Beginning Inference in Fourth Grade: Exploring Variation in Measurement

Lyn English
Queensland University of Technology
<l.english@qut.edu.au>

Jane Watson
University of Tasmania
<jane.watson@utas.edu.au>

This paper addresses one of the foundational components of beginning inference, namely variation, with 5 classes of Year 4 students undertaking a measurement activity using scaled instruments in two contexts: all students measuring one person's arm span and recording the values obtained, and each student having his/her own arm span measured and recorded. The results included documentation of students' explicit appreciation of the variety of ways in which variation can occur, including outliers, and their ability to create and describe valid representations of their data.

Numerous curriculum and policy documents highlight the importance of children working mathematically and scientifically in dealing with real-world data in the primary school years (e.g., Curriculum Corporation, 2006; National Council of Teachers of Mathematics, 2006). Limited attention however, is given to the statistical literacy that children need generally for decision-making in the 21st century. For our young students to become statistically literate citizens, they need to be introduced early to the powerful mathematical and scientific ideas and processes that underlie this literacy (e.g., English, 2012; Whitin & Whitin, 2011). Various definitions of statistical literacy abound (e.g., Gal, 2002; Watson, 2006). Watson's definition provides a comprehensive foundation: "the meeting point" of statistics and probability and "the everyday world, where encounters involve unrehearsed contexts and spontaneous decision-making based on the ability to apply statistical tools, general contextual knowledge, and critical literacy skills" (p. 11).

An important, yet underrepresented component of statistical literacy is beginning inference, which includes the basic components of variation, prediction, hypothesising, and criticising (English, 2010; Shaughnessy, 2006; Watson, 2006). Makar and Rubin (2009) identify three core components of beginning inference: generalising beyond the data, using data as evidence, and acknowledging uncertainty in the conclusion. As we start young students along this path we do not expect that all three aspects will be absorbed immediately. We begin with issues of variation within and between groups. Uncertainty is the second component picked up as variation influences the certainty with which one can make a decision. Finally, it is expected that students are able to generalise further than the data set at hand, with questions of reliability and validity that arise in the process. There is little research on beginning inference with young students. Our focus here is the foundational component of variation, as it occurs within a measurement activity.

Variation

Variation lies at the heart of statistical reasoning and is linked to all aspects of statistical investigations (Cobb & Moore, 1997; Garfield & Ben-Zvi, 2007; Konold & Pollatsek, 2002; Watson, 2006). Indeed, as Watson 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 understanding of variability is essential in the development of children's statistical literacy. This understanding should be integrated, revisited, and emphasised in statistics learning

In V. Steinle, L. Ball & C. Bardini (Eds.), Mathematics education: Yesterday, today and tomorrow (Proceedings of the 36th annual conference of the Mathematics Education Research Group of Australasia). Melbourne, VIC: MERGA.
© Mathematics Education Research Group of Australasia Inc. 2013

from 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). Explicit discussion on variation is needed throughout the primary school years, before students meet formal measures such as standard deviation in the secondary school years.

The research on young children's reasoning with variability and variation is limited; but Watson (2005) has found that young students do have a primitive understanding of these concepts. A good deal more research is needed on the nature of this understanding and effective ways to develop it in the primary school years. To this end, the study reported here introduced Year 4 students to experiences with variation, with the first activity engaging students in taking arm span measurements using scaled instruments in two contexts. The difference in variation for the two data sets and ways of representing this served to illustrate measurement "error" and measurement "approximation." Specifically, the research objective for this report relates to how children perceived variation in their measurement values, their identification of unusual values, how they represented and interpreted the values, and their assessment of measurement accuracy within and between contexts. Prior to addressing the study, we give brief consideration to perspectives on measurement and its links to other mathematics content areas.

Measurement

In their review of the geometry and measurement strand of the *Australian Curriculum: Mathematics* (Australian Curriculum, Assessment and Reporting Authority [ACARA], 2012), Lowrie, Logan, and Scriven (2012) lamented the lack of connectivity across these components as well as to other content areas, reflecting repeated calls for more links within and across topics and disciplines with similar conceptual underpinnings (e.g., Bobis, Mulligan, & Lowrie, 2009). Although measurement understandings have been linked to the development of geometry, number and algebra (e.g., Booker & Windsor, 2010), few studies have addressed connections with statistical literacy.

