Assessing and Fostering Children's Statistical Thinking

Graham Jones,
Carol Thornton,
Cindy Langrall & Ed Mooney
Illinois State University
Normal, IL 61790-4520
jones@ilstu.edu

Bob Perry
University of W. Sydney
Australia
b.perry@uws.edu.au

Steven Nisbet
Griffith University
Brisbane, Qld 41111 Australia
snisbet@gu.edu.au

Ian Putt
J. Cook University
Townsville, Qld 4811
Australia
Ian.Putt@jcu.edu.au

1. Overview

In response to the critical role that information plays in our technological society, there have been international calls for reform in statistical education at all grade levels (e.g., National Council of Teachers of Mathematics, 1989; School Curriculum and Assessment Authority & Curriculum and Assessment Authority for Wales, 1996). These calls for reform have advocated a more pervasive approach to the study of statistics, one that includes describing, organising, representing, and interpreting data. This broadened perspective has created the need for further research on the learning and teaching of statistics, especially in the elementary grades, where instruction has tended to focus narrowly on graphing rather than on broader topics of data handing (Shaughnessy, Garfield, & Greer, 1996).

Notwithstanding these calls for reform, there has been relatively little research on children's statistical thinking and even less research on the efficacy of instructional programs in data exploration. Although some elements of children's statistical thinking and learning have been investigated (Cobb, 1999; Curcio, 1987; Curcio & Artz, 1997; De Lange et al., 1993; Gal & Garfield, 1997; Mokros & Russell, 1995), research on students' statistical thinking is emergent rather than well established. Existing research on children's statistical thinking has certainly not developed the kind of cognitive models of students' thinking that researchers like Fennema et al. (1996) deem necessary to guide the design and implementation of curriculum and instruction.

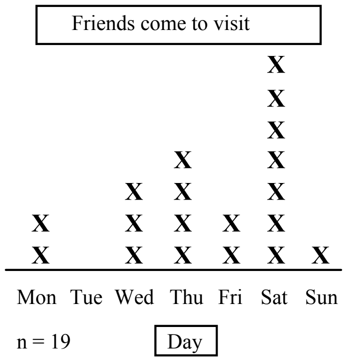
In this paper, we will discuss how our research has built and used a cognitive model to support instruction in data exploration. More specifically, the paper will: (a) examine the formulation and validation of a framework that describes students' statistical thinking on four processes; and (b) describe and analyse teaching experiments with grades 1 and 2 children that used the framework to inform instruction.

2. The Statistical Thinking Framework (Framework)

In generating the Framework (Figure 1), we identified four key statistical processes: describing data, organising and reducing data, representing data, and analysing and interpreting data. These processes which will be described below were modifications of similar processes identified by Shaughnessy et al. (1996). Based on our earlier work with number sense (Jones, Thornton, & Putt, 1994) and probability (Jones, Langrall, Thornton, and Mogill, 1997), the Framework was also formulated on the assumption that elementary children would exhibit four levels of statistical thinking in accord with Biggs and Collis's (1991) general development model, Structure of the Observed Learning Outcome (SOLO). These levels of statistical thinking were described as idiosyncratic, transitional, quantitative and analytical, and in a subsequent validation study with 20 target students, 4 from each of grades 1 through 5, we confirmed the existence of these four levels and refined the descriptors of children's thinking in the Framework.

Key Processes. The first process, describing data, incorporates what Curcio (1987, Curcio & Artz, 1997) calls "reading the data." Curcio notes that reading the data means extracting information explicitly stated in the display, recognising graphical conventions and making connections between context and data. Based on this definition and previous research (Beaton et al., 1996; Pereira-Mendoza & Mellor, 1991) we generated tasks to assess children's thinking on this process. A sample of these tasks is shown in Figure 2 (see D1 and D2). Organising and reducing data, incorporates mental actions such as ordering, grouping, and summarising data (Moore, 1997). As such, it also involves reducing data using notions of centre and spread. Some of the tasks used to assess this process (see O1 and O2 in Figure 2) were adapted from previous research (Strauss & Bichler, 1988; Mokros and Russell, 1995). Our third process, Representing data, incorporates constructing visual displays that sometimes require different organisations of data. Several studies (Beaton et al.; Zawojewski & Heckman, 1997) were helpful in building assessment tasks like the one labeled R1 in Figure 2. Analysing and interpreting data involves recognising patterns and trends in the data and making inference and predictions from the data. It incorporates what Curcio (1987) refers to as "reading between the data" and "reading beyond the data." The former involves using mathematical operations to combine and integrate data, while the latter requires students to predict from the data by tapping their existing schema for information that is not explicitly stated in the data. Curcio's research (Curcio, 1987; Curcio & Artz, 1996) was helpful in identifying assessment questions for this process. For example, see *Figure 2 (A2 through A6).*

Sam had some friends come to visit each day during one week in the summer. The number of friends and the days they visited were displayed like this:



D1 What does this picture tell you?
D2 How many friends came to visit each day?
A2 What can't you tell from this picture?
A3 Which day had the highest number of visitors?
A4 How many friends came to visit Sam during this week?
O1 About how many friends came to visit Sam each week?
O2 What is the average number of friends that came to visit Sam each day?
R1 Can you draw this graph in another way?
A5 How many friends would you expect to come to Sam's place every week during the vacation?
A6 How many friends would you expect to visit Sam during a 4-week month?

