Primary School Students' Statistical Thinking: A Comparison of Two Australian States

Ian Putt
James Cook University
<Ian.Putt@jcu.edu.au>

Bob Perry
University of Western Sydney
<b.perry@uws.edu.au>

Graham Jones, Carol Thornton, Cynthia Langrall, Edward Mooney
Illinois State University
<jones@math.ilstu.edu>

A framework for teaching and assessing statistical thinking comprising four constructs and four levels for each construct, has been developed and the framework validated using data from 20 US students in Years 1 – 5. The same validation procedures were implemented in two different cohorts, totalling 40 subjects, of Australian students in Years 1 - 5. Lower levels of coherence were found. This paper reports the Australian data, seeks to address reasons for the differences and compares the levels of performance between the Australian and US students.

Background

International pressure for reform in statistical education (e.g. Australian Education Council, 1994; Lajoie & Romberg, 1998; National Council of Teachers of Mathematics, 1998; School Curriculum and Assessment Authority and Curriculum and Assessment Authority for Wales, 1996) has stimulated research on statistical thinking. This has occurred largely in the primary grades, where the focus has been mainly on graphing rather than on broader aspects of data handling and analysis (Lajoie & Romberg, 1998; Shaughnessy, Garfield, & Greer, 1996). Mokros and Russell (1995) have investigated areas such as data organisation, while data modeling has been studied by Lehrer and Romberg (1996) and graph comprehension by Curcio (1987), and Friel, Bright, and Curcio (1997).

A study by Jones et al. (1998) has developed a framework of students' statistical thinking that could be used to inform instruction. To date, the validation of this framework has been carried out with data from US elementary school students. The study reported in this paper extends the validation by using data from Australia to: (a) validate the statistical thinking framework from an international perspective; and (b) compare the statistical thinking of Australian and US students in Years 1 - 5.

Theoretical Perspectives

The Statistical Thinking Framework (Jones et al., 1998) incorporates four key constructs adapted from Shaughnessy et al., (1996): describing data, organising and reducing data, representing data, and analysing and interpreting data. It is based on previous research (Aberg-Bengtsson, 1996; Curcio, 1987; Friel, Bright, & Curcio, 1997; Mokros & Russell, 1995) and has been validated in the US study (Jones et al., 1998).

The first construct, *describing data*, involves finding explicit information in a visual display, recognising graphical conventions, and making direct connections between the original data and the display (Curcio, 1987). The second construct, *organising and reducing data*, incorporates mental actions on data such as ordering, grouping, and summarising (measures of central tendency and spread) (Moore, 1997). The third construct, *representing data*, involves the construction of visual displays that represent different organisations of data. The fourth construct, *analysing and interpreting data* includes what Curcio (1987) referred to as "reading between the data" and "reading beyond the data" (p. 384). It involves recognising patterns in the data, and making inferences, interpretations, and predictions from the data.

The existence of four levels of statistical thinking across each of the four constructs was demonstrated in the US validation study which used a case study approach with 20 students from Years 1-5. Level 1 is associated with *idiosyncratic* thinking that is often unrelated to the given data and focuses on students' own personal data banks. Level 2 is seen to be *transitional* between idiosyncratic and quantitative thinking with students beginning to recognise the importance of quantitative reasoning and trying to make sense of the data. Level 3 reveals evidence of *quantitative* reasoning as a basis for statistical judgments with the beginning of analysis of data from multiple perspectives. Level 4 incorporates *analytical* and numerical reasoning in data exploration and shows evidence of being able to make connections between different aspects of the data (Jones et al., 1998).

Method

The two Australian samples each consisted of 20 students with four students in each sample being purposefully selected by class teachers from each of the cohorts of Years 1–5. Teacher assessment of student achievement in mathematics were the bases for selection with two students from the middle 50% and one from both the lower and upper quartiles of each grade level. Different schooling patterns resulted in variation in the students' ages among the two samples. The first Australian sample (Aus 1) (mean ages by grade - 5.5 years through 9.7) came from a private school in a provincial city in north Queensland. The second Australian (Aus 2) sample (mean ages by grade - 6.0 years through 10.2) was selected from a suburban public school in Sydney. Seventy-five percent of the students in the latter school came from families with a non-English speaking background, whereas the students in the private school were from families with English speaking backgrounds.

