# Probabilistic Reasoning and Prediction with Young Children

Virginia Kinnear
*Flinders University*
<virginia.kinnear@flinders.edu.au>

Julie Clark
*Flinders University*
<julie.clark@flinders.edu.au>

This paper reports findings from a classroom based study with 5 year old children in their first term of school. A data modelling activity contextualised by a picture story book was used to present a prediction problem. A data table with numerical data values provided for three consecutive days of rubbish collection was provided, with a fourth day left blank. Children were asked to predict the amount of rubbish collected on the fourth day and to explain their prediction. The results revealed children's intuitive probabilistic reasoning competencies and the influence of task design on their reasoning.

Statistical learning is a key social and curriculum concern that has been driven by the increased availability of quantitative data in society. Statistics is increasingly used to add credibility to the way data are presented and how data-based arguments are used to persuade (Ben-Zvi & Garfield, 2004). The inclusion and naming of the Statistics and Probability strand in the Australian Curriculum: Mathematics (ACARA, 2013) reflects international research foci on the role of statistics in 21st century decision making.

Variation is one of the most distinguishing features of statistics (Franklin et al., 2007) and its presence has the potential to profoundly influence the reasoning processes children employ when solving statistical problems (Masnick, Klahr, & Morris, 2007). Statistical reasoning necessitates dealing with variation through inference that engages inductive reasoning. Probability (as expected variation), prediction (as expected outcome), variation (as uncertainty) and inference are integrated in statistics. The unpredictable presence of variation in data is the area in greatest need of specific instruction in statistics education (Moore, 1990), particularly with young children. This paper presents findings from a study with 5 year old children in their first year of school, which explored children's knowledge and reasoning brought to a task designed to elicit prediction from data.

## Prediction and Reasoning

Drawing inferences is a core process in statistical problem solving. It requires making decisions that extend beyond the immediate data to a broader context and so engages inductive reasoning (Paparistodemou & Meletiou-Mavrotheris, 2008). Inference interacts and coordinates real world knowledge with data structures and representations to find a logical solution to the problem (Lehrer, Kim, & Schauble, 2007). Outcomes for uncertain phenomena, however, have observable random order over repeated measurements, and the mathematical description of measured randomness is probability (Moore, 1990). Probability quantifies or describes random variation that cannot be explained by causal relationships (Langrall & Mooney, 2005). The presence of variation in data, however, is about the presence of uncertainty and is accompanied by difficulties in assigning causes or explanations about its cause. If relationships or patterns in data cannot be found, prediction about outcomes can be made from the data that are an estimate based on existing, observable variation (Wild & Pfannkuch, 1999). Probability can quantify the likelihood of something happening based on existing data. Prediction on the other hand, is about determining an outcome based on existing data, without necessarily quantifying the likelihood or determining why. The ability to predict comes from being able to model and

2014. In J. Anderson, M. Cavanagh & A. Prescott (Eds.). Curriculum in focus: Research guided practice *(Proceedings of the 37th annual conference of the Mathematics Education Research Group of Australasia)* pp. 335–342. Sydney: MERGA.

interrogate variation (Reading & Shaughnessy, 2004) and is facilitated by the ability to read data representations (Curcio, 1987).

## Prediction and the Nature of Variation

The teaching focus on probability in school that relies on chance devices (Schwartz & Goldman, 1996) aligns with a corresponding focus in research on probability with young children that engages theoretical probability of events using artificial chance devices that focus on the likelihood of an event occurring (Greer, 2001). In contrast, *natural variation* is variability that is "inherent in nature" (Franklin et al., 2007, p. 6) and found in the diversity of human experience. Recognising natural variation is foundational to understanding concepts that underpin statistical reasoning with variation (Wild & Pfannkuch, 1999). Working with natural variation is about seeing that chance, not just deterministic reasons, can explain the existence of variation and that both explanations can be mathematically described as probability (Moore, 1990). Watson (2006) describes the word chance as a "precursor to probability" (p. 127) having more intuitive connotations and less formal connotations than probability, which quantifies chance.

