SUPPORTING YOUNG STUDENTS EMERGING UNDERSTANDINGS OF CENTRE THROUGH MODELLING

Jill Fielding-Wells and Judith Hillman
Learning Sciences Institute of Australia, Australian Catholic University
Education Queensland
jill.wells@acu.edu.au

'Centre' is a key conceptual understanding that students require to think and reason statistically. In this study, a class of young students (aged 9-10) who had predominantly procedural experience with statistics, was posed a complex problem regarding the best design for a catapult aeroplane. The context established an opportunity for students to engage authentically with data modelling. Analysis of lesson videotape and work samples provided insight into students' developing understandings as they worked with dot plots and hat plots to focus on the notion of 'middle'. Insights into students' thinking and reasoning are discussed as the students engage in constructing, interpreting and predicting from statistical models. This research has implications for statistics curriculum in the early formal years.

INTRODUCTION AND LITERATURE

Statistics at the primary/elementary school levels predominantly addresses graph construction and reading of a limited range of established graph formats: typically, bar/column graphs (including picture graphs) and pie charts. Common data used are favourite ice cream flavours or fruits, with data either pre-provided or collected from pre-populated options. The majority of student time is spent in constructing and colouring the graph before *reading* the graph to provide simplistic responses. These tasks are essentially procedural, resulting in students having little opportunity to develop conceptual understanding (Pfannkuch 2006). More accurately, this is not statistics but visual comprehension. Children begin to see graphs as illustrations rather than reasoning tools and key understandings, such as distribution, variation and centre, are overlooked. However, if young students were introduced to more authentic data uses, there is a likelihood that deeper understanding could be developed early, facilitating increasingly sophisticated understanding which would carry into later years. If students do not understand distribution, centre and variation, they cannot identify measures of centre (signal) from amid variation (noise) and therefore will be unable to infer to a population, or to make between group comparisons (Konold & Pollatsek, 2002). '*Graphs as tools'* – as distinct from 'graphs as pictures' – (Wild & Pfannkuch, 1999), and '*signal among noise'* are key understandings that students need early in their statistics education. One potential means for students to develop more appropriate conceptions of the purpose and use of statistics is through the production of models and engaging in the modelling process. In this way, representations are seen and employed as tools that can facilitate understanding of several of the 'big ideas' of statistics: distribution, centre and variability.

This paper focuses not on whether modelling can be used with young children, but rather addresses and aims to illustrate affordances of modelling for focusing on centre

Children's Statistical Understandings

A fundamental concept for students in the initial stages of engaging with statistics is the realisation that data is needed: data can be used to help in making decisions and solving problems, and data can be collected or generated. Building from this are several key statistical ideas that all students need to understand at a deep conceptual level: distribution; centre; and variation (Garfield et al. 2008a; Watson 2006). Conceptualization of distribution is quite difficult (Garfield & Ben-Zvi, 2007) as reasoning about distributions involves interpreting a complex structure that not only includes reasoning about features such as centre, spread, density, skewness and outliers, but also involves appreciation of related concepts of sampling, population, and chance (Pfannkuch & Reading, 2006, p. 4). Distribution of data is fundamental to statistical reasoning, and students need opportunities to work with data representations in learning contexts that explicitly focus on centre, variation, and shape of a data display, and to see that the clusters, outliers and gaps are important.

In M. A. Sorto, A. White, & L. Guyot (Eds.), *Looking back, looking forward. Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10, July, 2018), Kyoto, Japan.* Voorburg, The Netherlands: International Statistical Institute. iase-web.org

When introducing notions of 'centre', students are often exposed to the mean and later median as a procedural exercise without first developing a conceptual understanding of 'centre' or spread. Hence students learn to calculate measures of centre without a clear conception of what these measures tell us about data (Garfield & Ben-Zvi, 2007; Garfield et al., 2008b) and have difficulty in understanding the difference between median and mean (Garfield et al., 2008b), much less understand the term 'average', given both its statistical and commonplace meanings (Watson, 2006). When working with young students, maintaining a focus on distributions and their shape, along with use of informal language such as clumps and hills, draws students' attention to the key features of *centre* and enables them to visualize data sets.

