SECOND-GRADERS' PREDICTIVE REASONING STRATEGIES

Gabrielle Oslington

Macquarie University, Sydney, Australia

This paper reports predictive reasoning strategies used by ten second-graders in a classroom design study. A modelling activity based upon real data required students to predict maximum monthly temperatures for the current year using the natural variation provided by readings from the previous six years. The development of reasoning strategies was documented throughout the lesson sequence by analysis of responses to written prompts, videos of interviews and student drawn graphs. Student predictions reflected an emerging understanding variability, clusters and mean. Reasoning strategies became increasingly sophisticated using TinkerPlots, and with repeated opportunities for students to observe, represent, reflect upon trends in data.

STATISTICAL UNDERSTANDING IN THE ELEMENTARY YEARS

Recognising variability, grouping data according to attributes, and creation and representation of data sets are key competencies required for the development of statistical literacy in young students (English, 2012). However, it is widely understood that the development of statistical reasoning is a complex and not necessarily linear process, which is heavily reliant upon students' real-life experiences. Recent research suggests that young students draw upon personal context as well as the actual data values (Ben-Zvi & Aridor-Berger, 2016). Using data sets as evidence from which to make inferences proves a challenging process for young students, with many typically over generalising, or relying upon data such as very small sample sizes (Makar, 2016). Studies conducted in inquiry-based classrooms, where students have the opportunity to develop statistical investigations in a low stakes environment, allows exposure to data analysis strategies prior to formal instruction in statistics (English, 2012). By intentionally providing ambigious predictive tasks with multiple possible solutions, students can engage in meaningful statistical investigations and form predictions through a process of making sense of the information provided (Makar, Bakker, & Ben-Zvi , 2011).

Traditionally, predictive tasks for young students have centred on random response generators such as dice and spinners (Falk, Yedilevich-Assouline, & Elstein, 2012). However, prediction in 'real life' usually encompass both a degree of randomness alongside predictable, causal variation. Exposure to complex modelling activities including both these elements provides opportunities for developing reasoning skills in young students, alongside challenging their deterministic thinking and encouraging their observations of long-term trends. Examples of such tasks include modelling plant growth (Lehrer & Schauble, 2004; Mulligan, 2015) making predictions from picture books (English, 2012), examining first-graders' shoe size (Makar, 2016), and using

2018. In E. Bergqvist, M. Österholm, C. Granberg, & L. Sumpter (Eds.). *Proceedings of the 42nd Conference of the International Group for the Psychology of Mathematics Education* (Vol. 3, pp. 435-442). Umeå, Sweden: PME.

children's self-portraits to predict the age of the artist (Oslington, Mulligan, & Van Bergen, 2018). Natural phenomena such as weather and tides also provide rich opportunities for predicting and modelling variability (English, Fox, & Watters, 2005).

Konold and Pollatsek (2002) argue that data analysis from a student's perspective is fundamentally one of recognising a signal in a noisy environment, and this concept is accessible to students perhaps as early as the age of eight. The visual representations provided by TinkerPlots (Konold & Miller, 2005) software—an analytical tool designed to support students statistical reasoning —allows students to recognise both signal and noise for a given population. Through physically creating representations through drag and pull manipulations the software supports the development of informal statistical inference in the student. In the present study TinkerPlots was utilised with young children.

PURPOSE OF STUDY

The study reported is drawn from a larger design study incorporating four, multi-lesson instructional sequences examining students' development of reasoning skills and generalisation using TinkerPlots. The analysis presented here examines students' interpretation of a two-way table of data, and the strategies used for predicting missing values. The research questions were:

1. How do students use variability in given values to predict an unknown value?

2. What strategies are students using to justify their data choice?

3. How are students' interpretation of data reflected through their use of probabilistic language?

METHOD

Student participants

High achievers were selected from a cohort of 42 mixed ability second-graders attending an independent school in Sydney, Australia, through an assessment-based interview focused on pattern and structure (Mulligan, Mitchelmore, & Stephanou, 2015). Ten students (mean age 7 years 10 months) were withdrawn from the classroom for four weekly lessons using a design-based approach (Gravemeijer & van Eerde, 2009) with the researcher acting as teacher. The students had had previously exposure to TinkerPlots and were familiar with saving files, entering data and creating plots.

