



Queensland University of Technology
Brisbane Australia

This is the author's version of a work that was submitted/accepted for publication in the following source:

[English, Lyn D.](#) (2012) Data modelling with first-grade students. *Educational Studies In Mathematics*.

This file was downloaded from: <http://eprints.qut.edu.au/48400/>

© **Copyright 2012 Springer**

The original publication is available at SpringerLink
<http://www.springerlink.com>

Notice: *Changes introduced as a result of publishing processes such as copy-editing and formatting may not be reflected in this document. For a definitive version of this work, please refer to the published source:*

<http://dx.doi.org/10.1007/s10649-011-9377-3>

Data modelling with first-grade students

Lyn D. English

Abstract: This paper argues for a renewed focus on statistical reasoning in the beginning school years, with opportunities for children to engage in data modelling. Results are reported from the first year of a 3-year longitudinal study in which three classes of first-grade children (6-year-olds) and their teachers engaged in data modelling activities. The theme of Looking after our Environment, part of the children's science curriculum, provided the task context. The goals for the two activities addressed here included engaging children in core components of data modelling, namely, selecting attributes, structuring and representing data, identifying variation in data, and making predictions from given data. Results include the various ways in which children represented and re-represented collected data, including attribute selection, and the metarepresentational competence they displayed in doing so. The "data lenses" through which the children dealt with informal inference (variation and prediction) are also reported.

Keywords: Data modelling . Young learners . Informal inference . Statistical reasoning . Problem solving

1. Introduction

Young children are very much a part of our data-driven society, with early access to computer technology and daily exposure to the mass media where various displays of data and related reports can easily mystify or misinform, rather than inform, their inquisitive minds. 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; Shaughnessy, 2010). Rethinking the nature of young children's statistical experiences is imperative—we need to 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).

In this article, I first argue for the need to review young children's statistical experiences, with a focus on data modelling. Next, I describe the first year of a 3-year longitudinal study in which three classes of first-grade children and their teachers engaged in data modelling activities. Findings from two of the activities are then addressed, with a focus on:

1. How children structured, represented, and re-represented their collected data and their metarepresentational competence in doing so;

2. How children identified variation in a table of data and made predictions about missing values.

2. Data modelling with young learners

Data modelling provides a powerful vehicle for illuminating young children's learning potential (English & Watters, 2005) and for meeting the calls for early curriculum renewal in statistics. Such modelling engages children in extended and integrative experiences in which they generate, test, revise, and apply their own models in solving problems that are meaningful to them. The early work of Hancock, Kaput, and Goldsmith (1992) viewed data modelling as "a complete process of inquiry" (p. 338), using data to solve real-world problems and to answer genuine questions. Their research highlighted the importance of data creation and data analysis, considering them to be "two indispensable and mutually informing halves of data modelling competence;" data creation, however, was cited as "the neglected counterpart of data analysis" (p. 339). Later research by Lehrer and Schauble (e.g., 2005) and Lehrer and Lesh (2003) has focused on younger children and highlighted the developmental process of data modelling. The process begins with young children's inquiries and investigations of meaningful phenomena, progressing to deciding what is worthy of attention (i.e., identifying attributes of the phenomena), and then moving towards organizing, structuring, visualizing, and representing data. Data modelling also involves the fundamental components of beginning inference (Watson, 2006), which include variation and prediction, among others. In the remainder of this section, I address these core components of data modelling with young children.

2.1 Generating and selecting attributes

Early experiences with data modelling include the creation, analysis, and revision of data classification models. A fundamental element in creating these models is selecting attributes and classifying items according to these attributes. 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.

2.2 Structuring and representing data

Models are typically conveyed as systems of representation (Lehrer & Schauble, 2006). Structuring and displaying data are fundamental here, where "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 that they choose and to analyze and revise their creations are important in addressing these early difficulties. The need for classroom experiences that provide such opportunities has been emphasized over the years (e.g., Curcio, 2010; Lehrer & Schauble, 2007; Makar & Rubin, 2009; Russell, 1991), yet young children's typical exposure to data structure and displays has been through conventional instruction on standard forms of representation.

Constructing and displaying their 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 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 modelling. 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; diSessa, 2004). Yet, students 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.

