Teaching children to organise and represent LARGE DATA SETS in a histogram

STEVEN NISBET and IAN PUTT

report on their teaching investigation in which students explore the organisation and grouping of data. They highlight the type of thinking needed to represent grouped data in graphs.

Although some bright students in primary school are able to organise numerical data into classes (Konold & Higgins, 2000; Nisbet, 2001; Nisbet, Jones, Thornton, Langrall & Mooney, in press), most attend to the characteristics of individuals rather than the group, and 'see the trees rather than the forest'.

How can teachers in upper primary and early high school teach students to organise large sets of data with widely varying values into groups, and to represent the resulting grouped data appropriately in a histogram? How can we choose data sets that allow integration of different topics in the mathematics curriculum? What involvement can students have in the data collection?

These are questions that were addressed in a 'teaching experiment' with two year 7 classes in two different Queensland schools over a three-week period. (Year 7 is the final year of primary school in Queensland).

Data collection

One very effective and efficient way of obtaining a reasonably sized data set is to have students conduct a survey within their own class or across a number of classes. A strategy that worked well in introducing year 7 students to surveys was to supply each student with a data sheet containing the following four questions:

(a) Your height in cm,
(b) Your horizontal reach in cm (arms outstretched),
(c) Your estimate of the number of countries in the world, and

(d) Your estimate of the number of books you have read this year.

Students worked in groups of three or four to collect their own measurements and estimates and then separated the four answers into ice cream buckets labelled appropriately. The whole exercise took about twenty minutes by the time everyone had completed measuring, recording and sorting of the data sets. Tasks chosen ensured that the data were either discrete [the countries and the books] or continuous [height and horizontal reach]. Students were involved in working out the best strategies for measuring height and horizontal reach and in deciding how to arrive at the two estimates.

Part of the efficiency of this approach to data collection was that there were four sets of data that could be used in lessons. Students owned the data and could see where their own particular data were located within the overall representations that they developed in a series of lessons that followed over the next three weeks.

Data organisation and representation

It is one thing to collect a whole lot of data and quite another to organise that data in order to tell a story. The Queensland *Year 7 Teacher Sourcebook* (Department of Education, Qld, 1990) indicates that students will need help with organising data and drawing histograms, and that not all students will reach this level of proficiency in Year 7. This type of data organisation is not usually encountered in Queensland primary schools before Year 7, so we decided to initially test how Year 7 students would organise a large data set without having been taught.

We gave the students the following information and instructions:

> A group of 50 students did a survey and collected data on their heights.
> The following numbers are the 50 students' heights in centimetres.
>
> 168, 180, 139, 166, 152, 151, 176, 132, 155, 160, 179, 157, 135, 134, 147, 172, 171, 141, 164, 165, 140, 173, 168, 167, 137, 159, 155, 143, 154, 153, 172, 155, 145, 149, 171, 150, 151, 155, 170, 169, 165, 167, 149, 162, 159, 153, 148, 158, 154, 156.
>
> 1. In the space below, organise the data into groups — any groups that make sense to you.
> 2. Then, represent the data in any type of graph you like on the graph paper provided.

Interestingly, while about 70% of the students were able to organise the numbers into intervals of size 5 or 10, there were approximately 20% who had no idea how to handle such data. Of the 70% (n=40) who were able to organise into groups only four could represent the data in a way that showed the grouping. Figure 1 shows an example of a student who did make a grouping and did represent the groups but in a way that resembles a bar graph rather than showing the properties of a histogram. The student's graph clearly shows the distribution of the data and is good example of a bar graph. However, had the bars been touching to show the continuity of the scale of students' heights, it could be called a histogram.

On the basis of findings from a number of studies (Bright & Friel, 1998; Jones et al., 2001; Nisbet, 2003) and the above information, we decided to conduct a 'teaching experiment' that focused on implementing a series of

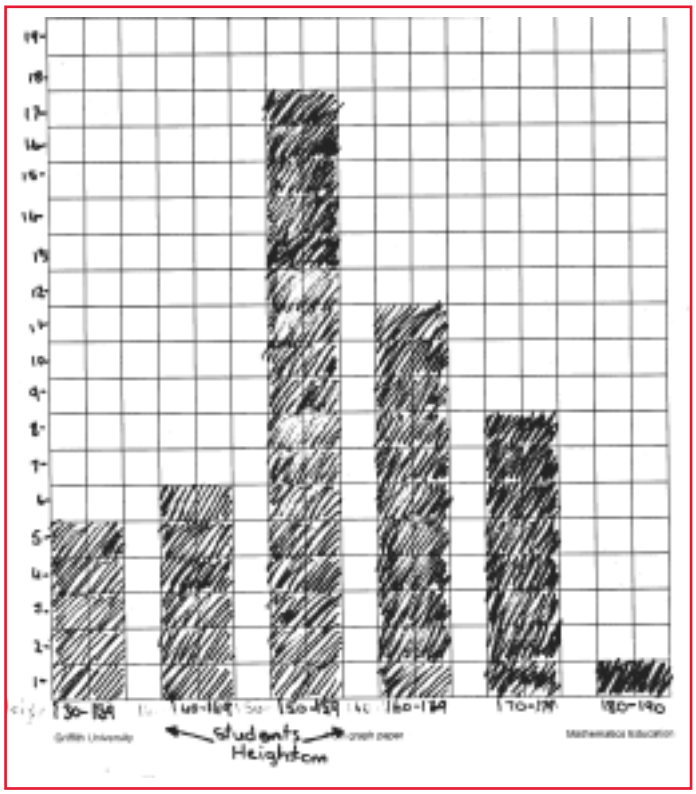

Figure 1. Graph showing organisation by groups.

three lessons on organising and representing such data sets, and noting the changes in the students' responses to the data-handling tasks.

