2007). An examination of the role of statistical investigation in supporting the development of young children's statistical reasoning. In O. Saracho & B. Spodek (Eds.), *Contemporary perspectives on mathematics in early childhood education* (pp. 215 –232). Charlotte, NC: Information Age Publishing

Lehrer, R., & Lesh, R. (2003) Mathematical Learning. In W. Reynolds & G. Miller (Eds.) *Comprehensive handbook of psychology* (Vol. 7, pp. 357 –390). New York: John Wiley.

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

Lehrer, R., & Schauble, L. (2000). Inventing data structures for representational purposes: Elementary grade students' classification models. *Mathematical Thinking and Learning, 2*(1/2), 51 –74.

Lehrer, R., & Schauble, L. (2005). Developing modeling and argument in the elementary grades. In T. Romberg & T. Carpenter (Eds.) & F. Dremock (Eds.), *Understanding mathematics and science matters* (pp. 29 –53). Mahwah, NJ: Lawrence Erlbaum Associates.

Lehrer, R., & Schauble, L. (2007). Contrasting emerging conceptions of distribution in contexts of error and natural variation. In M. C. Lovett & P. Shah (Eds.), *Thinking with data* (pp. 149 –176). New York: Taylor & Francis.

Lesh, R. A., & Kelly, A. E. (2000). Multi-tiered teaching experiments. In R. A. Lesh & A. Kelly (Eds.), *Handbook of research design in mathematics and science education* (pp. 197 –230). Hillsdale, NJ: Lawrence Erlbaum.

Lesh, R., & Lehrer, R. (2000). Iterative refinement cycles for videotape analyses of conceptual change. In R. Lesh & A. Kelly (Eds.), *Research design in mathematics and science education* (pp. 665 –708). Hillsdale, NJ: Lawrence Erlbaum.

Lesh, R., & Zawojewski, J. S. (2007). Problem solving and modeling. In F. Lester (Ed.), *The second handbook of research on mathematics teaching and learning* (pp. 763–804). Charlotte, NC: Information Age Publishing.

Moore, D. S. (1990). Uncertainty, In L. Steen (Ed.). *On the shoulders of giants: New approaches to numeracy* (pp. 95 –137). Washington, DC: National Academy Press.

National Council of Teachers of Mathematics (2006). *Curriculum Focal Points.* Retrieved October 20, 2008, from http://www.nctm.org/standards/content.aspx?id=270

Olson, D. R. (1994). *The world on paper*. Cambridge, UK: Cambridge University Press.

Shaughnessy, M. (2010). *Statistics for all: The flip side of quantitative reasoning*. Retrieved 14 August, 2010, from http://www.nctm.org/about/content.aspx?id=26327

Stillman, G., Brown, J., & Galbraith, P. (2008). Research into the teaching and learning of applications and modelling in Australia. In H. Forgasz, A. Barkatsas, A. Bishop, B. Clarke, S. Keast, W. Seah, & P. Sullivan (Eds.), *Research in mathematics education in Australasia, 2004 –2007* (pp. 141 –164). Rotterdam, The Netherlands: Sense Publishers.

Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: Ground theory procedures and techniques*. Los Angeles, CA: Sage.

Watson, J. (2006). *Statistical literacy at school: Growth and goals*. Mahwah, NJ: Lawrence Erlbaum.