Design and Procedure

For the four-lesson sequence described here, students used variable natural data in the form of maximum temperature readings sourced from the Australian Bureau of Meteorology as a scaffold for predicting, representing and explaining their understanding of variability. During Lesson 1 students were provided with a data table containing maximum monthly Sydney temperatures for 2010-2016. A final row, labelled 2017 was left blank. Students worked in teacher-selected pairs to determine preferred values, with one pair of students videoed during the process. Once completed, they responded to the prompt "Write down anything you noticed about the numbers" and then individually represented the table data through tables or drawings. After completion of the representations, eight students participated in semi-structured video interviews explaining their representations. In Lesson 2, students responded to three written prompts "Is there anything you noticed about the data?" "How did you decide to choose these numbers?" and "How certain are you that the numbers you have chosen are accurate?" Students were then shown a graph of the data in TinkerPlots, and responded to the prompt "What patterns do you notice?" In Lesson 3, students added their predicted values to the Lesson 2 TinkerPlots graph. In Lesson 4 students used their own TinkerPlots graphs to respond to the following prompts "What do you notice about your data?", "Do you want to change any of your numbers?" and "Why or why not?" Nine students then participated in individual semi-structured interviews. At all points within the lesson cycle, students were free to reflect upon and refine their data choices, thus providing the opportunity for the development of their own explicit mathematical interpretation (Mulligan, 2015).

Data sources and analysis

There were three data sources: student-constructed data sets and responses to written prompts (**written)**, transcripts (**oral)** from Lesson 1 and 4 semi-structured interviews and the transcripts of two videoed interviews in Lesson 1, and **representational** (hand-drawn graphs from Lesson 1 and TinkerPlots graphs from Lesson 3). In this preliminary report, an overview of emerging reasoning for each student is described as three levels of statistical reasoning: idiosyncratic, transitional and quantitative (Leavy, 2008).

Student-constructed data sets were coded on a three-point scale according to awareness of seasonal patterns and proximity of predictions to existing data set (no awareness, some awareness, close approximation). Student responses to written prompts (Lessons 1, 2a 2b and 4) were coded on a five-point scale of increasing generalised thinking, (nil/idiosyncratic response, individual scores observations only, observing differences between months, generalised comment about shape of data, generalised comment combined with awareness of mean, range or outliers). Oral responses from Lesson 1 were coded as for written prompts. Thematic coding was developed from the transcript of the videos from Lesson 1 and the semi-structured interviews in Lesson 4 (Flick, 2014). Hand-drawn graphs were coded according to levels of structural development, namely pre-structural, emergent, partial, structural and advanced structural (Mulligan, Hodge, Mitchelmore, & English, 2013). TinkerPlots graphs mirrored the data tables, and the same coding applied. Given the diversity of data sources, a range of analysis was used for trial purposes, with preliminary results only presented here.

RESULTS

Student predictions

The monthly predictions created by the five student pairs all showed at least some awareness of seasonality and the idea of range (Table 1). For pairs 1, 3 and 5, the seasonal pattern was distinct and the predicted values fell within or very close to the range of the data. Pair 2 showed a slight seasonal dip, however the predictions were well outside the range, by as much as 10ºC in the case of July. Similarly, Pair 4's winter dip was only apparent in June and August with other values outside the range.

Pair	Jan	Feb	Mar	April	May	Jun	July	Aug	Sept	Oct	Nov	Dec
1	43	40	36	32	29	25	27	28	31	35	38	40
2	45	40	39	35	32	31	36	42	42	43	43	45
3	38	36	29	26	23	21	24	27	30	36	39	41
4	47	41	31	30	27	24	29	20	32	39	41	42
5	35	41	30	31	26	22	21	22	24	27	32	39

