DATA MODELING IN ELEMENTARY AND MIDDLE SCHOOL CLASSES: A SHARED EXPERIENCE

Lyn D. English
Queensland University of Technology
Australia
l.english@qut.edu.au

This paper argues for a renewed focus on statistical reasoning in the elementary school years, with opportunities for children to engage in data modeling. Data modeling involves investigations of meaningful phenomena, deciding what is worthy of attention, and then progressing to organizing, structuring, visualizing, and representing data. Reported here are some findings from a two-part activity (Baxter Brown's Picnic and Planning a Picnic) implemented at the end of the second year of a current three-year longitudinal study (grade levels 1-3). Planning a Picnic was also implemented in a grade 7 class to provide an opportunity for the different age groups to share their products. Addressed here are the grade 2 children's predictions for missing data in Baxter Brown's Picnic, the questions posed and representations created by both grade levels in Planning a Picnic, and the metarepresentational competence displayed in the grade levels' sharing of their products for Planning a Picnic.

Introduction

The need to understand and apply statistical reasoning is paramount across all walks of life, evident in the variety of graphs, tables, diagrams, and other data representations that need to be interpreted. Elementary school 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; Watson, 2006). One approach in the beginning school years is through data modeling (English, 2010; Lehrer & Romberg, 1996; Lehrer & Schauble, 2007; Lehrer & Schauble, 2000).

Data modeling 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 modeling should be a fundamental component of early childhood curricula. Limited research exists, however, on such modeling and how it can be fostered in the early school years. Indeed, the majority of the research on mathematical modeling has been confined to the secondary and tertiary levels, with the assumption that elementary school children are not able to develop their own models and sense-making systems for dealing with complex situations (Greer, Verschaffel, & Mukhopadhyay, 2007).

In this paper, I first consider briefly the core components of data modeling relevant to the present activity, namely, structuring and representing data, and informal inference (specifically,

Wiest, L. R., & Lamberg, T. (Eds.). (2011). Proceedings of the 33rd Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education. Reno, NV: University of Nevada, Reno.

making predictions). I also consider the role of task context in data modeling. Specifically, I address the following questions:

- What was the nature of the three grade 2 classes' predictions for the missing values in a table of data for *Baxter Brown's Picnic*?
- What questions were posed and representations created by one second-grade and one seventh-grade class in *Planning a Picnic*?
- How was metarepresentational competence displayed in the sharing of products between the second-grade class and the seventh-grade class?

Structuring and Representing 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 analyze and assess their representations is important in addressing these early difficulties.

Constructing and displaying data models involves children in creating their own forms of inscription. By the first grade, children already have developed a wide repertoire of inscriptions, including common drawings, letters, numerical symbols, and other referents. As children invent and use their own inscriptions they also develop an "emerging meta-knowledge about inscriptions" (Lehrer & Lesh, 2003). Children's developing inscriptional capacities provide a basis for their 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). As Lehrer and Schauble (2006) emphasized, developing a repertoire of inscriptions, appreciating their qualities and use, revising and manipulating invented inscriptions and representations, and using these to explain or persuade others, are essential for data modeling. In a similar vein, diSessa has argued for the development of students' metarepresentational competence, which includes students' abilities to invent or design new representations, explain their creations, and understand the role they play (e.g., diSessa, Hammer, Sherin, & Kolpakowski, 1991).

Informal Inference: Making Predictions

There has been limited research on young children's abilities to make predictions based on data, an important component of beginning, informal inference. Although young children obviously do not have the mathematical background to undertake formal statistical tests, they nevertheless are able to draw informal inferences based on various types of data (Watson, 2007). Predictions can be based on aspects of the problem scenario and context, and children's understanding of the data presented. As pointed out by Watson (2006), one of the aims of statistics education is to help students make predictions that have a high probability of being correct. Yet in the real world, decisions are required where there is uncertainty and where several alternatives might be reasonable. Hence, young children's exposure to informal inference involving uncertainty is an important learning foundation if a meaningful introduction to formal statistical tests is to take place in secondary school.

The Role of Context

The nature of task design, including the task context, is a key feature of data modeling activities. 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). Moore emphasised that a data problem should engage students' knowledge of context so that they can understand and interpret the data rather than just perform arithmetical procedures to solve the problem.

The need to carefully consider task design is further highlighted in research showing that the data presentation and 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 (Cooper & Dunne, 2000; Pfannkuch, 2011). In designing the present activities, literature was used as a basis for the problem context. It is well documented that storytelling provides an effective context for mathematical learning, with children being more motivated to engage in mathematical activities and displaying gains in achievement (van den Heuvel-Panhuizen & van den Boogaard, 2008).