Variation in measurement is a fundamental understanding, yet there has been limited, if any, attention given to it in curriculum documents. One of the learning objectives of the present activity was the development of an appreciation of variation in measuring and measurements, and the need for accuracy in measurement. Children need to understand what it means to make an accurate measurement, why accuracy is important, and the variation we can expect in a measurement especially if it is repeated (Watson & Wright, 2008). The last understanding, in particular, is rarely addressed in the primary curriculum, yet as Konold and Pollatsek (2002) highlighted, it is an important context for various interpretations of average, an interpretation they refer to as "signal in noise" (p. 268). From this perspective, each measurement is an estimate of an unknown yet specific value. In sum, we argue that connecting statistical and measurement topics can provide a powerful tool for targeting these currently neglected core understandings in the primary curriculum.

Methodology

Background and design

Four Year 4 classes and one Year 4/5 class from a middle socio-economic school participated during the first year of a three-year longitudinal study (2012-2014). We report findings only on the Year 4 students (N=115; mean age = 9.5 years; 43% ESL).

This activity was created in collaboration with the teachers and formed part of their regular mathematics program in the areas of measurement and data. It was in line with the *Australian Curriculum: Mathematics* (ACARA, 2012), where the Year 4 Measurement Strand states students should "Use scaled instruments to measure and compare lengths" (ACMMG084, p. 25). As well for the Data Strand in Year 4, students should "Select and trial methods for data collection … Construct suitable data displays, with and without the use of digital technologies … Include tables, column graphs … Evaluate the effectiveness of different displays in illustrating data features including variability" (ACMSP095, 096, 097, p. 33). Professional development sessions were conducted with the teachers in preparation for their implementation of the activities; these were followed by debriefing sessions where we reflected on the students' and teachers' development as well as our own.

A design-based research approach was adopted, specifically, a design experiment, which involves engineering an innovative educational environment that supports the development of particular forms of learning and studying the learning that takes place in the designed environment (Kelly, Lesh, & Baek, 2008). Complementing the design-based nature of the longitudinal study, measurement of student progress was undertaken together with student and teacher interviews. We only report here on students' initial hands-on experiences in the classroom.

Measurement and Variation Activity

The implementation of the activity, "Measuring a person's arm span," varied in time allocation per class, with an average duration of 5 hours 10 minutes, spread across three days during one week for each class. Working in small groups, students made measurements of arm spans using scaled instruments in two contexts: all students measuring one person and recording the values obtained and, each student having his/her own arm span measured and recorded. Class data were recorded in each context. The students were supplied with various rulers, tape measures, string, and student workbooks.

In addition to the learning objectives cited previously, we also focused on developing careful attention to scale, gaining confidence in predicting representative measurements, describing the shapes of data sets, and determining which types of displays best show the variation in a data set. Various ways of representing the data were possible. An important learning feature was students' consideration of the most effective displays for showing the variation in the two data sets, with the emerging understanding that there is very likely to be measurement error in the first context and, hence, the measurement in the second context is an approximation.

Data collection and analysis

Data collection for this report was based on the scanned completed workbooks of consenting students. The student workbook responses for selected questions reported here were repeatedly analysed using iterative refinement cycles for analyses of children's learning (Lesh & Lehrer, 2000). Each response type was coded by each author, with codes refined, and finally checked by the senior research assistant; consensus was reached on all coding. Some students' responses encompassed more than one category, and some responses were incomplete; hence the number of responses reported varies across questions. In the remainder of this paper we report on selected responses from the student workbooks, specifically, five questions after completing measurements on the single student from the first context of the activity and one from the second context.

For the first context, students were asked the following questions: 1. Were all of the values the same? Why or why not? 2. Were you surprised at some of the values? Which ones? Why? 3. Write a summary of how accurate you think the measurements in the table are. What is your "best guess" of the arm span of the person the class measured? How confident are you of this value? 4. Create a graph or plot or picture to represent the values in the [results] table. 5. Write a summary statement about what your representation shows about the measurements your class made of the arm span of the person you measured. Think about the variation that is seen in your plot or picture. In the second context, while collecting measurements of class members, students were asked to answer the following question: 6. How accurate do you expect your results to be compared to our last lesson? As the data from each class were genuine, two classes had values that would be classified as outliers, two classes showed a large degree of variation, but no "certain" outliers, and one class had very consistent measurements.