Table 1: Student pairs predict monthly maximum temperatures for 2017

In their responses to "Write down anything you notice" (Lesson 1) six students referred to data shape (Pairs 1, 5) and identification of seasons (Pair 3) thus interpreting the table as representing one unit of information. In contrast, Pair 2 focused on individual data points. For example, "The hottest numbers were 46, 41, 42, 39 and 38" (Ashley) and "The hottest temperatures were all forties and thirties" (Julian). This difference in understanding was apparent in their representations: Fritz (Pair 1) and Lanni and Rhys (Pair 2) produced structural or advanced structural representations, showing awareness of equal spacing, partitioning, structured counting, shape and alignment and sequencing as described by Mulligan et. al. (2013). Rhys and Fritz's representations were presented as line graphs, showing their understanding of the data as a continuous sequence. Although these students were experienced in reading two-way tables, this task required interpretation of 72 data points, which could be viewed in rows (years) or columns (months), as well as a global set of repeating cycles. Most student pairs 'read' the table from top to bottom, with emphasis upon seasonality, using range, proximity to other values and seasonal knowledge to assist. Julian and Ashley, in contrast, highlighted individual data points with Julian tabulating each temperature, with the number of occurrences written below it and Ashley selecting and colour coding temperatures for each year. When prompted at interview, Julian was unable to identify seasonal patterns. Ashley focused upon rows (i.e. calendar years) rather than columns (months) of data, stating that "in 2016 there were two hot temperatures".

Students' reasoning behind predictions

All three data sources show students used personal experience, knowledge of seasons and the data table to make predictions. Although not all students could correctly name the months belonging to each season, these could be deduced from the table. Joseph for example, explained "Well, I found out that were (sic) the season are, so in the middle it is winter and then it heats up on the sides". Students' drew on memories of hot days of the year, weather reports and specific events, such as birthdays to link months and temperatures. Completing the table on a rather cold day at the end of August was complicated by students experiencing a relative heatwave the previous week. Rose wrote: "2017 was a very hot year. Winter felt like summer."

The most powerful predictor for most students was the data itself. Students' used the table to demonstrate an understanding of variability and emerging generalisations. Joseph and Stuart (Pair 3) reflected explicitly about their process: "I started with the middle and ceep (sic) going up by 2-3 degrees and on the left side going down by 3-2 degree" (Joseph) and "I started at June which divides the year in half, because I knew it would be cold then I did January and December which would be the hottest and I went in" (Stuart). Caden and Aaron (Pair 4) both describe selecting numbers around the other numbers. The transcript from Pair 5 reveals explicit discussion of range (March and April), months being "all in the 20s" (May), November as "like summer", and June being cold. This pair then observed an annual cycle by adjusting their December figure in line with the one they had produced for January.

Use of TinkerPlots

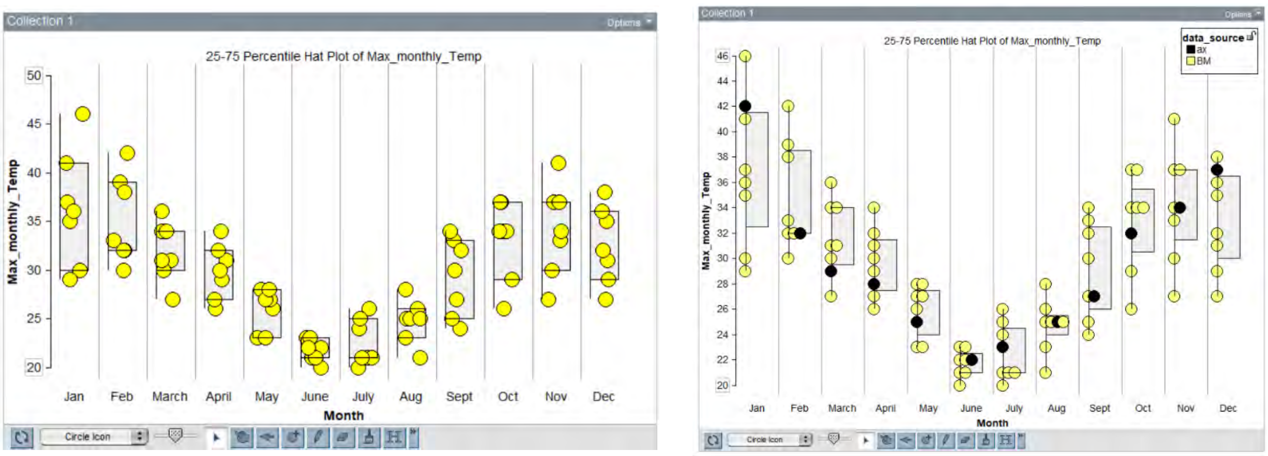

Figure 1 Maximum Sydney temperatures plotted using TinkerPlots software (left) and values including a student's estimations (right)