Methodology

Participants

The participants were from an inner-city Australian school, situated in a middle socio-economic area, with an enrolment of approximately 500 students from Prep (K) -7. The three first-grade classes (2009, mean age of 6 years 8 months) continued into the second year of the study, the focus of this paper (2010, mean age of 7 years 10 months, n=68). The grade 7 class (n=24), who participated in the *Planning a Picnic* activity, described below, had an age range from 12 to 13 years.

Research Design

A teaching experiment involving multilevel collaboration (English, 2003; 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.

Activities and Procedures

The final activity implemented in the second year of the 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 a story about Baxter Brown's picnic was created as the context for the first part of the activity (*Baxter Brown's Picnic*). For the *Planning a Picnic Activity*, two story picture books about foods and picnics were read to the grade 2 classes prior to their planning their own picnic.

For the *Baxter Brown's Picnic Activity*, the children (as a whole class) were presented with a table of six different items that he and each of his five canine friends chose to take on their picnic. The final column of the table was left blank, as indicated in Table 1 below. After discussing what they noticed about the values and variation in values across the table, the children were invited to predict the number of Oinkers that Baxter Brown and each of his friends might take on the picnic.

Table 1. Items Taken to Baxter Brown's Picnic

	Liver Straps	Beef Discz	Dentastix	My Dog Gourmet Beef	Bones	Oinkers
Baxter B.	3	5	2	1	3	
Monty	2	7	1	2	1	
Fleur	4	0	3	4	5	
Daisy	3	1	4	3	2	
Lilly	5	3	0	2	4	
Pierre	7	5	2	6	10	

For *Planning a Picnic* (grades 2 and 7), an initial whole-class discussion focused on questions the children might ask about planning a class picnic. In their groups, the children then listed five items they would like to take on the picnic, which were recorded by the teacher in a table on an interactive whiteboard. The children were subsequently asked what might be done with the data and what questions they might ask about the data. Each group's question was recorded on the board, with brief discussion on how some of the questions might be refined. In their groups, the children proceeded to answer their question and were to display their findings using whatever representation they liked. They were provided with a range of recording material including blank chart paper, grid paper, and chart paper displaying a circle shape. The children could use whatever of these materials they liked; no encouragement was given to use any specific recording material. On completion of the activity, the groups reported back to their class peers on how they answered their question. The grade 2 children were subsequently asked how their responses might compare with those of the other grade two classes, and were then asked to consider how the grade 7 classes in their school might respond to the activity. On the suggestion of one of the second-grade teachers, we administered the *Planning a Picnic* activity in one seventh-grade class. We then brought together the teachers and students from the second-grade class and the seventh-grade class for a sharing of how they worked the activity.

Data Collection and Analysis

In each of the second-grade classrooms, all whole-class discussions were videotaped and audiotaped; likewise, in each class, two focus groups (of mixed achievement levels and chosen by the teachers), were videotaped and audiotaped, with all tapes subsequently transcribed. There were 17 groups of second-grade children (3-4 per group), five in one class and six in each of the remaining two classes. For the seventh-grade class, the teacher chose mostly two-member groups, making 11 groups in total. The sharing of products between the two grade levels was videotaped and transcribed. All artifacts were collected and analyzed along with the transcripts. Where appropriate, iterative refinement cycles for analysis of children's learning (Lesh & Lehrer, 2000) were used, together with constant comparative strategies (Strauss & Corbin, 1990) in which data were coded and examined for patterns and trends.

Selection of Findings

Grade Two Children's Predictions for Baxter Brown's Picnic

In contrast to the children's use of informal inference in the first year of the study (English, in press), where they used the variation and range of values in a table of data to predict unknown values, the context of the present activity appeared to inhibit the children's ability to abstract the data from the context (Konold & Higgins, 2003). Each class initially identified the blank column as the first feature they noticed, with one child explaining, "Nobody wants Oinkers." In predicting how many Oinkers each of the dogs might take to the picnic, the children predicted small values less than 10, with their reasoning mainly based on the total number of other items each dog was bringing and the fact that if a larger number of Oinkers were brought to the picnic, the dogs "might get sick," "get a tummy ache," or "get fat." One child suggested zero, "because there has to be something that he doesn't like." There were some responses however, indicating an awareness of the need to consider the nature of the existing values, such as, "Because he (Monty) doesn't eat that much of anything else so he mustn't eat that much." In response to a child who predicted that Baxter Brown would take zero Oinkers, because he already has many other items, the teacher accepted the response as a reasonable prediction. Another student, however, disagreed, stating, "I don't think it's reasonable because he's pretty of a greedy guts so I think he would have more" (basing her decision on the existing item values for Baxter Brown).

