Year Seven Students' Representation of Numerical Data: The Influence of Sample Size

Steven Nisbet
Griffith University
<s.nisbet@mailbox.gu.edu.au>

Ten children in Year 7 of primary school were given two numerical data sets to represent graphically – one with 10 pieces of data, and one with 30. Increasing the size of the data set had little effect on the children's organisation of the data. In general, the children were no more likely to represent the larger data set in an organised form than the smaller data set. The majority of the children represented the smaller data set without any reorganisation. Similarly, the majority of the children required prompting to organise the larger data set. Mathematical ability did have some effect in that some of the more mathematically able children found it easier to organise the data than their less able counterparts. Possible explanations for the results are explored and the implications for teaching and the curriculum are discussed.

This paper concerns the ability of upper-primary children to represent numerical data in a graph, and explores factors which assist children to organise the data before drawing the graph. The term numerical data refers to counts or measures (e.g., number of CDs bought), in contrast to categorical data which refers to categories (e.g., eye colour).

The ability to draw an organised graph is one in a suite of skills expected of all students according to recent curriculum documents. The Australian Numeracy Benchmarks (Curriculum Corporation, 2000) include the ability of primary school students to organise, summarise, and display information in graphs. Similarly, the National Council of Teachers of Mathematics Standards (NCTM, 2000) has highlighted the need for students at all levels to be able to organise and represent data.

Research into children's ability to draw graphs has included the development of a framework for statistical thinking by Jones, Thornton, Langrall, Perry, & Putt (2000) (Framework). The third construct in the Framework - Representing Data - is the main issue under consideration in this study, and incorporates constructing representations that exhibit different organisations of the data. As with the other constructs, four levels of thinking have been proposed for this construct. The levels are defined by statements describing students' data displays in terms of the validity of a display when asked to complete a partial graph, and the degree of reorganisation of data when asked to produce a display. The evidence obtained in this study relates to the latter – the degree of reorganisation of data shown in the display.

According to the Framework, at Level 1 the student produces an idiosyncratic display that does not represent the data set. At Level 2, the student produces a display that represents the data but does not attempt to reorganise the data. At Level 3, the student not only produces a display that represents the data but also shows some attempt to reorganise the data. At Level 4, the student produces multiple valid displays, some of which reorganise the data.

Research in the area of students' representation of data is not extensive, however a small number of studies provide some background for this study. Lehrer and Schauble

B. Barton, K. C. Irwin, M. Pfannkuch, & M. O. J. Thomas (Eds.) *Mathematics Education in the South Pacific* (Proceedings of the 25th annual conference of the Mathematics Education Research Group of Australasia, Auckland, pp. 520-527). Sydney: MERGA.

(2000) investigated the process of data organisation with elementary school children in grades 1, 2, 4 and 5. They examined how these children developed and justified models to categorise (by grade level) drawings made by children in the same grade levels as themselves. Their results suggest that, at higher grades, children use more sophisticated strategies for organising data.

Nisbet (1999) examined the representations of categorical data generated by teacher-education students. The majority (99%) drew representations of the data showing some reorganisation of the data. However, the data was categorical, not numerical. Nisbet, Jones, Langrall and Mooney (submitted) analysed primary-school children's representations of categorical and numerical data. The study revealed that numerical data was significantly harder for children to organise and represent than categorical data. Children beyond Grade 1 can make connections between organizing and representing data when the data are categorical but generally not when the data are numerical. Whereas 60% exhibited Level 3 thinking with categorical data by reorganising the data, only 20% exhibited Level 3 thinking with numerical data. Two of the three Level-3 thinkers produced a tally table while the third drew a pictograph.

Another study (Nisbet, 2001) found that teacher-education students had similar difficulties with organising numerical data. All could produce an organised graph from categorical data, but only 19% could produce an organised graph from numerical data. For the latter, the majority of students merely drew separate bars for each individual piece of data without organising the data into numerical categories.

Why do more students find it difficult to represent numerical data in an organised way, compared to categorical data? It could be that the way to organise categorical data is obvious, but less obvious for numerical data. Maybe the need for organisation is not perceived to be great when there are only 10 items in the data set. Perhaps, if the data set was made larger, then the students would be more likely to see the need to organise the data, and subsequently draw an organised graph based on numerical categories. To test this hypothesis Nisbet (2002) presented students in Years 9 and 11 with two data sets, one small (10 pieces of data) and another larger (30 pieces of data) and asked them to represent the data graphically. With the small data set most students drew graphs showing no organisation of the data. However, increasing the size of the data set to 30, lead more students to organise the data and draw a representation based on number of times a value was recorded rather than individual values. There was an effect of mathematics ability with the Year 11 students. Most of these students in the higher ability group were able to organise and represent the data from the larger data set in an organised way without any prompting. However, only one student in lower ability group drew an organised graph without any prompting. There was no similar ability effect for the students in Year 9.