Transferring the data from a table to TinkerPlots format (Figure 1) assisted all students to recognise seasonality, and to display increasingly sophisticated statistical reasoning. Eight students interpreted the data spread as greater for warmer months: for example, Joseph responded, "It looks like the hotter the temperature is, the more spread out they are, like January or November, and the colder it is the more together they are like June". Caden described January as having a "big range, while June has a small, tight range". When working only from the table, Aaron wrote: "I notice no patterns", but after viewing the data graphically (Figure 3a) his response was "I see January is high and it gets lower and higher again." When compared with Leavy's (2008) three levels of statistical reasoning, eight of the ten students' explanations moved to a higher level after using TinkerPlots and five were at the third (quantitative) level, usually characteristic of students in grades 3 and above. The students' familiarity with the software potentially gave them access to concepts such as range and average, previously inaccessible for grade 2 students (Table 2).

Lesson	Fritz	Rose	Julian	Ashley	Joe	Stuart	Aaron	Cadel	Rhys	Lanni
1	trans	trans	trans	idio	trans	quan	trans	trans	trans	trans
2a	idio	idio	trans	idio	idio	idio	idio	trans	#	#
2b	quan	trans	quan	trans	quan	trans	quan	quan	trans	idio
4	quan	quan	quan	idio	quan	trans	trans	trans	trans	#

Table 2: Levels of second-graders statistical reasoning over four lesson sequence: idiosyncratic, transitional and quantitative # missing data

After plotting their own data against the table temperature (see sample Figure 2), students justified their data choice. Student explanations included "reasonableness" (Fritz), global warming (Julian), match to TinkerPlots data (six students), data located in the 'hat" (Fritz, Rose, Joseph), similarity to table figures (Rose, Caden) and seasonality (Stuart). When discussing their data, students either directly used or implied the following terms: outliers, clusters, spread, range, hats and quartiles, data shape, levels of certainty of prediction and discrepancies between their chosen values and the provided data table (Table 3).

Lesson	Fritz	Rose	Julian	Joe	Stuart	Aaron	Cadel	Rhys
Outliers	*					*		
Clusters	*	*				*		
Spread	*						*	*
Range	*						*	*
Hats	*	*	*	*	*			*
Data shape					*			*
Uncertainty	*		*	*	*			*
Certainty		*		*	*	*		
Discrepancies			*				*	

Table 3: Terms expressed by second-grade students when describing data variability

DISCUSSION

The task described here required students to predict temperatures for current and future dates by interpreting a two-way table data set. By using relevant data which involves natural variation, students could use their developing data-analysis strategies while simultaneously attending to other features relevant to them. These included recent experiences, knowledge of the seasons and weather reports. As in other studies (English, 2012), the data table itself formed the foundation for predictions, specifically student observations of maximums, minimums and range. When justifying their temperature figures, the TinkerPlots representation featured prominently in explanations, with students describing their data as "reasonable", "looking right" or located in the TinkerPlots hat. Previous studies (Ben-Zvi & Aridor-Berger, 2016) show that young students move between the content i.e. the provided data and their personal context when developing inferential reasoning skills. Early efforts at data modelling typically show a disconnect between the two worlds, with integration emerging over time and modelling experiences. In the study presented here, students integrated the content and the context appropriately. However, this process is still in development, as some students over generalised about climate change or linked a whole month from a single remembered day. Makar (2016) argues that repeated exposure to such predicting and organising activities provides a positive experience with informal statistics and a growing awareness of both variability and generalisation.