2.3 Variation and prediction

Variation lies at the heart of statistical reasoning and is linked to all aspects of statistical investigations (Cobb & Moore, 1997; Garfield & Ben-Zvi, 2007; Pfannkuch, 2005; Watson, 2006). Indeed, as Watson (2006) indicated, the reason data are collected, graphs are created, and averages are computed is to “manage variation and draw conclusions in relation to questions based on phenomena that vary” (p. 21). The importance of variation cannot be underestimated in the development of children’s statistical reasoning, beginning with the earliest grade levels (Garfield & Ben-Zvi, 2007). Unfortunately, this is not happening in many classrooms where teachers fail to make specific links to variation whenever they implement activities in data and chance (Watson, 2006). Research on young children’s reasoning about variation is limited, although the work of Watson (e.g., 2007) has indicated that young students do have a primitive understanding of variation.

There has also been limited research on young children’s abilities to make predictions based on data, another important element of beginning 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.

Given the limited research on informal inference in the beginning school years, Konold’s work on seeing data through different lenses provides a promising way of exploring how young children might deal with variation and prediction (e.g., Konold, Higgins, Russell, & Khalil, Data seen through different lenses. University of Massachusetts: unpublished manuscript). Along with others (e.g., Cobb, 1999; Rubin, Hammerman, & Konold, 2006), Konold has highlighted the difficulties students experience in seeing data from an entire aggregate perspective (the collection as a whole). Rather, students tend to focus on individual values, as pointers, case values, or classifiers. The first type, pointers, refers to the larger event from which the data

were drawn, without a focus on the actual data values (e.g., “I remember when we did that. We went down to the canteen.”).

Case values give information about the value of some attribute for individual cases, such as “That cross there is me; I go to the canteen three times a week.” Classifiers indicate the frequency of cases with a certain attribute value and without an overall view (e.g., “Lots of us go to the canteen three times a week”). The aggregate perspective is considered a unity comprising emergent statistical properties, such as distributional shape and spread (e.g., “Our class goes to the canteen from one to 5 days a week, but most of us go three times a week. Few go five times a week.”).

3 Methodology

3.1 Participants

Three classes of first-grade children and their teachers in an inner city Australian school participated in the first year of the study. The school is situated in a middle socio-economic area and has an approximate enrolment of 500 students from the preparatory year through to seventh grade. Each of the first-grade classes comprised 25 or 26 students, with a mean age of 6 years 8 months. The children's previous experiences in working with data were limited to sorting items (e.g., colored bears) and completing picture graphs (e.g., of favorite pets, hair color).

3.2 Research design

A teaching experiment involving multilevel collaboration (English, 2003; Lesh & Kelly, 2000) was adopted in this study. Such collaboration 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 maximize learning at each level. Given that the teachers' involvement in the study was vital, regular half-day professional development meetings were conducted with the first-grade teachers. These meetings introduced the teachers to the study, explored their current mathematics and science curricula, developed and refined activities, reviewed children's developments, and reflected on their professional development.

3.3 Task design

The nature of task design, including the task context, is a key feature of data modelling 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 emphasized 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). Picture story books that addressed the overall theme of "Looking after our Environment," a key theme in the teachers' curriculum at the time, were selected.

3.4 Activities and procedures

A series of three, multi-component problem activities was implemented in each class by the teacher, the researcher (author), and a senior research assistant. Each activity began with a teacher-led whole-class discussion on the associated story book, followed by the teacher explaining to the class the activity that was to follow. The children then worked the activity in small groups of three to four. As the children undertook the activities, we moved among the groups to assist in their recording, as the children had emerging writing

skills at the time. Our role was to facilitate, not give the children direct instruction. We were keen to see how the children developed their own approaches to working the activities.

Children's responses to the second and third activities, namely, Fun with Michael Recycle and Litterbug Doug, are the focus of this article. The Australian picture story books that served as the basis for these activities were Michael Recycle (Bethel, 2008) and Litterbug Doug (Bethel, 2009). The former tells the story of Michael Recycle who came from the sky to clean up a very dirty town, with his motto, "I'm green and I'm keen to save the planet." Litterbug Doug was originally a very dirty creature who lived in a pile of rubbish in a very clean town. A "green-caped crusader" then swooped to the Earth to reform Litterbug Doug. As a consequence, Litterbug Doug became the Litter Police for the town and enthusiastically monitored the town's environment.

Fun with Michael Recycle involved two lessons (lesson 1, average duration of 30 min and lesson 2, 60 min). The activity addressed posing questions, identifying and generating attributes, organizing and analyzing data, and displaying and representing data in different ways. Prior to the lessons, the storybook, Michael Recycle, was read and discussed, and one teacher's classroom (which was used in turn by the three classes) was set up with collections of reusable/recyclable and waste items. Next, each child in each group was given two Post-It notes, and the group was directed to explore the classroom for these various items. Each group member was to draw and name an item on each Post-It note. The groups subsequently returned to their group desk and proceeded to discuss the attributes of their items, then organize, analyze, and represent their data however they chose (on a large sheet of paper provided). On completion, the groups reported back to the class on how they represented their data. A brief whole-class discussion followed on the nature of the attributes the children had identified and how they had organized and represented their data (e.g., "Why did you decide to arrange your Post-Its on the page like that?")