The current study was therefore designed to test the same hypothesis with upper-primary children. Hence in this study, children in Year 7 were given the task of drawing two graphs – one graph to represent a small data set (n = 10), and second graph for a larger set (n = 30). The Framework was used to evaluate the graphs produced by the children's organisations and representations of the data, and an interview protocol was employed to ascertain the extent of prompting required before children realised that the larger set needed to be organised before it could be represented graphically.

Method

Participants

The target for this study was the upper-primary level, because this age group had not been included in previous studies on this topic by the researcher. A sample of 10 children at a suburban primary school was drawn from one of the Year 7 classes; their mathematics abilities varied across the range of ability, high, medium and low.

Tasks

Participants were given two tasks. The first required them to draw a graph representing the information in the following scenario.

Ten students were asked about the number of novels they read during the term. These are their answers.

NUMBER OF NOVELS: 5, 4, 1, 7, 5, 0, 3, 4, 5, 6

Draw a graph which represents the data.

The numbers were presented in a vertical list.

After the children had drawn their graphs, they were asked the following questions:

What sort of graph did you draw? and Why did you draw it that way?

The second task required the children to draw a graph representing this information.

Thirty students were asked about the number of CDs they bought during the year. These are their answers.

NUMBER OF CDs: 2, 4, 2, 7, 5, 0, 3, 4, 5, 1, 5, 4, 1, 7, 5, 0, 3, 4, 5, 6, 3, 4, 8, 6, 3, 2, 3, 4, 5, 6

Draw a graph which represents the data.

The numbers of CDs were presented in a vertical list. [Note that the data did not include names or any identifying information of the people who bought the CDs. This was done so that the children would see any loss of such data in the reorganisation process.]

If a student had drawn an organised graph successfully, he/she was asked:

What sort of graph did you draw? and Why did you draw it that way?

Table 1

Prompt No. 2 – Blank table of values (No. of CDs)

No. of CDs									
No. of people									

If the student was experiencing difficulty in working out what to do, a protocol comprising a series of prompts was available, and the extent of prompting necessary for the student to embark on the correct course of action was noted.

1. How many students bought no CDs? How many bought 1? Does that help with your graph?
2. Could you fill in a table of values like this?(Show blank table of values as in Table 1)
3. Could you draw a graph with this table of values? (Show completed table of values.)

If the first three prompts were not sufficient to assist the child to draw an organised graph, the child would be shown a graph of the data (Figure 1) and asked the following question.

4. Here's a graph drawn by someone else. What does it tell you?

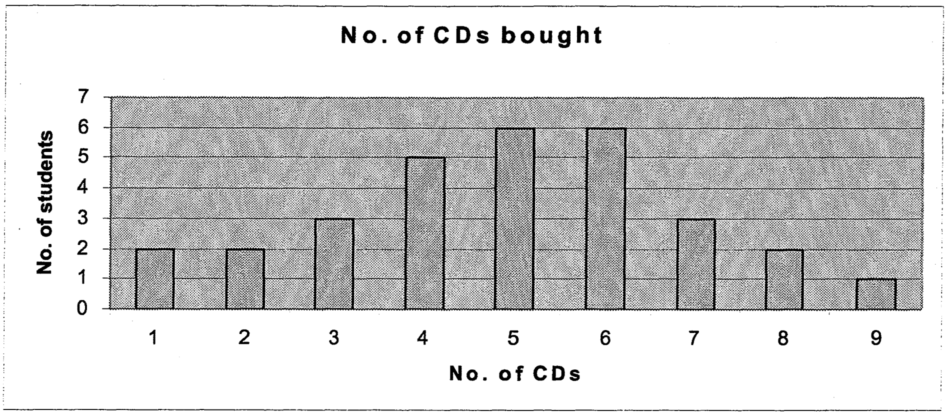

Figure 1. Graph shown in Prompt 4.

Each child was interviewed individually in a separate room away from any distractions of the class. All materials including sheets of graph paper, rulers, pencils and erasers were supplied by the researcher. The researcher kept notes of the answers given to the questions in the protocol, the amount of prompting required, and other comments made by the children in the interviews. All graphs drawn by the children were collected by the researcher for analysis.