The process used to validate the framework for these two samples was similar to that used in the earlier study (Jones et al., 1998). It involved three components: (a) interviewing and analysing target students' responses to a Statistical Thinking Protocol based on the Framework and comprising tasks from three separate contexts (see Appendix 1 for tasks from one of the contexts); (b) examining the stability of these students' thinking over the four constructs; and (c) elaborating the distinguishing characteristics of each thinking level. This validation compared the statistical thinking of students in each of the two Australian states and separately with that of students in the USA. Qualitative analysis was used in all three aspects of the validation.

The Statistical Thinking Protocol [STP] comprised seven questions associated with describing the data, seven with organising and reducing data, three with representing data, and six with analysing and interpreting data. It was administered individually to each of the 40 students by a member of the research team. Students' responses to the STP open-ended questions and probes were audio taped and transcribed, and student drawings and graphs were also collected. As well, notes were taken by each interviewer.

Jones et al. (1998) developed a coding rubric in the US validation study and this was used by the two Australian researchers. A double-coding procedure (Miles & Huberman, 1994) was followed in which the two Australian researchers independently coded all questions for each student's interview protocol. The coding rubric led to each question being coded by construct and the student's level of thinking. The two Australian researchers then met to compare and negotiate thinking levels on each question. Following this negotiation, each student's dominant level of thinking for each construct was determined by identifying the student's modal level of thinking for all questions associated with that construct. The reliability in coding between the Australian researchers on determining levels was 88% which is comparable to the US figure of 85%.

Results and Discussion

Tables 1-4 show the median levels by constructs within each grade for the three samples of primary school students. In using the median as the centre for each set of student data, we assumed that the levels data were interval, that is, that the intervals between thinking levels are equal. Given that the levels are consistent with Biggs' and Collis's (1991) developmental theory model, this assumption seems reasonable.

Table 1
Describing Data Displays: Median Statistical Thinking Levels

Sample/ Year	One	Two	Three	Four	Five
Aus 1	1.5	2	2	3	2.5
Aus 2	1	2.5	1.5	2.5	2.5
USA	1.5	1.5	3	3	3

Table 2
Organising and Reducing Data: Median Statistical Thinking Levels

Sample/ Year	One	Two	Three	Four	Five
Aus 1	1	1	2.5	2.5	3
Aus 2	1	1.5	2.5	2	2.5
USA	1.5	1.5	2.5	3	2.5

Table 3
Representing Data: Median Statistical Thinking Levels

Sample/ Year	One	Two	Three	Four	Five
Aus 1	1.5	1.5	3	3	3
Aus 2	1	1	3	3	2
USA	1.5	1.5	3	3	3

Table 4
Analysing and Interpreting Data: Median Statistical Thinking Levels

Sample/ Year	One	Two	Three	Four	Five
Aus 1	1.5	2	2	2	2
Aus 2	1	1.5	1.5	2	2
USA	1	1.5	2	2.5	3

Growth of Thinking Levels

From the four tables it can be seen that (a) the median thinking levels on all four constructs for the Australian and US samples generally increased or remained constant with increasing grade levels, although this was not always a smooth transition, especially in the Aus 2 sample; (b) the median levels for the Australian and US samples in Years 1 and 2 were similar on most constructs with the Australian samples being slightly higher by Year 2 on describing data displays but lower than the US sample in Year 5 on this same construct; (c) the pattern of median levels across the five grades for the Aus 1 and US samples was identical for representing data; (d) the median levels for the US sample in Years 3, 4 and 5 were similar to or higher than their Australian counterparts in the same grades; and (e) the Aus 2 sample generally performed at a slightly lower level than the Aus 1 or US samples.