Following this, the children were advised that Michael Recycle "really likes the different ways you have represented your recyclable/reusable and waste items but would like you to represent them in a different way on your chart paper." The children were given a second sheet of paper to do so and were to leave their initial representation sheet intact. On completion, the groups reported back to the class, during which they were encouraged to explain their new representation and indicate how it differed from their first.

The second activity, Litterbug Doug, was designed to engage the children in interpreting tables of data, identifying variations in the data, posing questions, and making predictions. The activity was implemented in one lesson, average duration of 75 min. Prior to the lesson, the children read and discussed the storybook, Litterbug Doug. The lesson began with the teacher explaining that "Now that Litterbug Doug has become the Litter Police, the townsfolk are interested to see what he collects in Central Park during his first 3 days. They also want to know if Litterbug Doug is doing a good job of collecting litter in Central Park." The children were then shown part of Fig. 1, that is, the table without the Tuesday, Wednesday, and Thursday columns. It was explained, "As a start, the town's mayor asked Litterbug Doug to show him what he collected on his first day, Monday. Litterbug Doug showed the mayor what he saw and what he collected in the park." Next, the children were posed questions to explore their interpretation of the table, given that they had had almost no exposure to such a table.

Next, it was explained to the children that "Litterbug Doug has now collected litter in Central Park for 3 days and the townsfolk are keen to see how much he has collected." The children were then shown Fig. 1. In their groups, children were to explore the second table, first noting the numbers of items collected on the

second and third days, then how the data varied across the first 3 days and why this might be the case. Their next task was to consider the blank Thursday column. The children were to predict how many different items Litterbug Doug might have collected on Thursday. On completion, the groups reported back to the class on the variation they noticed in the data and on their predictions for Thursday. Finally, the whole class was asked if the mayor and his townsfolk would have been happy with Litterbug Doug's collection of litter over the week.

Fig. 1 Litterbug Doug Table

What Litterbug Doug collected	Monday	Tuesday	Wednesday	Thursday
	2	5	4	
	4	3	2	
	2	6	3	
	1	4	2	
	2	3	0	

Given the young age of the children and their lack of experience in reading tables of data, a small data set was deliberately chosen. Although a statistician would not predict from such a small data set, it is important that young children be exposed to prediction with uncertainty and to appreciate, in due course, that one has to ask further questions, such as those regarding the sampling and context (Watson, 2007).

3.5 Data collection and analysis

In each classroom, two focus groups of students were videotaped and audiotaped. The focus groups were of mixed achievement levels and were selected by the teachers, who aimed to place a competent reader in each group. The artifacts of all student groups were collected and scanned, and all whole-class discussions and group presentations were videotaped and audiotaped (with the exception of those students without parent permission). Digital photo- graphs were also taken.

Data were drawn from the transcripts of the group work of two focus groups in each of the three classes, together with the artifacts and class presentations of all groups who had permission to participate in the study. In total, data from 15 groups were analyzed for the first activity and data from 13 groups for the second activity (the latter as a result of student absenteeism). Using iterative refinement cycles for analyses of children's learning (Lesh & Lehrer, 2000), the transcripts of the focus groups were reviewed many times in conjunction with their artifacts and class presentations, as were all group artifacts and whole-class presentations and discussions. The data were coded and examined for patterns and trends using constant comparative strategies (Strauss & Corbin, 1990).

Of particular interest in children's working of Fun with Michael Recycle were the following questions:

1. How did the children select attributes, and structure and represent their data in each attempt?
2. How did the children's representations and inscriptions change from their first to their second attempt?
3. What metarepresentational competence did the children display in working the activities?

For Litterbug Doug, the focus was on how the children identified variation in the table of data and made predictions about the missing values.

4 Results

4.1 Fun with Michael Recycle

4.1.1 Children's attribute selections, representations, and inscriptions

Given the nature of the task context, with its focus on recycle/junk/waste/reuse, it is not surprising that these were the attributes most of the children chose to classify the items they had collected. Four groups chose different attributes, however, such as paper, cardboard, and plastic. One group chose the attribute of shape to classify their items (square, rectangle, circle), and explained, "We sorted by shape."