Results

Overview of Results

The effect of size of data set. Increasing the size of the data set seemed to have little effect on the children's propensity to re-organise the data. With the smaller data set, the majority of the children drew a bar graph with separate bars for each piece of data. With the larger data set, the majority of the children needed prompting to group the data.

The effect of mathematics ability. There was a limited effect of mathematics ability shown and it was more evident with the larger data set. The children in the high mathematics ability group required fewer prompts to reorganise the data.

The effect of order of tasks. There appears to be no effect of order of tasks on the likelihood of the children to organise the data.

Results in Detail

The effect of size of data set. Increasing the size of the data set seemed to have little effect on the children's propensity to re-organise the data while drawing a graphical representation. Only a small proportion of the children was able to reorganise the data in both examples. One child out of the ten reorganised the smaller data set, and two out of ten reorganised the larger data set without prompting.

With the smaller data set concerning number of books read by students (n = 10), nine out of ten children drew bar graphs with individual bars – one bar for each student (Figure 2). These nine graphs were assessed to be at Level 2 according to the Framework because they did not show any reorganisation of the data. The other child drew a graph of grouped data without hesitation, showing how many students read zero books, how many read one book, two books, etc. (Figure 3). This graph was assessed to be at Level 3 because of the reorganisation of the data.

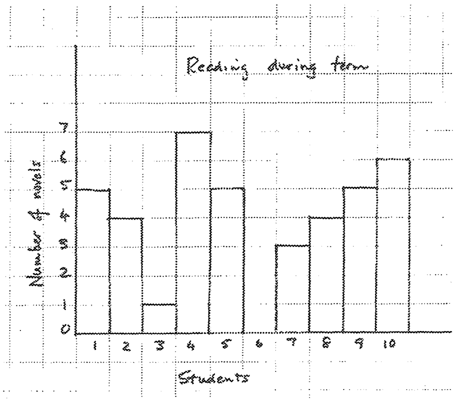

Figure 2. Typical Level 2 graph of smaller data set (number of novels read).

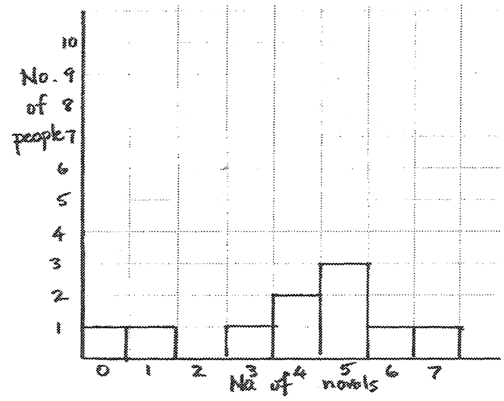

Figure 3. Level 3 graph of smaller data set (number of novels read).

[Note that the graphs shown in this paper have been redrawn by the researcher, as the originals by the children were drawn in pencil and did not scan well enough for reproduction.]

With the larger data set concerning the number of CDs bought by students (n = 30), eight out of ten children needed prompting (as described in the method section) before they thought of grouping the data according to how many CDs were bought. Seven of the graphs were valid and were assessed at Level 3 because they showed a reorganisation of the data The eighth graph was assessed at Level 2 because it was not a valid display (Figure 4). The other two children were able to draw an organised graph without any prompting at all. These two graphs were assessed to be at Level 3, but the difference with these two children is that they produced the organised graphs without any prompting (Figure 5).