On asking each class to consider the scenario of Baxter Brown taking 26 Oinkers, Monty 33, Fleur 50 etc., the majority of children used the task context to decide that these values were inappropriate. Comments such as, "They're um too big, the dogs would probably get a tummy ache and get sick" and "It's too heavy for them to carry to the picnic," were common. On the other hand, other responses suggested that some children were aware of the need to focus on the data itself, for example, "They would be bigger than all the numbers," "Ten is the highest number you can go up to," There's only one two-digit number," and "Because there would be too much."

Children's Questions and Representations for Planning a Picnic

The findings reported in this section focus on the responses of the selected second-grade class and the seventh-grade class. The table created by the grade 2 class appears below; a comparable table was developed by the grade 7 students.

Table 2. Picnic Items Chosen by the Grade 2 Class

Group1	Group 2	Group 3	Group 4	Group 5
choc chip cookies	sandwiches	blanket	food	cup cakes
fruit	fizzy drinks	fruit	picnic basket	cake
sausage rolls	cookies	cake	sunscreen	juice
cordial	fruit	esky	drinks	fruit soft drink
sandwiches	fruit pudding	soft drinks	chairs	carrots

The questions posed by each class are listed below. These questions addressed the table of data displaying the items each group would take on their picnic.

Grade 2

- Did everybody choose healthy items?
- Is there a most popular food?
- What are the different types of items?
- Did everybody choose the same items?
- Is there a most popular item?

Grade 7

- What percentage of foods are in each food group?
- How many different picnics brought 2 or more healthy foods?
- What percentage of foods are unhealthy?
- What is the most popular item on the list, soft drink or sandwich?
- How many items are processed foods in each picnic?
- What percentage of groups brought fruit on their picnic?

- What was the most popular food?
- What percentage of groups chose sandwiches compared to groups who chose fruit to bring on their picnic?
- Is there more healthy than unhealthy food?
- What food group does the majority of food from all of the picnics come from?

Not surprisingly, the grade 2 students' questions were less sophisticated than their older counterparts, resulting in a few difficulties in answering their question and representing their findings. Nevertheless, the younger children displayed a wider range of representations, albeit less sophisticated than their older counterparts. Each grade 2 group made use of the table of chosen items given to them (Table 1) and displayed a range of inscriptions in analyzing their data. For example, one group who addressed their question, "Did everybody choose healthy items?" placed a X on what they considered to be unhealthy items, a * on healthy items, a 0 on "things that aren't food," and a created symbol of mixed shapes for "fruit/sugar." This group also drew a food pyramid, with a focus on healthy and unhealthy items, and followed this with a third representation, a circle divided into halves displaying drawings and labels of "junk food" and "healthy foods." Four of the grade 2 groups made a list of selected items, before constructing a bar graph (3 groups) or a pie graph (cut into thirds; 1 group). One of the groups explained how their construction of a bar graph made them change their initial answer to their question:

> *Our question is, "Is there a most popular food?" There is, there is, the answer was, there is not any popular food because there were, there's 3. We, um, our finding things out was that all the things, we've made all the things that go together on the graph here and then we found out, we recorded how many different stuff there was and on one square it means that um, it means that there was one thing, on two squares it means that's there's two things and it keeps on going up to 6. And then we found out that there was no most popular food. There were 3 tying, drinks, cakes and picnic stuff…We wrote first that there was a popular thing but then when we ended up doing the graph, it ended up that there was, um, three populars.*

In contrast to the grade 2 children, all but one of the 11 grade 7 groups chose only one representation, with vertical bar graphs and pie graphs being equally popular (each chosen by 5 groups, with the display of percentages prominent). One group who created a pie graph also made a tally chart first. The remaining group created a line graph. When asked why they selected a line graph in preference to a bar graph, this group explained, "Well we thought because there are so many foods, drawing bars to make them seeable would be quite squishy; we just thought it would be easier to read if it was a line graph."

The children's foregoing explanations indicate a metarepresentational competence where they were able to explain and justify the representations they generated and also understand the role these played. Further evidence of such competence was evident in the sharing of products between the two grade levels.

Children's Metarepresentational Competence in Sharing Products for Planning a Picnic

As indicated in the methodology section, one grade 2 class and one grade 7 class came together to share their products for the *Planning a Picnic* activity. The grade 2 teacher initially asked her class to recall how they predicted the grade 7 students might work the activity. The children responded that "They won't have the same ideas," and "We said that they might be better because they'd had more years."