In their first attempt at representing their data, the majority of groups created pictographs with their sorted items pasted in either columns or rows within their respective categories (e.g., rows of recyclables on the left-hand side of a portrait-oriented sheet). Thirteen out of the 15 groups created a representation of this nature, which is an important foundational representation facilitating explicit links between the data collected and the task context (Konold & Higgins, 2003). The remaining two groups who did not make use of columns or rows placed their items randomly in their respective categories. One of

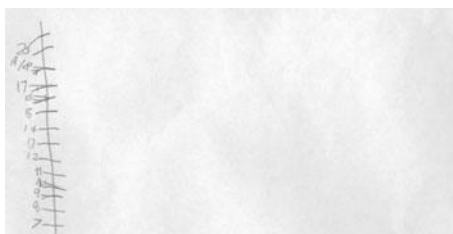
these groups justified their random placement by explaining, “cause we could fit more things in.”

4.1.2 Changes in children’s attributes, representations, and inscriptions

In moving from their first representation to their second, the children engaged in considerable debate over whether the attributes as well as the representations had to be changed. Seven groups chose to adopt new attributes to classify their items, such as the group who changed from the attribute of shape to the contextual attributes of reusable/recycle, compost. In the six focus groups, children’s debates drew forth a wider range of attributes, such as “heavy and light,” “hard and not hard,” “big and little,” and “things that fall down fast and things that fall down slow.” Children’s ability to look beyond the actual items and identify attributes that are not immediately apparent may be likened to what Lehrer and Lesh (2003) referred to as “lifting away from the plane of activity” (p. 377), a common feature of notational systems.

The children changed their representations on their second attempt in numerous ways, displaying changes in their pictographs (from rows to columns or vice versa), their inscriptions (using a mix of item names and drawings; item names only; drawings only; mix of ticks, crosses, and drawings), their paper orientation (from portrait to landscape or vice versa), and their selection of attributes. Seven groups changed their representation in one way only (e.g., used names of items only); four groups in two ways (e.g., changed from columns to rows or vice versa and used names of items only); one group in three ways (changed orientation, changed from columns to rows, and used a mix of names and drawings); and one group in four ways (changed orientation, changed from columns to rows, changed attributes, and used a mix of names and drawings). The remaining two groups changed their informal representation to more formal bar graphs, one of which is displayed in Fig. 2.

Fig. 2 A bar graph created by one group



There was a decline in the children's labelling of attributes, columns, or rows on their second attempt (from 11 instances to nine) but an increase in groups who recorded the number of each item type in their representations (from one instance to seven). Nevertheless, only half the groups failed to make numerical recordings suggesting the need for further learning experiences here, especially given that overlooking important or relevant information is one of the difficulties with early data modelling (Lehrer & Schauble, 2007).

4.1.3 Children's display of metarepresentational competence

Evidence of the children's metarepresentational competence can be seen in their explicit recognition of why they represented their data in the way they did; such competence guided both their mathematical thinking and how they communicated it (Lehrer & Lesh, 2003; Olson, 1994). For example, one group explained, "We put a line down to the bottom so we know which is junk and which waste is. So everyone knows which is junk and which is waste and we don't have to tell them" (the group did label their columns, however).

Another group, who placed their Post-It notes randomly in two appropriately labelled divisions on their sheet (first attempt), explained how they then manipulated their representation to cater for the difference in quantities of items:

Teacher: You left more space did you for your recycles?

Corey: Yes, cause there's going to be more of them.

Robert: More of those than that.

The group then continued to explain why they did not include duplicate items in their representation: "I went three there (pointing to the three Post-It notes across the bottom half of their sheet), cause I didn't have that one cause we already got it there (meaning there are two Post-It notes with cracked egg shells so he didn't position the duplicate item in

line with the others as it was a second drawing of the same item). It is interesting that this group did not include duplicate items on their second representation either, especially given young children's propensity to include redundant information in early data modelling (Lehrer & Schauble, 2007).

Another group, who created a vertical bar graph in their second attempt, also explained why they did not include duplicate items in their representation. They actually collected four recycle items and three waste items initially but in their second representation chose to only record one of the latter as "We had people that drew apple core, apple core, apple core. We um, so we made it one cause it was the same item." The groups' responses are interesting here. An awareness of the need to eliminate features that are not necessarily needed is an important goal of data modelling, but this is often difficult for young learners (Lehrer & Schauble, 2005).