Modelling

A statistical model can be considered that which enables the location, explanation or extraction of underlying patterns in data (Graham, 2006) and which adopts statistical ways of representing and thinking about the real-world (Wild & Pfannkuch, 1999). Through opportunities to generate models, young students can be introduced to, and encouraged to explore, initial conceptions of distribution, centre and variation. It is important that students learn to see a data distribution as an aggregate; observing the pattern and shape, rather than considering individual cases (Rubin, Hammerman, & Konold, 2006) and the use of models may provide a basis for this. Having students develop their own models from a problem question with identifiable outcomes, may assist in guiding students to use the process of modelling to focus on the aggregate, or signal (Konold & Pollatsek, 2002).

METHOD

The research question being addressed in this study is: What insights emerge about young students' statistical thinking and reasoning about *centre* when students engage in modelling to address an authentic, complex problem?

Design-based research (DBR)(Cobb, Confrey, diSessa, Lehrer, & Schauble, 2003) was utilized for the broader study as design research is characterized through implementation, reflection, and ongoing adjustment of successive iterations of an intervention in the classroom context. DBR is typified by practitioner and researcher working together to plan learning on a lesson-by-lesson basis, reflecting on progress and making progressive adjustments in planning with the intent to improve learning sequences.

The students involved in the study comprised a class of 26, 9-10 year-old students from a suburban government school in Australia. The school has a high-average Index of Community Socio-Educational Advantage (ICSEA) score. The class teacher, Ms Thompson, is an experienced teacher with no specialist training in statistics.

Multiple sources of data were collected, including videos of each entire lesson, student work samples and field notes. Video analysis was carried out following a process adapted from Powell, Francisco, and Maher (2003). The videos were viewed and logged to develop an overall record of the unit. Episodes which were pivotal in enhancing or demonstrating student understanding were identified and transcribed in full and linked to relevant representations/models in use. While the unit of learning focussed on distribution, centre, and variation, the findings that related to students developing conceptions of centre are the focus of this paper.

RESULTS

The unit introduced students to an authentic context for which they needed to ascertain the best sized piece of paper for making catapult aeroplanes. The students initially selected two sizes of paper (10 x 15cm and 20 x 30cm) and constructed and trialled the aeroplanes. They were introduced to dot plots for recording the data and shown by the teacher how to bin their data. Students constructed dot plots for each paper size. A lengthy class discussion was held on the overall shape and distribution of the data, including outliers, before consideration of variation and centre. Throughout the discussion, the teacher drew students' attention to the representation in terms of the context to ensure students were making the link from the interpretation of the graph to the aeroplane flights. The teacher then introduced TinkerPlots (Konold & Miller, 2005) by showing the students their binned data in a TinkerPlots display before showing them the same data unbinned

(Figure 1) after deciding that unbinned data would provide the more realistic distribution. She drew the students back to the context by asking whether, from looking at the data, the students could tell her which was the best sized piece of paper from which to make a catapult aeroplane. In response, one student argued that the smaller aeroplane (top plot in Figure 1) must be the better because there is a clump around 7 to 9 metres whereas the larger aeroplane was more clumped around 4 to 6 metres [Janice]. However, the majority of students disagreed, arguing that the spread of the larger aeroplane (bottom plot in Figure 1) was such that there wasn't really a clump.