Figure 2. Line plot of Sam's friends and questions

Thinking Levels. The validation process confirmed the existence of four levels of statistical thinking as postulated on the basis of the SOLO model (Biggs & Collis, 1991). Level 1 thinkers were consistently limited to idiosyncratic reasoning that was often unrelated to the given data and frequently focused on their own personal data banks. Level 2 thinkers were beginning to recognise the importance of quantitative thinking and they even used numbers to invent measures for centre and spread, albeit not always valid. Nevertheless their perspective on data was generally single minded and they seldom connected representations or analyses of the data to its context. Students exhibiting Level 3 thinking consistently used quantitative reasoning as the basis for statistical judgements and had begun to form valid conceptions of centre and spread. Level 3 students took a broader and more flexible approach when exploring data and seemed to be able to represent and analyse data from multiple perspectives even though they rarely made connections between different aspects of the data. Level 4 students used both analytical and numerical reasoning in exploring data and showed evidence of being able to make connections among different aspects of the data. In Figure 3, we show some exemplars of children's responses at each level. The questions refer to those in Figure 2.

Process/ Level	Level 1 Idiosyncratic	Level 2 Transitional	Level 3 Quantitative	Level 4 Analytical
Describing Data Displays (D)	•gives a description that is unfocused and includes idiosyncratic/irrelevant information; has no awareness of graphing conventions [e.g., title, axis labels] of the display •does not recognise when two displays represent the same data OR indicates some recognition but uses idiosyncratic/ irrelevant reasoning •considers irrelevant/subjective features when evaluating the effectiveness of two different displays of the same data	•gives a description that is hesitant and incomplete, but demonstrates some awareness of graphing conventions •recognises when two different displays represent the same data, but uses a justification based purely on conventions •focuses only on one aspect when evaluating the effectiveness of two different displays of the same data	•gives a confident and complete description and demonstrates awareness of graphing conventions •recognises when two different displays represent the same data by establishing partial correspondences between data elements in the displays •focuses on more than one aspect when evaluating the effectiveness of two different displays of the same data	•recognises when two different displays represent the same data by establishing precise numerical correspondences between data elements in the displays •provides a coherent and comprehensive explanation when evaluating the pros and cons of two different displays of the same data
Organising and Reducing Data (O)	•does not group or order the data or gives an idiosyncratic/ irrelevant grouping •does not recognise when information is lost in reduction process •is not able to describe data in terms of representativeness or "typicality" •cannot describe data in terms of spread; gives idiosyncratic / irrelevant responses	•gives a grouping or ordering that is not consistent OR groups data into classes using criteria they cannot explain •recognises when data reduction occurs, but gives a vague/irrelevant explanation •gives hesitant and incomplete descriptions of data in terms of "typicality" •invents a measure--usually invalid--in an effort to make sense of spread	•groups or orders data into classes and can explain the basis for grouping •recognises when data reduction occurs and explains reasons for the reduction •gives valid measures of "typicality" that begin to approximate one of the centres (mode, median, mean); reasoning is incomplete •uses an invented measure or description which is valid, but the explanation is incomplete	•groups or orders data into classes in more than one way and can explain the bases for these different groupings •recognises that data reduction can occur in different ways and gives complete explanations for the different reductions •gives valid measures of typicality that reflect one or more of the centres; reasoning is essentially complete •uses the range or invented measure that has the same meaning as range