Other evidence of children's metarepresentational competence was in their recognition that the quantities of items were conserved from one representation to another. Seven groups were able to recognize this conservation (e.g., "There's 2 there, 2 there, 4 there, 4 there (pointing to the first and second representations). Likewise, there was the group who displayed an understanding of "conservation of ideas" in creating their second representation. They created a grid in which an item was drawn and labelled in each square (e.g., eggs, pear, apple core). The group explained, "We've done the same ideas but we have done them differently. We've done them in rows, and like, um, we've done them in turns." The group then commented that their representation reminded them of a calendar and a graph.

4.2 Litterbug Doug

For the Litterbug Doug activity, findings from the analysis of all the groups are presented first. Next, the developments of two of the six focus groups

are detailed.

4.2.1 Children's identification of variation

Children's written and verbal responses revealed a number of different approaches to identifying variation: They totalled across rows (five groups), totalled each column and compared the totals (seven groups), compared values across rows (one group), identified items with the same value (one group), and totalled all values (five groups). Six groups displayed more than one of these approaches.

To gain insights into how the children analyzed the table, Konold et al.'s (2004; Data seen through different lenses. University of Massachusetts: unpublished manuscript) data lenses were applied, specifically, the case values, the classifiers, and the aggregate lenses, with modification made to the last lens. As indicated in the case studies, addressed next, children often switched lenses as they worked the activities. With Konold et al.'s case values lens, the unit of analysis is an individual case and the analysis focuses on considering the values of particular cases. Children's responses that suggested they viewed the data through such a lens included the following:

- Totalling across the rows and recording the number of each item collected;
- Totalling all the item values displayed in the table and recording the total;
- Identifying items with the same value (e.g., "There are three 3 s, have 3 there and 3 there and 3 here" [referring to three cans on Tuesday, three newspapers on Wednesday, and three cheese on Tuesday]);
- Noticing the increase in values for Tuesday ("The people have littered more and more on Tuesday."); and
- (In comparing values of items across rows) "It's little, big, little, except for this one (cans)."

The classifiers lens involves considering the frequency of cases with a particular value, without attention to the data collection as a whole. Children's responses that suggested they were using such a lens included: "He collected more apple cores on Tuesday and he collected less on Monday;" "On Monday he collected more drink cans than Wednesday;" and "He had less cheese on Wednesday and he had more on Tuesday."

Konold et al.'s (2004; Data seen through different lenses. University of Massachusetts: unpublished manuscript) final lens, the aggregate lens, is where the entire distribution of values is the perceptual unit. Although viewing through such a lens is difficult (Rubin et al., 2006), even for late primary and early secondary students, there appeared to be some evidence of what I term a pre-aggregate lens. That is, all of the data in the table were considered and frequencies compared and/or trends noted. Examples here include:

- Identifying Monday's and Tuesday's totals as "the same" and "Tuesday he had collected the most items;"
- (Comparing columns) "two didn't change and one did;"
- (Comparing column totals and values across the rows) "Monday and Wednesday are both the same but the rows are not the same; not the same in numbers;" and
- (Referring to column totals and applying contextual knowledge in doing so) "Well, first he didn't find that much cause it was his first day. And then he knew more so he found more and then he found so much that he couldn't find that much so it went down again."

4.2.2 Children's predictions

Predictions for the numbers of items Litterbug Doug might have collected on the Thursday suggested that the children had an informal awareness of the range and variation in the existing data. Twelve groups recorded

predictions of values ranging from 0 to 10. All but one of these 12 groups explicitly recognized that wild outliers (e.g., 56, 45) would be unlikely, as indicated later in the case study of Eric's group.

In their class presentations, five groups indicated that they considered the frequencies of the values across the rows of the table. They avoided repeating a quantity, or repeated a quantity, or gave a quantity that was not in the existing row. For example, when asked why a group recorded seven cans for Thursday, they explained, "Because there was no seven in that one. We did a number that wasn't in that line." One group justified their recording of four newspapers as, "Cause he found six but then he didn't find that much on the other 2 days so I thought to do four cause he didn't find that much on the other 2 days."

Two other groups displayed a more sophisticated awareness of trends in the data ("going up and down"). For example, "They went up and down (indicating Monday to Tuesday to Wednesday for the apple cores), then it kept counting down (referring to the three they added to the Thursday column). One child in this group actually did a corresponding hand motion to illustrate the trend. Other approaches (three groups) to predicting the values for the Thursday column included the use of patterns, numerical sequences (e.g., 4, 3, 2, 1), and odd and even numbers.