Results

Question 1: Perceptions of variation in the values (first context)

Five main categories of responses were identified in the analysis of the first question. Of the 96 responses to this question, 42% noted the measuring tools and how they were used. A further 29% mentioned how a tool was not used as accurately as possible. One response, for example, which incorporated both reasons stated, "No. Because the types of materials were different, it affected the measurement. Also overlapping the materials and not having them straight, changed it as well." Five percent of responses noted movement in the person being measured, as in the response, "One of the reasons I think the values were diffrent (sic) might be because P.'s arms might have goten (sic) tyerd (sic)." Eight percent of responses referred to the use of different units of measurement, such as, "No. The measurements were all not the same because some were not quite accurate as others and some people used cm and some used m and cm." The remaining responses (15%) cited various more nebulous "differences" with respect to variations in values and personal ways of measuring.

Question 2: Identification of unusual values (first context)

Analysis of the children's responses (N=94) to the question about surprising values yielded four categories (in addition to an irrelevant/uninterpretable category). Overall 63% of responses identified an outlier or extreme/unusual value including why it was the case, for example, "Yes, T. did 99 cm and everybody else did over 110 cm." Another 19% noted variation in the values, such as, "Y. and M. because they had a big difference: 13cm! B.'s arm couldn't grow that fast!" and "Yes, I was surprised at the 146 cm measurement and 159 measurement because of the veration (sic)." Only one student mentioned that different measuring tools could be responsible for the surprising values, whereas the remaining responses (13%) stated a general lack of surprise with reasons including reference to values considered "more or less the same," or values close to an estimated arm span, such as, "I knew everyone would measure at approx. 150 cm". Given the different data across classes, these percentages are not as important as the fact that so many students (83%) appreciated the variation involved in the process.

Question 3: Consideration of accuracy and "best guess" (first context)

In considering the accuracy of their measurements, 100 children's responses were classified in 8 groupings. Children's assessment of their measurement accuracy was based primarily on the mode or frequencies of values (35% of responses), such as, "I think the measurements in the forties would of (sic) been close to the exact measurement because it's the most popular measurements." Eleven percent of responses displayed an awareness of average or central tendency, with one student saying, "I think the best answer was 127 cm because many people got 126 cm and 128 and just 2 people got 127 cm but it is in the middle. I am kind of confident." The observation of variation in assessing accuracy, as occurred with the mention of values being "a bit more apart," was noted by 12% of students. Nine percent made reference to a visual approximation of what an arm span should be in Year 4, one measured student saying, "I would have thought that my arm span would be around 1m 48." Fewer references were made to an outlier (4%) for this question, for example in the identification of a "fake" value such as 146 cm "because the number can't go up to 146." The accuracy of the measuring tools was mentioned by 4% of students, for example, "My best guess is 154cm because the ruler in my opinion is the best measurement unit because it cannot bend." Another 4% of responses noted the position or movement of the student being measured, as in "I thought the measurements were inaccurate because sometimes he stretched as far as he could and sometimes he didn't." Twenty-three percent of responses, however, simply gave a limited statement of accuracy such as, "No, I don't think the measurement was that accurate."

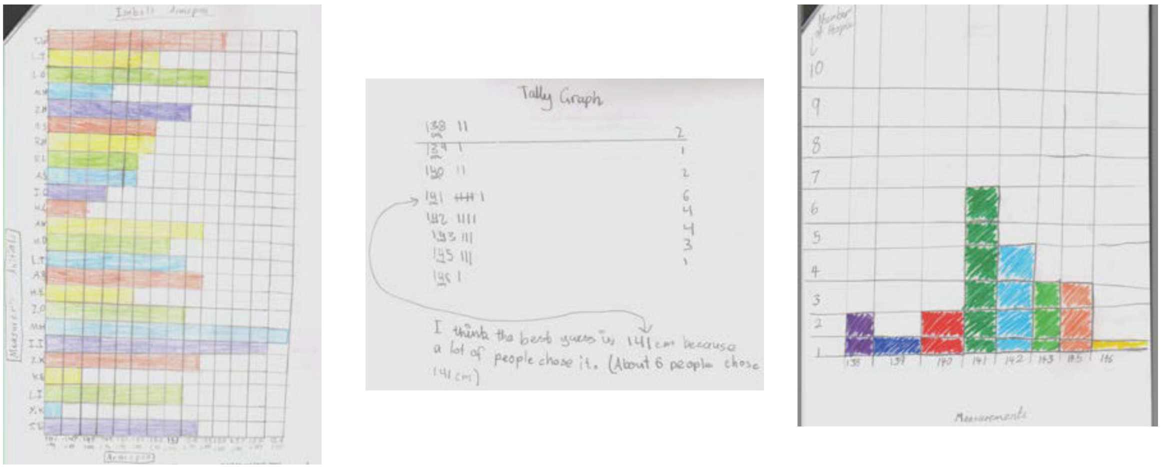
Question 4: Representations (first context)