Studies on young children's probabilistic reasoning have built on the work of Fischbein and the concept of young children's intuitions about probability. Intuitions are subjective, described as "a feeling of obviousness, of intrinsic certainty" (Fischbein & Schnarch, 1997, p. 96) resulting from experiences with human behaviour, where estimations, prediction and random events are engaged. Studies have highlighted the strength and vulnerability of young children's use of their life experiences in reasoning and decision making. Early experiences that children have with artificial chance devices, such as dice and coins, can lead to the development of specific, deterministic reasons for chance events (Moore, 1990). Young children tend to use deterministic and subjective knowledge to judge or reason about random events in ways that affect probabilistic understanding (Langrall & Mooney, 2005). Studies have found, however, that tasks can engage young children's intuitive, informal understandings of probability in the absence of prior formal instruction (English, 2012; Mousoulides & English, 2009).

Research suggests that young children's intuitive responses may be influenced by task design, although research on tasks that engage young children in data prediction is limited. Most studies use tasks that require children to predict from graphs (e.g., Asp, Dowsey, & Hollingsworth, 1994 (Grades 4, 6 & 8); Watson & Kelly, 2002 (age 6 years); Watson & Moritz, 2001 (Preparatory to Grade 10)). Overall findings from these studies were that prediction was generally difficult to make and explanations for predictions were speculative or drew from personal knowledge.

## Method

### *Design and Procedure*

Participants were drawn from a State government primary school in South Australia. One class of fourteen children comprising nine boys and five girls in their first term of school attendance (mean age 5 years 2 months) and their teacher participated. A qualitative design-based research method, informed by the Models and Modeling perspective (Lesh & Doerr, 2003) underpinned the study. The classroom based study was undertaken over a ten week school term. Four separate data modelling activities incorporating picture story books that addressed a theme of recycling were implemented.

The picture story book used to contextualise the third modelling activity, implemented in week seven of data collection was Litterbug Doug (Bethel, 2009). The main character, Litterbug Doug, was lazy and messy until he was taught to recycle and he became a Litter Policeman, picking up rubbish for recycling. In the data activity problem scenario, Litterbug Doug had tidied up a town by collecting rubbish in the park.

A data table was used (Figure 1) to show numerical data values in columns for three consecutive days of rubbish collection for five items. The fourth day column, Thursday, was left blank. The data modeling problem asked the children to predict what amount of rubbish Litterbug Doug collected on Thursday for each item, and to say why they thought that amount would have been collected. Children worked independently in four small groups of three or four to find an agreed solution to the problem. The children were encouraged to represent the predicted values for Thursday in any way they liked.

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

*Figure 1.* Litterbug Doug data modelling problem table.

## *Data Collection and Analysis*

The group work data were captured using digital video cameras and audio recording devices on small portable stands on the group tables. All audio recordings were transcribed in full by the researcher including descriptions believed to be relevant to communication and understanding such as vocal emphasis, body movements and facial expressions.

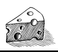
Inductive and deductive analysis across iterative cycles (Lesh & Lehrer, 2000) enabled data to be reduced, an adequate framework for interpretation to be developed, and core meanings that were theoretically based or emerged inductively from patterns or themes in the data to be identified.

# Results

## *Children's Data Prediction Models*

All four groups recorded a prediction for each of the missing values for Thursday. Children took turns within the groups to suggest a value for an item which was discussed and agreed to. Table 1 shows the numerical data values predicted for Thursday by each group of children. There was some replication of predicted values found in a relevant row that were taken from various positions across the row. All four groups' predicted values were between zero and ten.