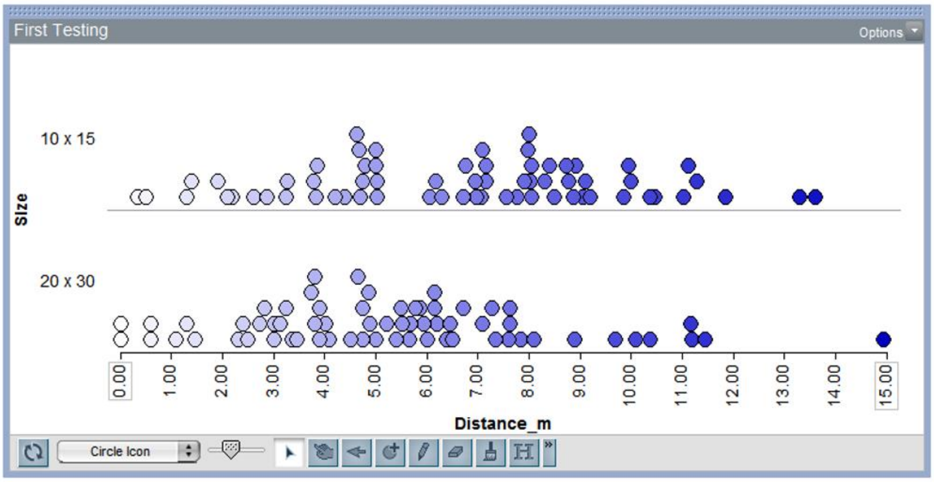

Fig. 1. TinkerPlots comparison of 10 x 15cm and 20 x 30cm flight test data

When the students were unable to make a decision, the teacher suggested that finding a *middle* might assist. One student was aware of the mean from the context of cricket batting averages, however, Ms Thompson steered him away from this suggestion and challenged the class to find the middle 50% of the data, handing out printed versions of Figure 1 with which to work. The students found this quite difficult due to limited knowledge of percentages. Most students related this to one-half and set about trying to identify the midpoint of the data to identify half-way to split the data into two halves. Janice was the first to work this out and explained the process to her group. Her representation (Figure 2) and explanation are below:

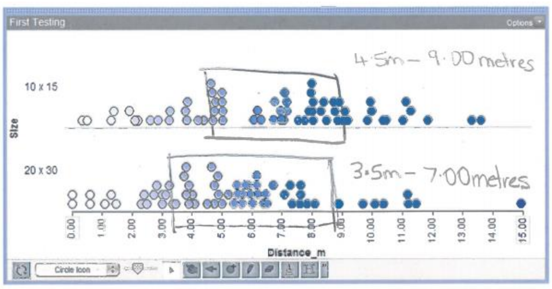

Fig. 2. Student identified centre (Janice)

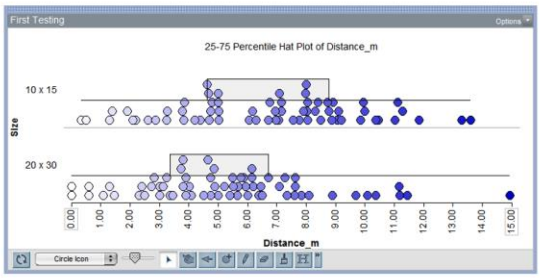

Fig 3. Comparing TinkerPlots Hats to students' hat plots

Janice	There is 65 altogether in this, of the data. And then but you need to halve 65 to get the middle which is 32.
Callum	32 and a half.
Janice	32 and a half. We're just going to work with 32. But then if we want to get, 32 plus 32 is 65 so we would need to have… half 32 is 16 so we would need to have 16 on each side to get that second 32.
Callum	But why do we need to get 32?
Isaac	Where did the 16 come from?

Isaac eventually comprehended what Janice was saying and proceeded to find one-quarter of the total number of data points, identifying the top quarter and the bottom quarter and then the two quarters in the middle as making the half or middle 50%. Janice marked where she thought the middle 50% was on the handout (Figure 2) and then summarizes her conclusion:

Janice So, I personally think 10 by 15 is better because it's, as you can see, the average is closer to the higher numbers. So, where the average for 20 by 30 [cm] is a lot lower than the 10 by 15 [cm]. Because it is a whole 2 metres further.

After Janice's explanation to the class, most students were able to locate and the middle 50% and began discussing the data in terms of centre rather than individual data points. The teacher then formally introduced the hat plots feature of TinkerPlots and encouraged students to compare the generated hat plot (Figure 3) with their own. The ordering of the activities was done to ensure the students could conceptually link the hat plot to their data rather than merely 'reading' the generated representation.

Round 2

The students concluded that, based on the data, the (10 x 15 cm) was the better aeroplane. The students then used this conclusion to inform their choices of paper size for their next iteration of testing. They also engaged in lengthy discussion about how they could control the variables to increase accuracy. The students decided that 6 x 9cm, 8 x 12cm, 10 x 15cm and 12 x 18cm would be the sizes to test. The results of the second investigation were recorded and entered into TinkerPlots and the students provided with the display shown in Figure 4. The data that resulted from this second round of testing enabled the students to identify the best aeroplane overall (the one with the highest 'hat'). The students were asked to provide a written justification of their response. One response, Grant's, is provided below as an example of a more articulate response:

The best sized piece of paper for a catapult plane would be the 12 x 18 plane because it has a middle 50% of 7.50m to approximately 10m. All the other planes had a smaller (lower) middle 50% because their ranges of distance are between 2m-4.50m (6 x 9 plane), 4m–7m (8 x 12 plane), and 6m to 7.50 m. We use the middle 50% because it is kind of like an average where half of the planes landed in a cluster and that is easy to read and easy to figure out. … Overall, with all the things to think about, like variables, outliers and the middle 50%. This has affected my decision, but I still choose the 12x18 paper plane because even though the 10x15 was more consistent, the 12x18 middle 50% (which gives you the range where the most planes landed) was the largest [highest] range and that gave me a good idea of what is usually thrown. [Grant]

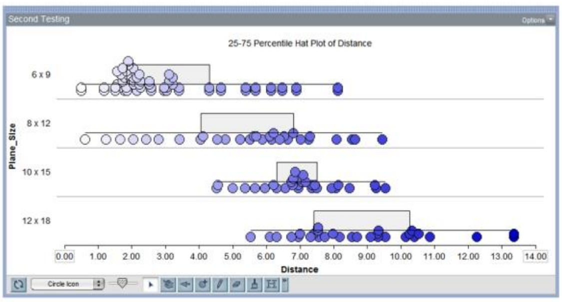

Fig. 4: Data from second round of testing comparing 6 x 9cm, 8 x 12cm, 10 x 15cm and 12 x 16cm paper sizes with hat plots

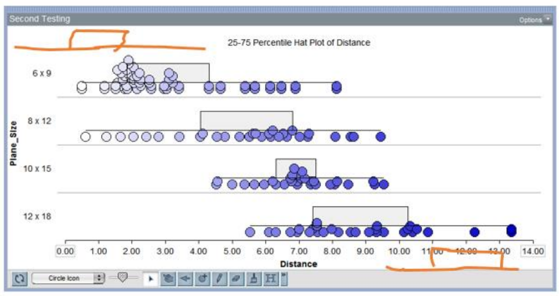

Fig. 5: Predicting the 'hats' for aeroplanes one size smaller and one size larger

To ensure that the students were individually interpreting the dot/hat plots, the teacher asked the students to make predictions about the likely outcomes of aeroplanes made from one size smaller and one size larger pieces of paper. The students' responses were varied, however, the majority showed hats that would follow from the existing pattern in the data representations. Cindy's response (one of the clearest for reproduction) has been provided to illustrate (Figure 5).

DISCUSSION AND CONCLUSION

The purpose of this paper was to illustrate possible affordances of statistical modelling for developing young students' statistical thinking and reasoning about centre in the context of an authentic, complex problem. Wild and Pfannkuch (1999) describe a statistical model as a way of representing and thinking about the real world. In this paper, students were afforded opportunities to represent the outcomes of the tests of various sizes of paper aircraft to make a decision about the best size. In generating their own data, for an authentic purpose, students constructed and interpreted representations based on data they could meaningfully connect to its source. In making decisions by drawing on the representations, students used their representation as a tool, or *model*, to make predictions about possible outcomes of further trials, drawing on patterns they identified in these models.

Previously noted tendencies for students to focus on individual data scores rather than aggregates or patterns (Wild, 2006) were observed in this research. However, the teacher, after giving the students the opportunity to observe and discuss individual scores, focussed students on 'patterns' and 'shape' in the data displays to draw students' attention to centre or signal. In considering where the 'clumps' of data lay; the students focussed on what was 'typical' in the data. In the first instance, the data was quite spread and so the students could not find a way to identify a useful difference between the two data sets. The teacher's challenge to find the middle 50% shifted the representation from requiring *reading* to requiring manipulation in order to be useful. Once the students had worked through the process of finding the middle 50%, they were able to quickly identify differences between two data spreads. It was this need to separate the signal from the noise (Konold & Pollatsek, 2002) that served as an authentic prompt for the teacher to introduce the idea of *middle 50%* and then hat plots. The students thus had a genuine need for calculating the hat and, because the first iteration had them plotting and calculating by hand, they were able to make links between the hats and the data sources, thereby facilitating interpretation. In this, the importance of having students work with data that they have sourced and represented – in response to a purposeful question - is foregrounded. The purposeful question provided not only a meaning for the data but a need to represent it and draw on statistical measures to interpret it. While students were conceptually recognizing the need to separate the centre from the variation, they were not introduced to the terms signal and noise. In hindsight, these may have been useful terms to establish.