Students were then given a blank page in their workbooks and asked to "create a graph or plot or picture" to represent the measurement values collected by the class. Of the 83 students, 6 students created two representations, making a total of 89. Of the 89, 7% could not be interpreted, 36% focussed on the actual measured values and 57% focussed on the frequency with which the measured values occurred. Two types of representations were created for the measured values: lists (either unordered [6%] or ordered [3%]) or value plots (all of which were unordered [27%]). The plot in Figure 1(a) is of this latter type. Of the frequency representations, 4% were frequency plots that did not accurately show the correct totals from the class data; 16% of representations were tables of tallies either unordered (7%) or ordered (9%). An example of the latter is shown in Figure 1(b). The rest of the graphs were frequency plots, either unordered (8%) or ordered (29%). Although there were several variations, the plot shown in Figure 1(c) was the most typical.

Question 5: Interpretation of representations (first context)

Students were then asked to write a summary statement about what their representations showed about the measurements, keeping in mind variation seen in the plot. Of the 83 responses, 11% were not actually related to the data, for example mentioning colour or describing people. Thirty-three percent of responses gave a strict description of the graph with no summary of difference (variation) or confirmation of the arm span of the student measured (expectation), for example, "My graph shows the measurements as well as who measured them." Twenty-eight percent of responses, perhaps due to the instructions, mentioned variation in terms of difference, range, smallest, and/or largest. Five students (6%) noted "most" or "centre" or "likely," whereas 17% of students mentioned aspects of both variation and expectation. An example with the same data as

Figure 1(b) including both was: “most people measured that their[s] are the 141 cm and the lowest number was 138 cm and the biggest is 146 cm.”



(a) Unordered value plot (b) Ordered tallies (c) Ordered bar chart

Figure 1. Examples of students’ plots.

Question 6: Comparison of accuracy between contexts (second context)

Six categories of responses (N=84) were identified for this question (a focus on ‘more accurate’ data in second context). The most frequent number of responses (33%) mentioned measuring tools and how they were used, such as, “I think the results are accurate because we were measuring on a f lat surface and the tape was put in place accuratly (sic).” Next most frequent (26%) were responses noting the reliability of the person undertaking the measurement. Knowledge or practice gained from the first context was a feature of 11% of responses, for example, “Yes, I think they will be because now we have had practice at measuring we might be more accurate than last time,” and “We have learnt more about measurement.” Other responses made nebulous comments on “difference” in individuals, strategies, or frequencies of values (12%); gave unexplained percentage or numerical values (4%); or appeared irrelevant or uninterpretable (14%).

Discussion

This study took place across five classrooms, each collecting its own measurement data. It is acknowledged that in some ways this is a limitation because if all classes had been presented with the same data set, we could have expected more consistent response types across classes. It was felt strongly, however, by the researchers and teachers that the students needed to own their data; feedback from all involved, including students, indicated this decision was appropriate. The analysis presented here is hence global in nature, covering all classes and allowing us to make general suggestions about the ability of students in Year 4 to handle the concept of variation in a measurement context.

The representations produced were of two types, those repeating in some way the actual measurements recorded by the class and those further refining the data to represent the frequency of occurrence of each measurement. This process of moving, for example, between a value plot (Figure 1(a)), perhaps through a recording of frequency (Figure 1(b)), to a frequency bar chart (Figure 1(c)), is an example of "transnumeration" (Wild & Pfannkuch, 1999) –"changing representations to engender understanding" (p. 227). What is shown in this study is that most Year 4 students could begin this process and 55% could produce the higher level of a summary representation. Future analyses will allow consideration of responses in relation to specific classroom settings and also of the association between responses on related items.

The responses to Question 6, from the second context of measuring all class members' arm spans, were expected to reflect the variation experienced when many measurements were made of a single student in the first context and leading to the realisation that there could be similar variation from the actual value with the single measurement of each student in the class. Another result, however, of operating in an actual school situation, was that students' responses reflected other more obvious, and less theoretical, aspects of the measuring process. Due to time constraints in some classes, the researchers had to assist in the measuring and measured against a flat wall. Both of the aspects stood out and were legitimate reasons for belief in greater accuracy for the second context.