Consideration is now given to two case studies, which provide more in-depth examples of how children worked the Litter Bug Doug activity.

4.3 Case studies

4.3.1 Trina's group

The group commenced the activity by first viewing the data through a case values lens, adding the total number of items Litterbug Doug collected for each day. Switching to a classifiers lens, the group drew the conclusion that "On Tuesday he did the most" and that Monday and

Wednesday were a “tie,” and recorded “Tuesday has the most. It has 21 things.” The group then used their contextual knowledge of looking after the environment to identify the recyclable and rubbish items, stating that, on Tuesday, Litterbug Doug has collected more rubbish than you can recycle (indicating that that banana skin, cheese, and apple core are not recyclable). In considering how the values of the items changed across the 3 days, the group reverted to a case values lens, noting that “there was a pattern,” (11, 21, 11 [in the column totals]).

Continuing through a case values lens and applying context knowledge in doing so, the group viewed the values of individual items across the rows, and in doing so, created a rating scheme, namely, “a kind of good sign,” “a good sign,” “a really good sign,” “a bad sign,” and “a really bad sign.” The following excerpt illustrates their deliberations here:

Trina: Um, on Monday for the cheese he had 2 and on Tuesday for the cheese he had 3 and on Wednesday he had zero, so that’s a good sign...And the banana skin on Monday he collected 1 banana skin which was a good sign as well...And on Tuesday he collected 4 banana skins, which is a really good sign...And 2 (referring to Wednesday) is sort of a good sign cause it’s 1 more than 1 (referring to Monday).

Aaron: No, that’s actually a very good sign.

Trina: That’s good (referring to the one banana skin for Monday).

Harry: That is because there is only 1 banana skin lying on the ground

Trina: That’s good Aaron, that’s good if you have one or zero. On Monday they had 2 newspapers which is sort of good and 6 is really bad but you can recycle.

The group’s interpretations of their ratings appeared to change as they considered the different values. For example, the larger values were

classified as both “good” and “bad,” as were the smaller values. The children appeared to have two types of ratings, depending on whether the items were recyclable or not. For example, they considered the larger amounts of apple cores (Tuesday and Wednesday) to be “not very good,” but “2 is a pretty good sign” (Monday). In contrast, for the recyclable items, the four drink cans on Monday were considered “a good sign because we can recycle them.”

When asked if there is anything else they noticed about the data, the group commented that Litterbug Doug did not collect anything on Thursday and suggested “It could mean he didn’t find any things” or “Maybe because those days (Monday–Wednesday) he found everything!” Prior to considering possible data for Thursday, the group continued working through a case values lens and decided to add all the data, concluding that Litterbug Doug had collected 43 items altogether from Monday to Wednesday.

Predicting values for the Thursday column generated considerable group debate. Tina suggested one apple core while Aaron wanted to record 14 items of cheese, because this was the total number of cheese items across the 3 days. Continuing through a case values lens, Trina noted the nature of the item values across the rows and disagreed with Aaron, stating that amount was “silly” and suggested just four pieces of cheese. When her group members wanted to next record 15 cans for Thursday, Trina again objected and said “It’s just too much and really silly” and recorded five instead.

4.3.2 Eric’s group

Like Trina’s group, Eric’s group initially explored the data through a case values lens, but commenced by comparing individual item values, such as noting that there is only “one zero” (referring to the zero cheese on Wednesday) and “there is only one six and no other sixes” (referring to the six newspapers on Tuesday). Still viewing the data through a case

values lens, one group member, Jacob, suggested “Let’s count how many there is altogether” and proceeded to record 11 under the Monday column. He then decided to add all the values, and, to assist him here, he drew arches that connected each value across each row. Claiming “there’s 43 altogether,” he drew three lines, one from the bottom of each column, to connect to his recorded numeral, 43. His actions here support past research demonstrating that young children do invent a variety of inscriptions designed to meet particular goals and purposes (Lehrer & Lesh, 2003; diSessa et al., 1991); in this case, Jacob used his inscriptions to first assist him in totalling item values and then to indicate from where his recorded total was derived.

When asked what they noticed about the values as the days progressed, Jacob continued to use a case values lens, stating, “There’s a two there, five there, and a four there (referring to the row of apple cores) and they, like one, two, three, four, they aren’t in order.” Other group members noted similar cases: “No threes in there (referring to the Monday column) and threes in there (referring to the Tuesday column).”

The group then switched to a classifiers lens, considering the respective column totals and stating that more items had been collected on Tuesday than Monday. It was also noted that “he collected more newspapers than anything else.” When asked for further observations, the group reverted to a case values lens, comparing individual item values (e.g., “there’s only one 1” [referring to the Monday column], “six is the biggest” [the Tuesday column], and that each row had “different numbers.”