Figure 1. Statistical Thinking Framework

Continued Process/ Level	Level 1 Idiosyncratic	Level 2 Transitional	Level 3 Quantitative	Level 4 Analytical
Representing Data (R)	•constructs an idiosyncratic or invalid display when asked to complete a partially constructed graph associated with a given data set •produces an idiosyncratic or invalid display that does not represent or reorganise the data set	•constructs a display that is valid in some aspects when asked to complete a partially constructed graph associated with a given data set •produces a display that is partially valid, but does not attempt to reorganise the data	•constructs a display that is valid when asked to complete a partially constructed graph associated with a given data set; may have difficulty with ideas like scale or zero categories •produces a valid display that shows some attempt to reorganise the data	•constructs a valid display when asked to complete a partially constructed graph associated with a given data set; works effectively with scale, zero categories, •produces multiple valid displays, some of which reorganise the data
Analysing and Interpreting Data (A)	•makes no response or an invalid/irrelevant response to the question, "What does the display not say about the data?" •makes no response or gives an invalid/incomplete response when asked to "read between the data" •makes no response or gives an invalid/ incomplete response when asked to "read beyond the data"	•makes a relevant but incomplete response to the question, "What does the display not say about the data?" •gives a valid response to some aspects of "reading between the data" but is imprecise when asked to make comparisons •gives a vague or inconsistent response when asked to "read beyond the data"	•makes multiple relevant responses to the question, "What does the display not say about the data?" •gives multiple valid responses when asked to "read between the data" and can make some global comparisons •tries to use the data and make sense of the situation when asked to "read beyond the data;" reasoning is incomplete	•makes a comprehensive contextual response to the question, "What does the display not say about the data?" •gives multiple valid responses when asked to "read between the data" and can make coherent and comprehensive comparisons •gives a response that is valid, complete, and consistent when asked to "read beyond the data"

Figure 1. Statistical Thinking Framework continued

Question	Level 1	Level 2	Level 3	Level 4
D1	It tells how many crosses there are.	It tells you how many friends came over and it tells you the day.	It tells you 2 visited on Monday, 0 visited on Tuesday …	It tells you about friends Visiting Sam. Two visited on Monday, …
O2	7, that's how many came to my place.	Between 7 and 0. It's somewhere there but I don't know.	Two, because two came on 2 of the days. It's the most.	About 3. Share them. Take 3 away from here [7] and give it to the day with 0, [and so on], they'll all have about 3.
R1	Drew a graph with snakes that had no resemblance to the original data.	Drew the same graph with circles instead of crosses; one day incorrect.	Drew the graph in a valid manner and turned it through 90 degrees.	Drew a bar graph in a manner that was valid and complete.
A5/A6	A5: 6, otherwise they'd trash the house.	A5: May be 19. But I'm not sure.	A5: Between 10 and 15 because I don't think 19 would come every week. A6: Maybe 55--three weeks of 15 and 1 week of 10. Cause 19 a week is too high.	A5: Well since there were 19 this week, that's a good estimate. I'd say 15 cause there's so many on Saturday. A6: 4 by 15-that's 60. Kind of a lot but 60 per month.

Figure 3. Exemplars of Thinking at Each Level of the Framework

3. The Teaching Experiments

A teaching experiment has been defined as a methodology that is aimed at capturing and documenting students' thinking over time (Steffe & Thompson, in press). During a teaching experiment, researchers develop sequences of instructional activities or learning trajectories (Simon, 1995) and analyse students' mathematical learning as it occurs in the social situation of a classroom or a small group (Cobb, 1999). In our teaching experiments, the learning trajectories (goals, tasks, and expected learning sequence) were based on the Framework, which was also used to trace changes in students' learning during the intervention. The grade 2 teaching experiment comprised 9 sessions each of 40 minutes and we used categorical and numerical data from the class's Butterfly Garden Project as the context for each session. The grade 1 teaching experiment comprised 4 sessions each of 40 minutes and tasks were based on a data set generated from the "number of teeth" lost by the children in the class. All children in both classes were assessed prior to and immediately following the teaching experiments using the same protocol that had been used to validate the Framework.

<u>Effect of the teaching experiment: Quantitative Analysis</u>. For the <u>grade 2 class</u>, a Wilcoxin Signed Ranks Test (Siegel & Castellan, 1988) revealed significant growth between the pre and post intervention levels of the 19 students on each of the four statistical thinking processes: describing data ($p<.001$); organising and reducing data data ($p<.001$); representing data ($p<.002$); and analysing and interpreting data ($p<.004$). For the <u>grade 1 class</u> of 18 children, the Signed Ranks Test revealed significant growth for two of the four statistical process: describing data ($p<.01$) and organising and reducing data ($p<.02$); the other two processes did not produce significant differences.

<u>Effect of the teaching experiment: Qualitative Analysis.</u> Several learning patterns emerged from the analysis of instruction and in particular a detailed case-study analysis of 4 target students in each of the grades 1 and 2 classes. These learning patterns are described by statistical process. With respect to <u>describing data</u>, children bring varying degrees of prior knowledge about meanings and conventions associated with contextual data displays. Experiences with different kinds of data during instruction seemed to focus their thinking and produce less idiosyncratic descriptions. Categorical data was more troublesome than numerical data. Children's intuitive thinking with respect to <u>organising and reducing data</u> was problematic. Although they were reluctant to use paper and pencil to reorganise data (especially categorical data), technology proved very helpful in stimulating their organising strategies. Our results also show that collectively children revealed conceptual knowledge of average and spread that

was multifaceted and useful in informing instruction. The difficulty for the teacher is in deciding how and when to use children's different representations of center. Children's prior knowledge in representing data appears to be constrained by limited accessibility to pervasive sorting and organising schemas. However, instruction that incorporated technology or the use of unfinished graphs showed potential in stimulating children's sorting schema and ipso facto their capability for constructing representations. With respect to analysing and interpreting data, children's thinking, prior to the intervention, was more normative on tasks that involved reading between the data than on tasks that involved reading beyond the data. The intervention revealed some unanticipated problems with tasks that focused on reading between the data and also highlighted the importance of context in relation to tasks that involved reading beyond the data.