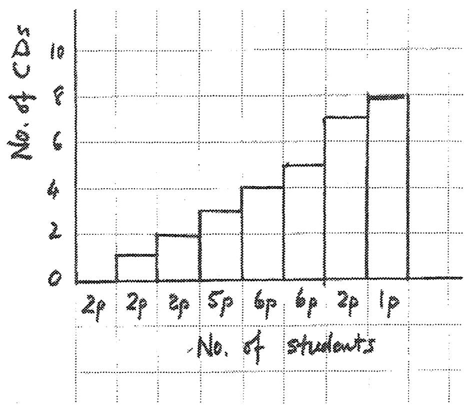

Figure 4. Invalid graph of larger data set

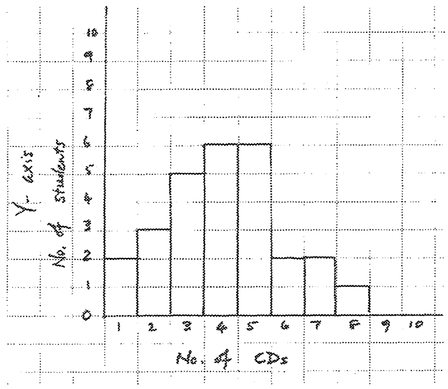

Figure 5. Level 3 graph of larger data set

The effect of mathematics ability. The effect of mathematics ability was demonstrated only to a limited extent. With the smaller data set, the only child to draw an organised graph was a child of high mathematics ability. However, the two other high mathematics ability children drew graphs without reorganisation, similar to the children of low and medium mathematics ability (Table 2).

Table 2
Number of Children and Level of Thinking with the Smaller Data Set by Ability Group.

Mathematics ability group	Level of thinking (representing data) 1 (Idiosyncratic)	2 (Transitional)	3 (Quantitative)	Total Children
Low	0	3	0	3
Medium	0	4	0	4
High	0	2	1	3
Total children	0	9	1	10

The effect of mathematics ability. This was more evident with the larger data set. Although the number of children in each ability group is small, the average number of prompts for each group gives an indication of the general trend: the more mathematically able children required fewer prompts to reorganise the data than their less able counterparts. The average number of prompts was 0.33 for the high ability group, 1.75 for the medium ability group, and 2.67 for the low ability group (See Table 3). The only children to draw an organised graph without any prompting were high ability children.

Table 3
Number of Children Requiring Prompts for Task 2 by Ability Group.

Mathematics ability group	Number of prompts to organise the data 1	2	3	4	Average no. of prompts
Low	1	0	1	1	2.67
Medium	2	1	1	0	1.75
High	1	0	0	0	0.33
Total children	5	2	2	1	1.9 (overall)

The effect of order of tasks. There appears to be no effect of order of tasks on the likelihood of the children to organise the data. Towards the end of the interviews with the children, the researcher thought about the possible effect of order of tasks. Perhaps the smaller data set and no apparent need to reorganise was influencing the children to maintain a 'no reorganisation' stance. So the final two children were given the larger data set before the smaller set. However both of these children required prompting to organise the larger data set. Then, having completed the organised graph for the smaller data set, they reverted to an unorganised graph (bars for individual people) when presented with the smaller data set! This effect needs further investigation.

Other Points Noted During the Investigation

The mechanics of drawing graphs. All the children appeared to be very familiar with the mechanics and conventions of drawing simple graphs. When drawing the graph for the smaller data set, there was little hesitation in drawing the axes and deciding that the number of novels went on the vertical axis (referred to confidently by one child as the Y-axis). One child did label the vertical axis as the 'number of students' but then quickly changed it to 'number of novels'. Two children omitted zero from the scale of number of novels, but inserted it when they realised from the data that it was necessary.

Alternative ways of reorganising the data. Although most children were able to reorganizing the larger data set and draw a graph of number of people (vertical axis) against number of CDs (horizontal axis), albeit with prompting, three children came up with alternative ways of reorganizing the data. One child made 15 pairs from the data (n = 30), and drew a bar graph with 15 bars – one for each pair of students. Another combined the numbers of CDs for each number bought; he drew a bar of height 6 for the two students who bought three CDs, a bar of height 24 for the four students who bought six CDs, and so on (See Figure 6). A third child attempted to group together the students who bought the same number of CDs. However, the graph that he drew was not a valid display (Figure 4). It was merely a sequence of bars showing the number of CDs bought without any representation of the number who bought each number of CDs.

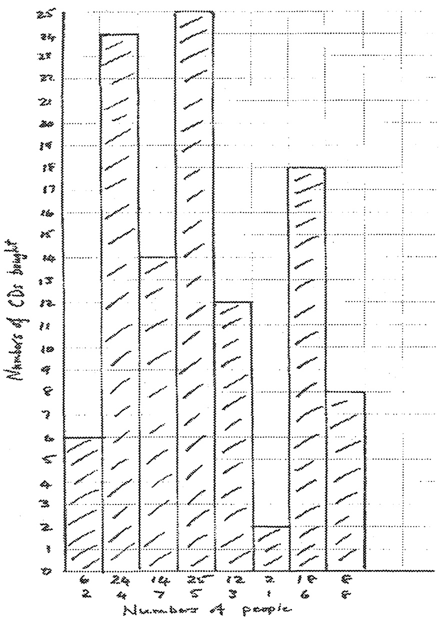

Figure 6. An invalid display.