Table 1
*Group Predicted Values*

| What Litterbug Doug Collected | Group 1 | Group 2 | Group 3 | Group 4 |
|---|---|---|---|---|
|  | 4 | 4 | 7 | 10 |
|  | 7 | 3 | 1 | 6 |
|  | 0 | 5 | 5 | 3 |
|  | 1 | 3 | 2 | 1 |
|  | 8 | 2 | 10 | 5 |

### *Considering Range and Frequency When Predicting*

The children were conscious of the range of data values provided in the table, found in their actions and discussion as they made prediction decisions. Most children were observed to touch existing numbers with their fingers or pencils and to visually scan along rows at various times as they considered what numeral to write in the blank Thursday column. These actions indicated that the children were considering the range and frequency of the existing data before making a predicted value decision. For example, in Group 1, Isabel carefully scanned across the "tin can" row as she said, "I think Litterbug Doug, um, I think tin cans, I think Litterbug Doug collected um, (scans back and forth along the row of numbers 4, 3, and 2) 5 tin cans". In Group 2, Eliot predicted how many apple cores were collected on Thursday, explaining:

> Um, I think that, um, Litterdut Bug [sic] (scans the data table up and down columns) … Litterbug Doug collected zero apple cores on … no, um, um er, (scans along the apple core row) I think Litterbug Doug, ah, collects that many (points to the number 4 in the apple core row). Guys, what do you think? That Litterbug Doug collected 4 on that day (points to the Thursday column)?

Predicted values were explained based on existing values already provided in the data table. The following exchange took place between the group members in Group 1 as they determined a predicted value for the number of tin cans:

> Isabel: I think Litterbug Doug collected um, (scans the data table) five tin cans.
> Carl: I think two (pauses, scans the data table), no I think five as well.
> Toby: (scans the rows) I think six.
> Isabel: (scans the data table) Actually, I think seven, because there's no seven.

Similar consideration of the range and frequency of existing table values was found in Group 2 where Jade explained her prediction of four tin cans, saying, "Ah, because um I think he collected um, 4 tin cans on that day" (taps pencil on the Monday column of the tin can row). Jade's explanation suggests that if four tin cans had been collected on one day in the week, it was not unreasonable to suggest that this quantity could be collected again.

All explanations for predictions made by the children as they worked in their groups were drawn from the existing data values in the table. This suggests that the range and frequency of the data provided strong evidence for the children's predictive reasoning (Reading, 2009). It was only when an explanation for the predicted values was requested by the teacher or researcher that knowledge from the picture story book was drawn to support a decision. The character's needs and likes were considered; for example, Bryce explained, "Because, because he ate apple cores", Sam explained, "Because, um, he um, he like-ed the tin cans so he collected from that day", and Eliot stated, "Er because he um likes to know what to do, so he just collects them”. When asked why he might have collected five newspapers, Jade explained, “Yep because um because he because I think he needed five of those”. Other explanations considered availability of items for the character Litterbug Doug to collect. Gina said, “Because um, um, Litterbug Doug just found them”, Isabel stated, “That’s because I’m thinking he collected them all in the dump”, and Carl suggested that only two cheeses were collected “because um he couldn’t find any more”. Eliot suggested that four apple cores were collected because “er, its popoolar, and he likes, he might like four”. These examples highlight that information gleaned from the picture story book influenced the knowledge children drew from to explain their predictions.

## Discussion

### *Using Data to Predict*

The development of the children’s models provided evidence of probabilistic reasoning, indicating that the task activated intuitions about chance (Watson, 2006). The study found the children used existing data, knowledge of the data context contextualised by the picture story book, and probabilistic reasoning to make predictions. The children’s predictions suggest that given the data provided for three days’ rubbish collection, it is possible that more, less or equivalent amounts could be predicted to be collected on the fourth day. This supports prior studies showing that probabilistic reasoning “consists of drawing conclusions about the likelihood of events based on available information or personal knowledge or beliefs” (Morsanyi, Primi, Chiesis, & Handley, 2009, p. 210). As the children did not have any previous formal instruction in chance or probability, the finding suggests that a solution to the prediction problem was found using emerging probabilistic intuitions and reasoning capacities and competencies drawn from individual experiences outside formal instruction.