The language used by the students included specific statistical concepts such as hats, quartiles, spread, range, clusters and outliers. The complexity and ambiguity of the task along with visual accessibility provided by TinkerPlots, supported student reasoning at level unexpected for such young students. According to Makar et al. (2011) inferential reasoning requires a statement of generalisation beyond the data, using data to support the generalisation and recognition of uncertainty. Through their generalisations about temperature cycles and the use of a data table as a principal resource for predicting, and through balancing certainty and uncertainty, these students are well on the way to meeting these criteria. Prolonged and focused activities, to which these students had had previous exposure, greatly supported their statistical reasoning. Multiple studies have reported on the development of students' data analysis strategies in the early school years (English, 2012; Leavy, 2008; Mulligan, 2015, Oslington et al., 2018). This study extends this research by demonstrating the potential of TinkerPlots as a viable tool for developing data analysis strategies by students as young as the second-grade.

CONCLUSION

This study affirms previous research with young students demonstrating their capacity to utilise a data table to predict missing values and to use context to support their predictions. TinkerPlots allowed students to increase their level of statistical reasoning. This is a small, specialised study conducted with capable students, and thus the results cannot be generalised to all second-grade students. Nevertheless, they suggest that TinkerPlots can be used as a viable tool for data analysis by second-grade students, and the study confirms the feasibility of providing rich modelling tasks for developing statistical reasoning in young students.

References

Ben-Zvi, D., & Aridor-Berger, K. (2016). Children's wonder how to wander between data and context. In D. Ben-Zvi, & K. Makar, *The teaching and learning of statistics: International perspectives* (pp. 25-36). Springer .

English, L. (2012). Data modelling with first-grade students. *Educational Studies in Mathematics, 81*(1), 15-30.

English, L., Fox, J., & Watters, J. (2005). Problem posing and solving with mathematical modeling. *Teaching Children Mathematics, 12*(3), 156-163.

Falk, R., Yedilevich-Assouline, P., & Elstein, A. (2012). Children's concept of probability. *Educational Studies in Mathematics, 81*, 207-233.

Flick, U. (2014). *An introduction to qualitative research* (5 ed.). London, UK: Sage.

Gravemeijer, K., & van Eerde, D. (2009). Design research as a means for building a knowledge base for teachers and teaching in mathematics education. *The Elementary School Journal, 109*(5), 510-524.

Konold, C., & Miller, C. D. (2005). *TinkerPlots: Dynamic data exploration.* Emeryville, CA: Key Curriculum Press.

Konold, C., & Pollatsek, A. (2002). Data anaylsis as the search for signals in a noisy process. *Journal for Research in Mathematics Education, 33*(4), 259-289.

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, *Contemporary perspectives on mathematics in early childhood education* (pp. 215-232). Charlotte, NC: Information Age Publishing.

Lehrer, R., & Schauble, L. (2004). Modeling natural variation through distribution. *American Educational Research Journal, 41*(3), 635-679.

Makar, K. (2016). Developing young children’s emergent inferential practices in statistics. *Mathematical Thinking and Learning, 18*(1), 1-24.

Makar, K., Bakker, A., & Ben-Zvi , D. (2011). The reasoning behind informal statistical inference. *Mathematical Thinking and Learning, 13*(1-2), 152-173.

Mulligan, J. (2015). Moving beyond basic numeracy: data modeling in the early years of schooling. *ZDM Mathematics Education, 47*, 653-663.

Mulligan, J., Hodge, K., Mitchelmore, M., & English, L. (2013). Tracking structural development through data modelling in highly able grade 1 students. In V. Steinle, L. Ball, & C. Bardini (Eds.), *Proceeding of the 36th Annual Conference of the Mathematics Education Research Group of Australasia* (pp. 530-536). Melbourne, Vic: MERGA.

Mulligan, J., Mitchelmore, M., & Stephanou, A. (2015). *Pattern and Structure Assessment: an assessment program for early mathematics.* Camberwell, Victoria: ACER Press.

Oslington, G., Mulligan, J., & Van Bergen, P. (2018). Young children's reasoning through data exploration. In V. Kinnear, M. Y. Lai, & T. Muir, *Forging connections in early mathematics teaching and learning.* Springer.