As the grade 7 class presented their products to their younger counterparts, there were several displays of metarepresentational competence at both grade levels. One grade 7 group reported that they solved their question using a bar graph that showed percentages of the particular items targeted in their question. When asked why they chose this representation, the group explained, "We tried a pie graph but we couldn't like split it into the right amount of groups." When invited to define a pie graph for the grade 2 children, one group member explained, "A pie graph is a circle that you put lines into and then color sections which is what, yeah, is what you chose." When the grade 2 children were asked to compare their bar graph representations with the grade 7 group, they responded that theirs was easier to read as "They (grade 7) used percentages and we don't know about percentages yet." However, the younger children were able to interpret the grade 7 representation when asked what the most popular and least popular item was:"Cause it's got the names at the bottom (labels under X axis). I was looking at the fruit one and I knew that it was the most…cause it's got the highest thing (bar) that goes up." In answering their question, "How many different picnics brought two or more healthy foods?" another grade 7 group justified their selection of a bar graph in preference to a pie graph by explaining, "Cause if you did like a pie graph… you wouldn't really show each group and how many items each individual group brought."

A follow-up grade 2 class discussion on how their working of the activity compared with the grade 7 students included comments such as: "We took more healthy food than they did;" "They were really bad choices;" "They did pie graphs and we didn't know like how to;" and (they did) "The line graph." In a follow-up question, the grade 2 children commented that 100% means "all of it" (circle) and "to understand the pie, we can look at it and see if it adds up to 100%."

Discussion and Concluding Points

Three main issues arising from the children's responses are worth highlighting here—the role of task context in the grade 2 children's predictions, the nature of the questions and representations created by the grade 2 and 7 classes, and the metarepresentational competence displayed in their sharing of products.

As previously noted, children need to appreciate that data are numbers in context, while at the same time abstract the data from the task context. Although context provides meaning in statistics (Garfield & Ben-Zvi, 2008), it can create both obstacles and supports in student's statistical reasoning (Pfannkuch, 2011). The purposefully created context of Baxter Brown and his canine friends organizing a picnic appeared to hinder the children's analysis of the table of data (Table 1). Only a few children justified their predictions by considering the nature (range and/or variation) of the values displayed, with the majority making contextual inferences such as the need to consider the dogs' health. The role and impact of task context require careful consideration in designing statistical activities; clearly a good deal more research is needed here to guide the development of data modeling in the early years.

Posing questions about the class selection of picnic items was a comparatively new learning experience for the second-grade children and did present some difficulties, resulting in discussion on how some of the questions might be refined. Such difficulties can be expected—transforming initial questions into more specific statistical questions is not an easy step, especially for young children (Konold & Higgins, 2003). Not surprisingly, the grade 7 students generated more sophisticated questions, applying mathematical understandings they had developed during their additional years of schooling. Nevertheless, both grade levels displayed metarepresentational competence in generating, describing, explaining, and justifying their representations. Interestingly, most of the grade 2 children, in contrast to their grade 7 counterparts, created more than one representation and could identify the links between their representations. The sharing of products was a rich learning experience for both grade levels, providing opportunities for appreciating different approaches to dealing with data and for questioning, explaining, and interpreting the data models of others. Consideration should be given to creating such sharing opportunities across grade levels.

References

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

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

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

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

English, L. D. (In press). Data modelling with first-grade students. *Educational Studies in Mathematics Education*.

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.

Garfield, J, B., & Ben-Zvi, D. (2008). The discipline of statistics education. In J. B. Garfield & D. Ben-Zvi (Eds.), *Developing students' statistical reasoning; Connecting research and teaching practice*. Springer.

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

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

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

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

Lehrer, R., & Lesh, R. (2003) Mathematical Learning. In W. Reynolds & G. Miller (Eds.) *Comprehensive Handbook of Psychology*, Volume 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*, 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.

Lehrer, R., & Schauble, L. (2006). Cultivating model-based reasoning in science education. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 371-386). NY:

Cambridge University Press.
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). NY: 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 Associates.
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 Associates.
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.
Pfannkuch, M. (2011). The role of context in developing informal statistical reasoning: A classroom study. *Mathematical Thinking and Learning, 13*(1-2), p. 27-46.
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
Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: Ground theory procedures and techniques*. CA: Sage.
van den Heuvel-Panhuizen, M., & van den Boogaard, S. (2008). Picture books as an impetus for kindergartners' mathematical thinking. *Mathematical Thinking and Learning: An International Journal, 10*, 341-373.
Watson, J. M. (2006). *Statistical literacy at school: Growth and goals*. Mahwah, NJ: Lawrence Erlbaum Associates.
Watson, J. M. (2007). Inference as prediction. *Australian Mathematics Teacher, 63*(1), 6-11.

Acknowledgement: This project 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.