From the overall profiles for the Australian and US students there appears to be similar growth patterns and trends. Differences between the data of the Australian and US samples may be attributed to age differences between corresponding grades (in Australia students are about a year younger than their counterparts in the same grade in the US), different curriculum emphases at each grade (for example, sample Aus 2 does not meet probability in its curriculum, perhaps significantly decreasing the students' potential exposure to data handling), and differences in the sampling, such as the high number of non-English speaking background students in the Aus 2 sample.

Stability Across Constructs

With respect to stability across the four constructs, the raw data for all 40 Australian students were examined to determine the numbers of students who exhibited the same level of thinking on at least three of the four constructs. An example of this consistency can be seen in Figure 1 which shows the levels for the Year 1 students in sample Aus 2.

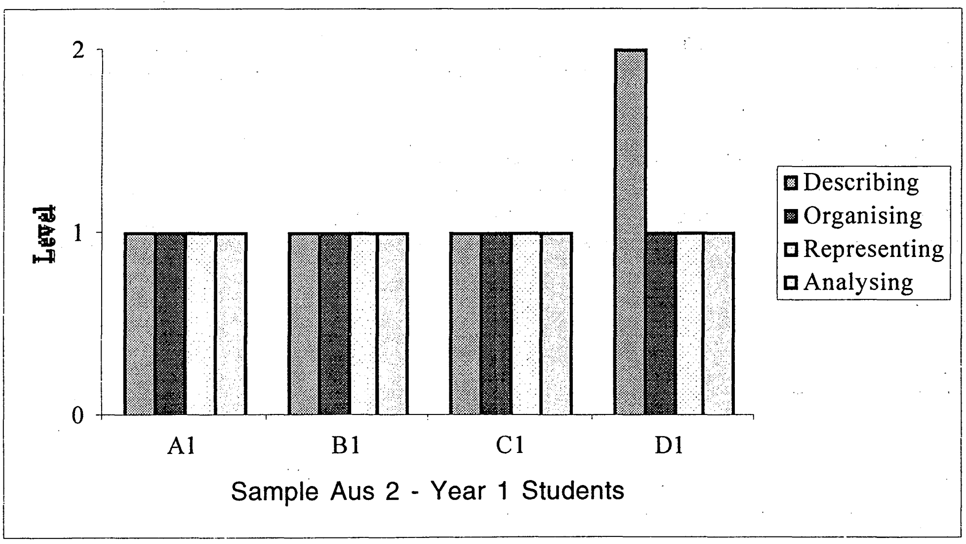

Figure 1. Levels on each construct for Year 1 students in sample Aus 2.

The consistency figures (at least three levels equal) for the Australian samples were: Aus 1 - 65% and Aus 2 - 45%. The corresponding consistency data from the US study (Jones et al., 1998) was 80%. Hence, while the levels for sample Aus 1 were relatively stable and coherent across all four constructs, those for sample Aus 2 were less so. In the case of the

Aus 2 sample, there were a number of language and cultural issues which need to be considered when interpreting these data. Figure 2 shows these less consistent levels for the Year 2 students' in sample Aus 2.