All children used the available data to develop their prediction models. This finding supports the view that making predictions is about seeing relationships in the data that are separated from the event that created it and “using those relationships as a basis for making predictions about new cases” (Lehrer & Schauble, 2002, p. 23). Further, the children drew exclusively from contextualised knowledge of the picture story book to explain their predicted values when asked. These findings contrast with prior studies where children’s attempts to explain when drawing inferences from data were often grounded in personal experiences and not the data or the data context. Pereira-Mendoza (1995) found that 7 year olds could interpret graphical information but could not use the information to make realistic predictions. Watson and Moritz’s (2001) study used pictographs to estimate missing values and included 6 year olds who were not able to provide responses that referred to the given data, were unwilling to predict because of insufficient information, and engaged personal knowledge to explain their reasoning. In contrast, the children in this study were able to make use of the available data values to predict a reasonable missing data value and draw from the context of the problem to explain their decisions. This indicates that task design that combines embedded prediction tasks that meaningfully contextualise the problem and require engagement with the structure of data tables can support young children’s entry into data prediction.

### *Task Design and Reading the Data*

The findings suggest that the task design, where data were provided in a table, influenced probabilistic reasoning by supporting the children to “read the data”, “read between the data" and "read beyond the data", descriptions of levels for graph reading described by Curcio (1987). As the children were able to track rows and columns and isolate individual values associated with particular items and locate information, the structure of the table, with picture labels of categories heading lists of values in a row, was a format that made the information accessible and able to be comprehended (Friel, Curcio & Bright, 2001). Providing the data in this form contrasts with previous studies of prediction tasks with young children that employed bar graphs or pictographs. The finding in this study suggests that the children's predictions were supported by being able to engage in a literal reading of the data and had developed knowledge of the form and structure of a data table to do this. Knowledge of conventions for representing content supports comprehension and the ability to predict (Curcio, 1987).

The prediction decisions also suggest that the children were able to "read between the data" (Curcio, 1987, p. 384); that is, to interpret, integrate and find relationships in the information available in the data table (Curcio, 1987). What is notable is that although prior studies have found children's prediction to rely on observed patterns in data (Watson & Moritz, 2001), the data table provided to the children did not have numerical patterns that may have assisted seeing or forming such connections. Further, the children's predictions showed replication of values found in the row for each object, and a reasonableness of the range, revealing some understanding or intuitions of reasonable distributional variation, even in the absence of pattern in the data. This again suggests that existing data were taken into account (Leavy, 2008).

When predicting, the children were able to draw conclusions about the data and generalise beyond them using data to support their decisions, suggesting the inclusion of elements of statistical inference (Makar & Rubin, 2009; Reading, 2009). It is not suggested here that the children's data based explanations fulfill the requirements for informal inferential reasoning as the children's explanations did not impliedly or implicitly acknowledge uncertainty (Watson & Neal, 2012). The children's visible reasoning, however, suggests that existing data values were salient for prediction. The children's models also showed reasonable predicted values, given the existing data. These results suggest the children's engagement with aspects of variation and distribution which are building blocks for informal inferential reasoning (Reading, 2009).

## Conclusion

The study found that the children had intuitive knowledge of representation conventions for data tables and intuitive appreciation for variation and probabilistic intuitions. The children engaged probabilistic reasoning to make judgments about the likelihood of events using available data. Accordingly, inductive reasoning was employed as the children encountered statistical concepts, such as encountering variation in the data that triggered a need to manage uncertainty. The children's explanations revealed the use of data and data context knowledge for making prediction decisions and reasoning with data to answer questions.

The task design supported eliciting children's intuitions about probability to form meaningful predictions from data. The children were not asked to identify the probability of outcomes of the event, but to make a prediction about a possible outcome, which they chose to express as a data value. Children borrow from the experiences and concepts they have, including their messy everyday experiences with chance (Greer, 2001) and their propensity for causal reasoning. The finding suggests that task design features that tapped into natural frequencies of an event and the human context of behaviour supported children to use the data they were provided with to problem solve.