After making further comparisons of individual item values, Kristy stated that Litterbug Doug had collected the most items on Tuesday, however, did not use column totals to determine this. Rather, she used both case values and classifiers lenses to draw her conclusion: “Cause he’s got two 3’s on it and there’s the six (referring to the Tuesday column) and “so it’s more than that (Monday’s column), and more than that one (Wednesday’s column).

The group progressed to deciding on values for the Thursday column. Hamish began by claiming the new values should be 4, 10, 20, 30, and 40 (referring to the cheese, banana skins, newspapers, cans, and apple cores, respectively). Kristy, however, disagreed with his suggestion, claiming that “it’s kind of too high.” She was considering trends in the existing data to support her claim:

Cause if he could collect, if he could have collected that many, some of them might have been, it might have been on here (meaning those larger values would have appeared on the previous days). So it’s too many.

Kristy explained further that, since the larger values do not appear in the existing data, then “they’d have to be lower than that” and suggested five banana skins, three newspapers, four cans, six apple cores, and two cheese. It would seem that Kristy was viewing the data through a pre-aggregate lens in making her prediction as she was taking into consideration the nature of the entire values displayed in the table. Switching to a classifiers lens, the group compared their Thursday prediction with the existing data. Kristy noted that “Thursday is more than Wednesday cause it’s got six apple cores and this one (Wednesday) doesn’t have six.” The other members of the group then commented that Thursday is “way more than Monday.”

5 Discussion and concluding points

This paper has argued for a renewed focus on statistical reasoning in the early school years, with opportunities for children to engage in data modelling. Data modelling is a powerful means of illuminating young children’s learning potential; it engages children in extended and integrative experiences in which they generate, test, revise, and apply their own models in solving meaningful problems. The goal of the present research was to investigate young children’s data modelling where they select attributes, structure and represent (and re-represent) collected

data, and deal with informal inference (variation and prediction).

With respect to the children's attribute selections, the Fun with Michael Recycle activity focused on items that featured a recycle/junk/waste/reuse focus, so it is not surprising that most groups used these attributes in structuring and representing their data on their first attempt. On their second attempt, however, over half the groups changed their attributes along with their representations. Although not directed to do so, children's generation of new attributes demonstrated their ability to switch their attention from one item feature to another. That is, they needed to consider what was worthy of attention and what needed to be placed in the background, reflecting Lehrer and Lesh's (2003) notion of "lifting away from the plane of activity" (p. 377).

Children's representations for Fun with Michael Recycle were predominantly pictographs, which was likely influenced by the task design. As noted previously, task presentation and context can create both obstacles and supports in developing children's statistical reasoning. Here, the initial use of Post-It notes likely limited the forms of representation the children created. Nevertheless, the children did display an awareness of the structure of their pictographs, making effective use of rows and columns, and appropriate inscriptions. These early representations are important in assisting young children in abstracting or simplifying information they have gathered from their data collection (Konold & Higgins, 2003).

Divergent ways of creating re-representations were observed, with children again displaying a repertoire of inscriptions, including drawings, written text, numerical symbols, and other referents (ticks and crosses). Data modelling engages young children in creating their own forms of inscription and their responses here revealed their ability to change and incorporate several inscriptions in their re-representations.

Metarepresentational competence is an important factor in young

children's development of data modelling. Such competence, albeit emerging, was evident in the children's use of inscriptions, their structuring and displaying of data, their detection of redundant information, their awareness of the need to eliminate unnecessary features, and their conservation of ideas and quantities of items. The children had not received direct instruction on these components; their seemingly naturally developing metarepresentational competence appeared to play a substantial role in shaping their learning and reasoning in working the activity (Lehrer & Lesh, 2003). Although a good deal more research is needed here, the development of young children's metarepresentational competence should receive greater attention in early mathematics curricula.

Children's working of the Litterbug Doug activity indicated that these young children could deal with informal inference in analyzing a table of data, specifically, identifying variation and making predictions. Although some might argue that variation and prediction cannot be made from such small data sets and that various contextual factors can influence children's predictions, it is important that young children be given opportunities to draw informal inferences from situations involving uncertainty. As previously noted, children can draw inferences on aspects of the problem scenario and context and their understanding of the data presented. Opportunities for thinking imaginatively beyond the problem context, in conjunction with thinking about the data, should be an acceptable part of beginning, informal inference.