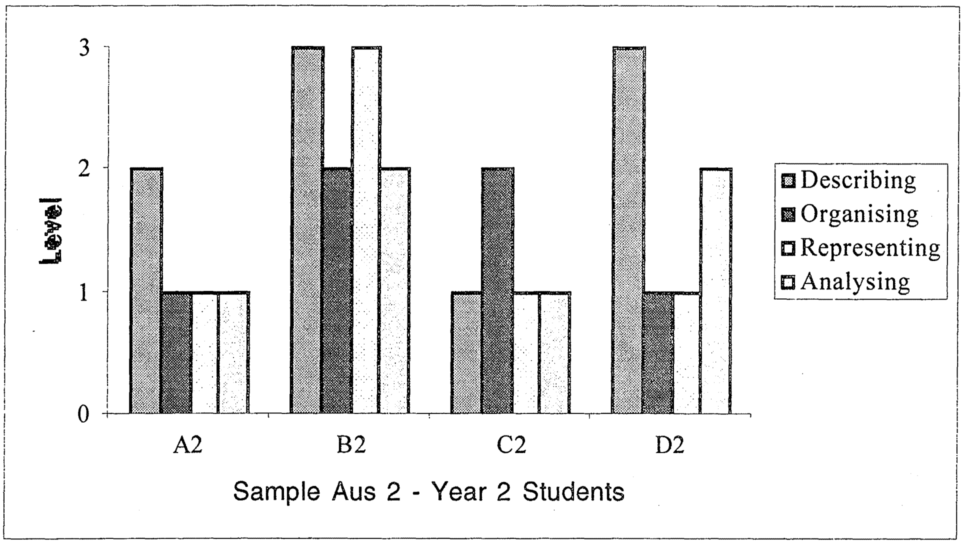

Figure 2. Levels on each construct for Year 2 students in sample Aus 2.

Exemplification of Levels of Thinking

Analysis of the students' statistical thinking across all three samples exemplified the distinctive features of each level of the Statistical Thinking Framework. These are illustrated in responses of students from the different samples.

Level 1 - Idiosyncratic. In answer to the question "About how much did Pete score each day?", a Year 1 student from the Aus 1 sample said "Ten ... maybe he likes ten", while another from Aus 2 suggested "Some days - on Monday, Tuesday, Wednesday, Friday and Thursday, they go home and play after school". When asked "About how many friends would you expect to come to Sam's place each week during the summer holidays?", two Year 1 students from the Aus 1 sample suggested "Two ... because his Mum wouldn't be so busy" and "One ... because ... like you're having too much friends over".

Level 2- Transitional. In response to "What's the average number of Beanie Babies for each child?", a Year 4 student from the Aus 1 sample answered "four … 'cause there's two people with four". This student seemed to have a developing concept of average as the mode. In response to "How many Beanie Babies does each child have?", a Year 5 student from the Aus 1 sample explained that "they should have around 4 or 5 or 3 ... cause if the kids are little, instead of having really hard toys and hurting themselves, they all have soft toys". Here we see a mixture of idiosyncratic and quantitative responses with this student trying to make sense of the data as best she could.

Level 3 - Quantitative. In response to the question about Susie's average score, a student in Year 2 from sample Aus 2 concentrated on the mode and said "3, 3 is the most number in there". A good example of a quantitative response was given by a Year 5 student in the Aus 1 sample in answering the question "About how many friends came to visit Sam each day?" She said "About 3. ... Well lots of numbers are close to three. Its about an average of three". This is approaching a level four response as it does exhibit some analytical and numerical reasoning.

Level 4 - Analytical. In response to "What was Pete's average score?", a Year 2 student from Aus 2 said "Made the average, plussing all together and divide by 5 - I learned that at Vietnamese school".

Educational Importance of the Study

A National Statement on Mathematics for Australian Schools (1991) emphasised Chance and Data as one of the strands that should be part of mathematics curricula in Australia. Various state education departments have developed or are developing revised curricula to reflect recent research into enhancing students' learning of areas such as Chance and Data. While there has been a small number of research studies into how children think about statistical concepts, there is a need for further research with a focus on informing instruction (e.g., Fennema et al., 1996). As well as seeking to address the void in research-based knowledge of students' statistical thinking, this study investigated students' statistical thinking across different cultures in accord with Shaughnessy's (1992) recommendation.

The Australian sample Aus 2 contained a range of children from different ethnic groups whose mathematical experiences were likely to be more varied and who would have had English as a second or third language. This is likely to have accounted for the greater variation in the results for this group compared with sample Aus 1 and the US sample who were mainly first language English speakers. One aspect of this study that needs to investigated further is the relevance of the different contexts in which the probes are embedded. The Beanie Babies and the Bean Bag Game may have been less relevant for the students in the Aus 1 and Aus 2 samples although there are similar toys available in Australia. One of the members of the research team is planning to conduct a similar study with children in a remote Indigenous community in Australia to investigate their statistical thinking and to explore the appropriateness of the tasks.

References

Aberg-Bengtsson, L. (1996). Studying primary school children's learning of graphics: Some experiences using video recordings as data. Paper presented at the 24th NFPF Congress, Lillihammer, Norway.