The children's explanations revealed important information about the contextual basis for the prediction decisions they made. The children's focus on data based reasoning when making prediction decisions indicates that although children have a propensity to attribute causal effects or deterministic modes of reasoning to chance (Langrall & Mooney, 2005), this was not the children's immediate explanatory response to solving the prediction problem. In addition, non-data explanations revealed that the picture story book was a source of knowledge the children drew on to account for what they observed in the data. This finding is consistent with research that finds that analysis and interpretation of data are dependent on interaction with contextual knowledge (Langrall, Nisbet, Mooney, & Jansem, 2011), and that discovering meaning in data requires conjecturing about the context of the problem based on the data. The significance of the children's use of the data context is that it suggests that children have the capacity and ability to draw meaningfully from data context knowledge to explain data observations, if the connection to the data context source is meaningful. These findings demonstrate that children's intuitive probabilistic reasoning competencies are significantly underestimated in research and curriculum expectations.

## References

Australian Curriculum, Assessment and Reporting Authority [ACARA]. (2013). *The Australian Curriculum: Mathematics, Version 4.1*, 31 January 2013. Sydney, NSW: Author.

Asp, G., Dowsey, J., & Hollingsworth, H. (1994). Students' understanding of pictographs and bar graphs. In G. Bell, B. Wright, N. Leeson, & G. Geeke (Eds.), *Challenges in mathematics education.* (Proceedings of the 17th Annual Conference of the Mathematics Education Research Group of Australasia, pp. 57-65). Lismore, NSW: MERGA.

Ben-Zvi, D., & Garfield, J. (2004). Statistical literacy, reasoning and thinking: Goals, definitions, and challenges. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 3-16). Dordrecht, Netherlands: Kluwer Academic Publishers.

Bethel, E. (2009). *Litterbug Doug*. Mascot, Australia: Koala Books.

Curcio, F. (1987). Comprehension of mathematical relationships expressed in graphs. *Journal for Research in Mathematics education, 18*(5), 382-393.

English, L. D. (2012). Data modelling with first grade students. *Educational Studies in Mathematics, 81*(1), 15-30. doi: 10.1007/s10649-011-9377-3

Franklin, C., Kader, G., Mewborn, D. S., Moreno, J., Peck, R., Perry, M., & Scheaffer, R. (2007). *Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report: A pre-K-12 Curriculum Framework*. Alexandria, VA: American Statistical Association.

Fischbein, E., & Schnarch, D. (1997). The evolution with age of probabilistic, intuitively based misconceptions. *Journal for Research in Mathematics Education*, *21*(1), 96-105.

Friel, S. N., Curcio, F. R., & Bright, G. W. (2001). Making sense of graphs. *Journal for Research in Mathematics Education, 32*(2), 124-158.

Greer, B. (2001). Understanding probabilistic thinking: The legacy of Efraim Fischbein. *Educational Studies in Mathematics, 45*, 15-33. doi: 10.1023/A:1013801623755

Langrall, C. W., & Mooney, E. S. (2005). Characteristics of elementary school students' probabilistic reasoning. In G. Jones (Ed.), *Exploring probability in school: Challenges for teaching and learning* (pp. 95-119). Dordecht, The Netherlands: Kluwer Academic Publishers.

Langrall, C., Nisbet, S., Mooney, E., & Jansem, S. (2011). The role of context expertise when comparing data. *Mathematical Thinking and Learning 13*(1&2), 47-67. doi: 10.1080/10986065.2011.538620

Leavy, A. (2008). An examination of the role of statistical investigation in supporting the development of young children's statistical reasoning. In O. N. Saracho & B. Spodek (Eds.). *Contemporary perspectives on mathematics in early childhood education* (pp. 215-232). Charlotte, NC: Information Age Publishing.