The affordances of TinkerPlots were considerable in this activity. Once the students had worked out how to create dot plots and find the hat by hand, they were introduced to TinkerPlots. The ability to have the hats illustrated saved considerable time, which the students appreciated, but also facilitated multiple manipulations. In this respect the model, in combination with the affordances of TinkerPlots, overcame the 'graph as illustration' perceptions so often noted by students (Wild & Pfannkuch, 1999).

When students are introduced to measures of centre in primary schooling, this is often approached as a formulaic measure only through mean, median and mode. Representing centre as a single digit serves to mask distribution and thus limits development of students' conceptions of centre. By focussing on the middle 50% the teacher was maintaining the visual aspects of the data distribution and the relationship between the centre and distribution was clear and able to be visualised. In this research, it appeared that adopting a modelling approach that maintained the visual distribution facilitated conceptual development and offered far more potential to develop statistical thinking and reasoning than a formulaic procedural approach. By working with distributions, children become more familiar with a variety of data 'shapes', appreciating what the shape of the data is saying. Having such an appreciation before formal, procedural methods are introduced in essential.

The modelling activity undertaken here served to focus students on key understandings of centre – supporting the contention that young children should address variation and centre through modelling at early ages (Garfield et al., 2008c) by demonstrating that the capacity to do so is well within their reach. Perhaps even more heartening is that this teacher had no special training in statistics or mathematics beyond that of a generalist primary teacher.

ACKNOWLEDGEMENTS

This work was supported by funding from the Australian Research Council under DP170101993. The authors wish to gratefully acknowledge the contributions of the teacher and students engaged in this research.

REFERENCES

Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher, 32*(1), 9-13.

Garfield, J., and Ben-Zvi, D. (2007). How students learn statistics revisited: A current review of research on teaching and learning statistics. *International Statistical Review / Revue Internationale de Statistique, 75*(3), 372-396.

Garfield, J., Ben-Zvi, D., Chance, B., Medina, E., Roseth, C., and Zieffler, A. (2008a). Creating a statistical reasoning learning environment. In J. Garfield and D. Ben-Zvi (Eds.), *Developing students' statistical reasoning: connecting research and teaching practice* (pp. 45-63). Dordrecht: Springer Netherlands.

Garfield, J., Ben-Zvi, D., Chance, B., Medina, E., Roseth, C., and Zieffler, A. (2008b). Learning to reason about centre. In J. Garfield and D. Ben-Zvi (Eds.), *Developing students' statistical reasoning: connecting research and teaching practice* (pp. 187-200). Dordrecht: Springer Netherlands.

Garfield, J., Ben-Zvi, D., Chance, B., Medina, E., Roseth, C., and Zieffler, A. (2008c). Learning to reason about variability. In J. Garfield and D. Ben-Zvi (Eds.), *Developing students' statistical reasoning: Connecting research and teaching practice* (pp. 201-214). Dordrecht: Springer Netherlands.

Graham, A. (2006). *Developing thinking in statistics*. London: Paul Chapman Publishing.

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

Konold, C., and Pollatsek, A. (2002). Data analysis as the search for signals in noisy processes. *Journal for Research in Mathematics Education, 33*(4), 259-289.

Pfannkuch, M. (2006). Comparing box plot distributions: A teacher's reasoning. *Statistics Education Research Journal, 5*(2), 27-45.

Pfannkuch, M., and Reading, C. (2006). Reasoning about distribution: A complex process. *Statistics Education Research Journal, 5*(2), 4-9.

Powell, A. B., Francisco, J. M., and Maher, C. A. (2003). An analytical model for studying the development of learners' mathematical ideas and reasoning using videotape data. *Journal of Mathematical Behaviour, 22*(4), 405-435.

Rubin, A., Hammerman, J. K. L., and Konold, C. (2006). Exploring informal inference with interactive visualization software. Paper presented at the *Seventh International Conference of Teaching Statistics*, Salvador, Brazil.

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

Wild, C. J. (2006). The concept of distribution. *Statistics Education Research Journal, 5*(2), 10-26

Wild, C. J. and Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review 67*(3), 223-265.