Australian Education Council. (1994). *Mathematics: A curriculum profile for Australian schools*. Melbourne: Curriculum Corporation.

Australian Education Council. (1991). *A National Statement on Mathematics for Australian schools*. Melbourne: Curriculum Corporation.

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

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

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

Friel, S. N., Bright, G. W., & Curcio, F. R. (1997). Understanding students' understanding of graphs. *Mathematics Teaching in the Middle School, 3*, 224-227.

Jones, G. A., Thornton, C. A., Langrall, C. W., Mooney, E. S., Perry, B., & Putt, I. (1998). Students' statistical thinking. In S. Berenson, K. Dawkins, M. Blanton, W. Coulombe, J. Kolb, K. Norwood, & L. Stiff (Eds.), *Proceedings of the nineteenth annual meeting, North American Chapter of the International Group for the Psychology of Mathematics Education* (Vol. 1, pp. 371-376). Columbus, OH: ERIC Clearinghouse for Science, Mathematics, and Environmental Education.

Lajoie, S. P. & Romberg, T. A. (1998). Identifying an agenda for statistics instruction and assessment in K-12. In S. P. Lajoie (Ed.), *Reflections on statistics: Learning, teaching, and assessment in grades K-12* (pp. xi -xxi). Mahwah, NJ: Lawrence Erlbaum.

Lehrer, R. & Romberg, T. (1996). Exploring children's data modelling. *Cognition and Instruction*, 14 (1), 69-108.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis*. Thousand Oaks, CA: Sage.
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 edition). New York: Freeman.
National Council of Teachers of Mathematics. (1998). *Principles and standards for school mathematics: Discussion draft*. Reston, VA: Author.
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. (1992). Research in probability and statistics: Reflections and directions. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 465-494). New York: Macmillan.
Shaughnessy, J. M., Garfield, J., & Greer, B. (1996). Data handling. In A. J. Bishop, M. A. Clements, C. Keitel, J. Kilpatrick, & C. Laborde (Eds.), *International handbook of mathematics education*, 382-393.

Appendix 1: The Beanbag Game and Questions

Susie and Pete play a beanbag game each day after school. In this game they try to throw a beanbag into a circle. Each takes 5 turns every day. The number of times each child landed the bag in the circle last week is shown below.

Scores for Susie and Pete

Day	Susie	Pete
Monday	3	3
Tuesday	4	0
Wednesday	3	5
Thursday	2	3
Friday	3	4

Describing data displays

Q. 1 What do these two sets of scores tell you?

Analysing and interpreting data

Q. 6 How are Susie's scores and Pete's scores alike? Explain.
How are Susie's scores and Pete's scores different? Explain.

Organising and reducing data

Q. 1 About how much did Susie score each day? How did you figure that out?
What was Susie's average score? How did you figure that out?

Q. 2 About how much did Pete score each day? How did you figure that out?
What was Pete's average score? How did you figure that out?

Q. 4 What set of scores (Susie's or Pete's) has the greatest spread, or do they both have the same spread? Explain.

Appendix 2: Statistical Thinking Framework

Construct / 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 recognize when two different displays represent the same data OR indicates some recognition but uses idiosyncratic/irrelevant reasoning •considers irrelevant or 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 •recognizes 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 •recognizes 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	•recognizes 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
Organizing and Reducing Data (O)	•does not group or order the data or gives an idiosyncratic/ irrelevant grouping •does not recognize when information is lost in reduction process •is not able to describe data in terms of representativeness or "typicality" •is not able to describe data in terms of spread or gives idiosyncratic/relevant responses	•gives a grouping or ordering that is not consistent OR groups data into classes using criteria they cannot explain •recognizes 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 •recognizes when data reduction occurs, and can explain the reasons for the reduction •gives valid measures of "typicality" that begin to approximate one of the centers (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 •recognizes 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 centers; reasoning is essentially complete •uses the range or an invented measure that has the same meaning as the range
Representing Data (R)	•constructs an idoisyncratic 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 reorganize 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 reorganize 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 reorganize the data	•constructs a valid display and 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 reorganize the data
Analyzing 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"