4. Conclusion

Given the prior knowledge and growth that children demonstrated on all four statistical processes, there is evidence that they can accommodate a broader approach to data exploration. However, if instruction on data exploration is to reach its full potential in the elementary grades, there is a need for further research to build learning trajectories that link the different levels of children's statistical thinking identified in the Framework.

References

Beaton, A. E., Mullis, I. V. S., Martine, M. O., Gonzalez, E. J., Kelly, D. L., & Smith, T. A. (1996). *Mathematics achievement in the middle school years: IEA'S third international mathematics and science study (TIMSS).* Chestnut Hill, MA: Center for the Study of Testing, Evaluation, and Educational Policy, Boston College.

Biggs, J. B. & Collis, K. F. (1991). Multimodal learning and intelligent behavior. In H. Rowe (Ed.), *Intelligence: Reconceptualization and measurement* (pp. 57-76). Hillsdale, NJ: Erlbaum.

Cobb, P. (1999). Individual and collective mathematical development: The case of statistical analysis. *Mathematical Thinking and Learning*, 1, 5-43.

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

Curcio, F. R. & Artz, A. F. (1997). Assessing students' statistical problem-solving behaviors in a small-group setting. In I. Gal & J. B. Garfield (Eds.), *The assessment challenge in statistics education*. Amsterdam, The Netherlands:

IOS Press.

De Lange, J., van Reeuwijk, M., Burrill, G., & Romberg, T. (1993). *Learning and testing mathematics in context. The case: Data visualization*. Madison, WI: National Center for Research in Mathematical Sciences Education, University of Wisconsin.

Fennema, E., Carpenter, T. P., Franke, M. L., Levi, L., Jacobs, V. R., & Empson, S. B. (1996). A longitudinal study of research to use children's thinking in mathematics instruction. *Journal for Research in Mathematics Education*, 27, 403-434.

Gal, I. & Garfield, J. B. (1997). *The assessment challenge in statistics education*. Amsterdam, The Netherlands: IOS Press.

Jones, G. A., Langrall, C. W., Thornton, C A., Mogill, A. T. (1997). A framework for assessing and nurturing young children's thinking in probability. *Educational Studies in Mathematics*, 32, 101-125.

Jones, G. A., Thornton, C. A., Putt, I. J. (1994). A model for nurturing and assessing multidigit number sense among first grade children. *Educational Studies in Mathematics*, 27, 117-143.

Mokros, J. & Russell, S. J. (1995). Children's concepts of average and representativeness. *Journal for Research in Mathematics Education*, 26, 20-39.

Moore, D.S. (1997). *Statistics: Concepts and controversies* (4th ed.). New York: Freeman.

National Council of Teachers of Mathematics (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA:,The Council.

Pereira-Mendoza, L. & Mellor, J. (1991). Students' concepts of bar graphs: Some preliminary findings. In D. Vere-Jones (Ed.), *Proceedings of the Third International Conference on Teaching Statistics* (Vol. 1, pp. 150-157). The Netherlands: International Statistical Institute.

School Curriculum and Assessment Authority & Curriculum and Assessment Authority for Wales. (1996). *A guide to the national curriculum.* London: School Curriculum and Assessment Authority.

Shaughnessy, J. M., Garfield, J., & Greer, B. (1996). Data handling. In A. J. Bishop, K. Clements, C. Keitel, J. Kilpatrick, & C. Laborde (Eds.), *International handbook of mathematics education* (Part 1, pp. 205-238). Dordrecht, The Netherlands: Kluwer.

Siegel, S., & Castellan, N.J. (1988). *Nonparametric statistics for the behavioral* *sciences* (2nd ed.). New York. McGraw-Hill.

Simon, M. (1995). Reconstructing mathematics pedagogy from a constructivist perspective. *Journal for Research in Mathematics Education* ,(26), 2, 114-145.

Strauss, S. & Bichler, E. (1988). The development of children's concepts of the arithmetic average. *Journal for Research in Mathematics Education*, 19, 64-80.

Zawojewski, J. S. & Heckman, D. S. (1997). What do students know about data analysis, statistics, and probability? In P. A. Kenney & E. A. Silver (Eds.), *Results from the sixth mathematics assessment of the national assessment of educational progress* (pp. 195-223). Reston, VA: National Council of Teachers of Mathematics.