Children's responses to this activity revealed a variety of approaches to identifying variation in the item values. Applying Konold et al.'s (2004; Data seen through different lenses. University of Massachusetts: unpublished manuscript) data lenses, it appeared that children were using both case values and classifiers lenses in identifying variation, often switching between the two. The children focused on the value of individual cases (e.g., number of newspapers collected) and operated on

these values (e.g., totalling the number collected). They also considered the frequencies of several cases and compared these (e.g., more apple cores on Tuesday and less on Monday). There also appeared to be an emerging aggregate lens in the children's viewing of the data, which I have termed, a pre-aggregate lens. That is, children considered all of the values displayed, compared frequencies, and identified trends, such as the group who noted that two columns had the same totals but the rows had different values. Viewing through a pre-aggregate lens was also apparent in Kristy's (Eric's group) prediction for the Thursday values, where she took into consideration the entire range of values in the table to justify why wild outliers would be unlikely. Clearly, more research is needed in exploring the lenses through which young children view and analyze data; it would seem, however, that they utilize multiple lenses in dealing with variation and prediction. Activities that encourage different lens use, including a focus on perceiving data through a pre-aggregate lens, would enhance young children's statistical development.

Despite the increased calls for renewed attention to statistical learning in the early school years, research examining young children's developments here remains in its infancy. Data modelling provides one promising avenue for enriching and extending young learners' abilities to work with data and reason statistically.

Acknowledgements

The project reported here is supported by a 3-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 enthusiastic participation of the classroom teachers and their first-grade students, as well as the excellent support provided by my senior research assistant, Jo Macri. Professor Jane Watson's advice (personal communication) on the statistical learning of

young children is also gratefully acknowledged.

References

- Bethel, E. (2008). Michael recycle. Mascot, Australia: Koala Books.
- Bethel, E. (2009). Litterbug Doug. Mascot, Australia: Koala Books.
- Cobb, P. (1999). Individual and collective mathematical development: The case of statistical data analysis. *Mathematical Thinking and Learning*, 1(1), 5–43.
- Cobb, G. W., & Moore, D. S. (1997). Mathematics, statistics, and teaching. *The American Mathematical Monthly*, 104, 801–823.
- Cooper, B., & Dunne, M. (2000). Assessing children's mathematical knowledge: Social class, sex and problem solving. Buckingham, UK: Open University.
- Curcio, F. R. (2010). Developing data-graph comprehension in grades K–8. Reston, VA: National Council of Teachers of Mathematics.
- diSessa, A. A. (2004). Metarepresentation: Native competence and targets for instruction. *Cognition and Instruction*, 22(3), 291–292.
- diSessa, A. A., Hammer, D., Sherin, B., & Kolpakowski, T. (1991). Inventing graphing: Metarepresentational expertise in children. *The 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., & Watters, J. J. (2005). Mathematical modelling in third-grade classrooms. *Mathematics Education Research Journal*, 16(3), 59–80.

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., & Ben-Zvi, D. (2007). How students learn statistics revisited: A current review of research on teaching and learning statistics. *International Statistical Review*, 75(3), 372–396.

Hancock, C., Kaput, J. T., & Goldsmith, L. T. (1992). Authentic inquiry with data: Critical barriers to classroom implementation. *Educational Psychologist*, 27(3), 337–364.

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, NY: Teachers College.

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 ed., 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.

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

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). New York, 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.

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

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.

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

Pfannkuch, M. (2005). Thinking tools and variation. *Statistics Education*

Research Journal, 14(2), 5–22. Pfannkuch, K. (2011). The role of context in developing informal statistical inferential reasoning: A classroom study. *Mathematical Thinking and Learning*, 13, 1–2.

Rubin, A., Hammerman, J., & Konold, C. (2006). Exploring informal inference with interactive visualization software. In: A. Rossman & B. Chance (Eds.), *Working cooperatively in statistics education. Proceedings of the Seventh International Conference on Teaching Statistics*. Salvador, Brazil. [CDROM]. Voorburg, The Netherlands: International Statistical Institute.

Russell, S. J. (1991). Counting noses and scary things: Children construct their ideas about data. In D. Vere-Jones (Ed.), *Proceedings of the Third International Conference on the Teaching of Statistics* (pp. 158–164). University of Otago: Dunedin, NZ.

Shaughnessy, J. M. (2010). *Statistics for all: The flip side of quantitative reasoning*. Retrieved 14 August, 2010, from <http://www.nctm.org/about/content.aspx?id026327>

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 kinder- gartners' 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.