Lehrer, R., Kim, M-J., & Schauble, L. (2007). Supporting the development of conceptions of statistics by engaging students in measuring and modelling variety. *International Journal of Computers for Mathematical Learning, 12,* 195-216. doi: 10.1007/s10758-007-9122-2.

Lehrer, R., & Schauble, L. (2002). Children's work with data. In R. Lehrer & L. Schauble (Eds.), *Investigating real data in the classroom: Expanding children's understanding of math and science* (pp. 1-26). New York: Teachers College Press.

Lesh, R., & Doerr, H. M. (2003). Foundations of a models and modeling perspective on mathematics teaching, learning and problem solving. In R. Lesh & H. M. Doerr (Eds.), *Beyond constructivism: Models and modeling perspectives on mathematics problem solving, learning and teaching* (pp. 3-33). Mahwah: Lawrence Erlbaum Associates.

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

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

Masnick, A. M., Klahr, D., & Morris, B. J. (2007). Separating signal from noise: Children's understanding of error and variability in experimental outcomes. In M. C. Lovett & P. Shah (Eds.), *Thinking with data* (pp. 3-26). New York: Lawrence Erlbaum Associates.

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

Morsanyi, K., Primi, C., Chiesis, F., & Handley, S. (2009). The effects and side-effects of statistics education: Psychology students' (mis-)conceptions of probability. *Contemporary Educational Psychology, 34,* 210-220. doi:10.1016/j.cedpsych.2009.05.001

Mousoulides, N. G., & English, L. D. (2009). Kindergarten students' understanding of probability concepts. In M. Tzekaki, M. Kaldrimidou, & H. Sakonidis (Eds.), *In search for theories in mathematics education. Proceedings of the 33rd Conference of the International Group for the Psychology of Mathematics Education* (PME33, Thessaloniki, Greece). Retrieved from http://www.pme33.eu

Papaistodemou, E., & Meletiou-Mavrotheris, M. (2008). Developing young children's informal inference skills in data analysis. *Statistics Education Research Journal, 7*(2), 83-106.

Pereira-Mendoza, L. (1995). Graphing in the primary school: Algorithm versus comprehension. *Teaching Statistics, 17*(1), 2-6. doi: 10.1111/j.1467-9639.1995.tb00704.x

Reading, C. (2009). Cognitive development of informal inferential reasoning. *57th Session of the International Statistical Institute - Statistics: Our Past, Present & Future*, Durban, South Africa, 16th - 22nd August, 2009.

Reading, C., & Shaughnessy, J. M. (2004). Reasoning about variation. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 201-226). Dordecht, The Netherlands: Kluwer Academic Publishers.

Schwartz, D. L., & Goldman, S. R. (1996). Why people are not like marbles in an urn: An effect of context on statistical reasoning. *Applied Cognitive Psychology, 10*, S99-S112.

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

Watson, J. M., & Kelly, B. A. (2002). Emerging concepts in chance and data. *The Australian Journal of Early Childhood, 27*(4), 24-28.

Watson, J. M., & Moritz, J. B. (2001). The role of cognitive conflict in developing students' understanding of chance measurement. In J. Bobis, B. Perry, & M. Mitchelmore (Eds.), *Numeracy and beyond.* (Proceedings of the 24th Annual Conference of the Mathematics Education Research Group of Australasia, pp. 523-530). Sydney, NSW: MERGA.

Watson, J. M., & Neal, D. (2012). Preparing students for decision-making in the 21st century: Statistics and probability in the Australian Curriculum: Mathematics. In B. Atweh, M. Goos, R. Jorgensen, & D. Siemon (Eds.)*, Engaging the Australian National Curriculum: mathematics: Perspectives from the field* (pp. 89-113). Online publication: Mathematics Education Research Group of Australasia.

Wild, C., & Pfannkuch, M. (1999). What is statistical thinking? *Proceedings of the 5th International Conference on Teaching Statistics (ICOTS5, Singapore).* Retrieved from http://www.stat.auckland.ac.nz/~iase/publications/2/Topic3c.pdf