No table of values. All of the children who required only one prompt to organise the larger data set omitted the drawing of a table of values in the process of generating the organised graph. Instead, for each value of numbers of CDs, they looked down the list of numbers to count how many people bought that many CDs. I assume that these primary children did know about tables of values, because most of those who were shown the table of values in prompt 2 recognised it. However, that step was omitted, leaving the children open to a possible source of error.

Language. It is possible that there exists an ambiguity in the phrase 'number of students', or 'number of children'. For instance, one of the children who drew an unorganised graph for the smaller data set (novels read), wrote 'no. of children' along the horizontal axis, along with the digits '1' to '10' when these digits actually referred to an implied student number, i.e. 1st student, 2nd student, etc.

Discussion

The minimal effect of size of data set on Year 7 children's propensity to reorganise the data raises some questions about (a) the formal primary school curriculum in data handling, and (b) the data handling activities that are given to children throughout the primary grades. It is hoped that the current rewriting of the mathematics syllabus in Queensland will place greater emphasis on data organisation, and its role in the early steps of preparing data for representation. Reference to the Framework by Jones et al (2001) in particular to the constructs 'Organising and reducing data' and 'Representing data' would highlight the importance of having children group data into classes and articulate the basis of the groupings.

Regarding the data-handling activities employed in primary classes, one can only speculate that a mechanistic approach (Ernest, 1989) involving the use of rules and formulae and data found in a textbook may be prevalent in schools. A more beneficial approach may be the dynamic approach (Russell & Friel, 1989) in which students investigate an issue of interest to them and collect their own data, then analyse it, organise it and represent it. Students need experience in organising their own data and representing it, rather than representing data already collected and organised by a textbook author. It would seem worthwhile for students to organise their data by making rank orders, drawing line plots and stem-and-leaf plots, and making tables of grouped data. Further research, in the form of teaching experiments, needs to be carried out to ascertain the impact of such learning activities on children's ability to draw an organised graph.

The fact that the only children to reorganise the data without prompting were high mathematics ability children seems to imply that a component of mathematics ability is the tendency to organise information. However, not all children identified by the school as high mathematics ability did reorganise the data. This issue needs further investigation with respect to both sides of this issue – the defining of high mathematics ability and the ability of teachers to correctly identify children with such ability.

Another point worth considering is the fact that the equivalent study conducted with secondary students (Nisbet, 2002) revealed a greater effect of increasing the size of the data set especially with high mathematics ability students. Further studies could investigate whether this is a curriculum issue or a developmental issue.

References

Curriculum Corporation (2000). *Australian Numeracy Benchmarks*. Melbourne, Vic.: Author.

Ernest, P. (1989). The impact of beliefs on the teaching of mathematics. In P. Ernest (Ed.), *Mathematics Teaching: The state of the art*. New York: Falmer.

Jones, G., Thornton, C., Langrall, C., Mooney, E., Perry, B., & Putt, I. (2000). A framework for characterizing students' statistical thinking. *Mathematical Thinking & Learning, 2*, 269-307.

Lehrer, R., & Schauble, L. (2000). Inventing data structures for representational purposes: Elementary grade students' classification models. *Mathematical Thinking and Learning, 2* (1 & 2), 51-74.

National Council of Teachers of Mathematics (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

Nisbet, S. (2002). *Students representing numerical data: the influence of sample size*. Paper to be presented to the Annual Conference of the International group for the Psychology of Mathematics Education, Norwich, UK, July 2002.

Nisbet, S. (2001). Representing categorical and numerical data. In J.Bobis, M.Mitchelmore, B.Perry (Eds.) *Proceedings of the 24th Annual Conference of the Mathematics Education Research Group of Australasia*, Sydney, NSW, 1-4 July 2001. Sydney: MERGA.

Nisbet, S. (1999). Displaying statistical data. *Teaching Mathematics, 24,* (1).

Nisbet, S., Jones, G., Langrall, C., & Mooney, E. (submitted). Children's representation of data.

Russell, S., & Friel, S. (1989). Collecting and analyzing real data in the elementary school classroom. In P. Trafton & A. Schulte (Eds.) *New Directions of Elementary School Mathematics, 1989 Yearbook.* Reston, Va: National Council of Teachers of Mathematics.

Wood, D., Bruner, J., Ross, S. (1976). The role of tutoring in problem solving. *British Journal of Psychology, 66, vhv*181-191.