Conclusion

Overall, the findings indicate that fourth-grade students can begin to think about the inferences that can be made from data collection and even the uncertainty involved in their decisions about measurements taken. The students in this study could see that variation in values obtained in the first context was due in large part to the measuring tools and how they were used. Identification of outliers was recognised and appropriate explanations given, indicating an appreciation of the variation involved in the measuring process. Although students were able to begin the transnumeration process, with many advancing to a higher level, the written interpretations of their representations were somewhat limited in their reference to core concepts such as "centre," likely," etc. As noted, the activities presented to these Year 4 students were in direct alignment with ACARA (2012). Children's capabilities in drawing informal inferences need to be recognised, with increased exposure to a range of statistical representations requiring interpretation and explanation beyond basic descriptions. If children are not exposed to informal inference in the primary school, the introduction of formal statistical tests in later schooling can become a meaningless experience because students will not have developed an intuition about the story conveyed by data.

Acknowledgement

This project reported is supported by a three-year Australian Research Council (ARC) Discovery Grant (120100158). Any opinions, findings, and recommendations expressed here are the authors'. We acknowledge the excellent support from the SRA, Jo Macri.

References

Australian Curriculum, Assessment and Reporting Authority (ACARA). (2012). *The Australian Curriculum: Mathematics, Version 3.0, 23 January 2012*. Sydney, NSW.

Bobis, J., Mulligan, J., & Lowrie, T. (2009). *Mathematics for children: Challenging children to think mathematically* (3rd ed.). Frenchs Forest, NSW: Pearson Education.

Booker, G., & Windsor, W. (2010). Developing algebraic thinking: Using problem solving to build from number and geometry in the primary school to the ideas that underpin algebra in high school and beyond. *Procedia- Social and Behavioral Sciences, 8*, 411-419.

Cobb, G. W., & Moore, D. S. (1997). Mathematics, statistics, and teaching. *American Mathematical Monthly, 104*, 801-823.
Curriculum Corporation. (2006). *Statements of Learning for Mathematics/Statements of Learning for Science.* Carlton South: Author. Retrieved 22 February 2013 from http://www.curriculum.edu.au/verve/_resources/StmntLearning_Science_2008.pdf
English, L. D. (2010). Young children's early modelling with data. *Mathematics Education Research Journal, 22*(2), 24-47.
English, L. D. (2012). Data modelling with first-grade students. *Educational Studies in Mathematics, 81*(1), 15-30.
Gal, I. (2002). Adults' statistical literacy: Meanings, components, responsibilities. *International Statistical Review, 70*, 1-51.
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*, 372-396.
Kelly, A. E., Lesh, R. A., & Baek, J. Y. (Eds.). (2008). *Handbook of design research methods in education*. New York: Routledge.
Konold, C., & Pollatsek, A. (2002). Data analysis as the search for signals in noisy processes. *Journal for Research in Mathematics Education, 33*(4), 259-289.
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.
Lowrie, T., Logan, T., & Scriven, B. (2012). Perspectives in geometry and measurement 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*. Online Publication: MERGA, pp. 71-88.
Makar, K., & Rubin, A. (2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal, 8*(1), 82-105.
National Council of Teachers of Mathematics. (2006). *Curriculum focal points for prekindergarten through grade 8 mathematics: A quest for coherence*. Reston, VA: Author.
Shaughnessy, J. M. (2006). Research on students' understanding of some big concepts in statistics. In G. Burrill & P. Elliott (Eds.), *Thinking and reasoning with data and chance* (pp. 77-98). Reston, VA: National Council of Teachers of Mathematics.
Watson, J. M. (2005). Variation and expectation as foundations for the chance and data curriculum. In P. Clarkson, A. Downton, D. Gronn, M. Horne, A. McDonough, R. Pierce & A. Roche (Eds.), *Building connections: Theory, research and practice* (pp. 35-42). (28th annual conference MERGA).
Watson. J. M. (2006). *Statistical literacy at school: Growth and goals*. Mahwah, NJ: Lawrence Erlbaum.
Watson, J., & Wright, S. (2008). Building informal inference with TinkerPlots in a measurement context. *Australian Mathematics Teacher, 64*(4), 31-40.
Whitin, D. J., & Whitin, P. E. (2011). *Learning to read numbers: Integrating critical literacy*. New York: Routledge.
Wild, C., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review, 67*(3), 223-248.