The 'teaching experiment'

Lesson 1

This lesson involved the collection of data described earlier in the article. Our approach was to record the students' height data on the chalkboard as they were randomly drawn out of the corresponding bucket. Organisation of the data involved finding the smallest and largest height and then using these values to determine the range — a concept that many of these primary school students had never heard of before. Once students had arranged the heights from smallest to largest and then calculated the range, they were ready to explore the notion of intervals. The number of intervals and an appropriate size for these data was subject to a valuable discussion as we looked at what would be practical and meaningful. Issues such as overlapping of intervals were addressed and the process of rounding values became one that had an application in this task.

Construction of a grouped frequency distribution was modelled by the teacher and then students prepared their own distribution from the class data on heights. What proved helpful initially was to have a recording sheet that gave some guidance to the students in remembering the sequence of steps in handling the data, (see Figure 2 for an example).

Name: Belinda Tallman

Original order	Smallest to largest	Lowest value = 132 Highest value = 180 Range = 48		
168	132	Possible interval size(s):		5
180	134	Intervals	Tallies	Frequencies
139	135	130–134	/ /	2
166	137	135–139	/ / /	3
152	139	140–144		
etc.	etc.	etc.		

Figure 2. A recording sheet for organising the height data.

Following this, the teacher modelled the process of representing the grouped data in a histogram, namely:

(a) allocate axes (frequencies on vertical axis, height on horizontal axis);
(b) determine range required for each axis;
(c) work out a suitable scale for each axis; and
(d) construct contiguous bars for a histogram.

Care was taken to emphasise the contiguous nature of histograms compared with the non-contiguous nature of bar graphs that was characteristic of the graphs of those students who grouped the data in the pre-instruction task (see Figure 1). Students were given the opportunity to construct their own histogram on a sheet of centimetre graph paper with discussion of cosmetic features of the histogram such as inserting labels on the axes and a title. The teacher then discussed possible ways of describing and interpreting the resulting graph.

Lesson 2

An excerpt from a video (Moore, 1992) on histograms that utilises data on 'the times of lightning strikes' and 'school-arrival times' was viewed as a stimulus to show how and why histograms are used in real life. (This excerpt clearly showed how a frequency distribution is built up from raw data, and the effect of the choice of various interval sizes.)

The procedure from Lesson 1 was modified with the data on the 'number of countries'. A handout with the data listed in random order was given to each student who then cut out each of the numbers and assembled them in

order from smallest to largest. Once completed, the teacher recorded these ordered data on the board and a class discussion ensued about the possible groupings (intervals) that might be used to form the frequency distribution and the corresponding histogram. Two challenges with these data were the magnitude of the numbers and the range of the numbers (from 50 to 64 000). Students realised that a problem existed with drawing a histogram with these values. Two alternatives, that resulted in different histograms, were considered by the different classes, namely, putting all the outliers in a 300+ group, or dropping the outliers that were not considered realistic estimates of the number of countries. In either case, the interval (group) sizes were much larger than for the height data that had a range of only 48.

Once interval size was agreed upon, students constructed their own histograms. Then they were encouraged to write a paragraph on what the graphs told them, which was followed by sharing of their interpretations with the whole class.

Lesson 3

The 'horizontal reach' data were presented unordered on a sheet similar to Figure 2 and students were asked to identify highest and lowest and range, complete the ordering, decide on a possible grouping, and then represent the data in a completed histogram. Students' groupings varied — some five and some ten — so different histograms were produced which led to fruitful discussion (facilitated by the teacher) when students were trying to interpret the story portrayed by the histogram. Figure 3 shows one student's histogram produced with 5 cm interval sizes (140–144, 145–149, 150–154, etc.).

A similar procedure was followed with data on the 'number of books read'. When time permitted, students were also introduced to other data sets to discuss and make decisions on groups and interval sizes without drawing the histograms.

Checking the outcomes

In the week following the three lessons, the two classes completed a task where they were given a different data set (students' vertical reach — continuous data) and instructions similar to the pre-instruction situation. They were asked to organise the data into any groups that made sense to them and then to represent the data in any type of graph they liked on the graph paper provided.

What changes had taken place?

Organisation of the data into groups (intervals) rose from 40 (70%) to 49 (86%) from the pre-instruction task to the post-instruction task. However, the most encouraging result occurred in the area of representing the data in a histogram with an increase from 4 out of 57 (7%) to 42 (74%) over the

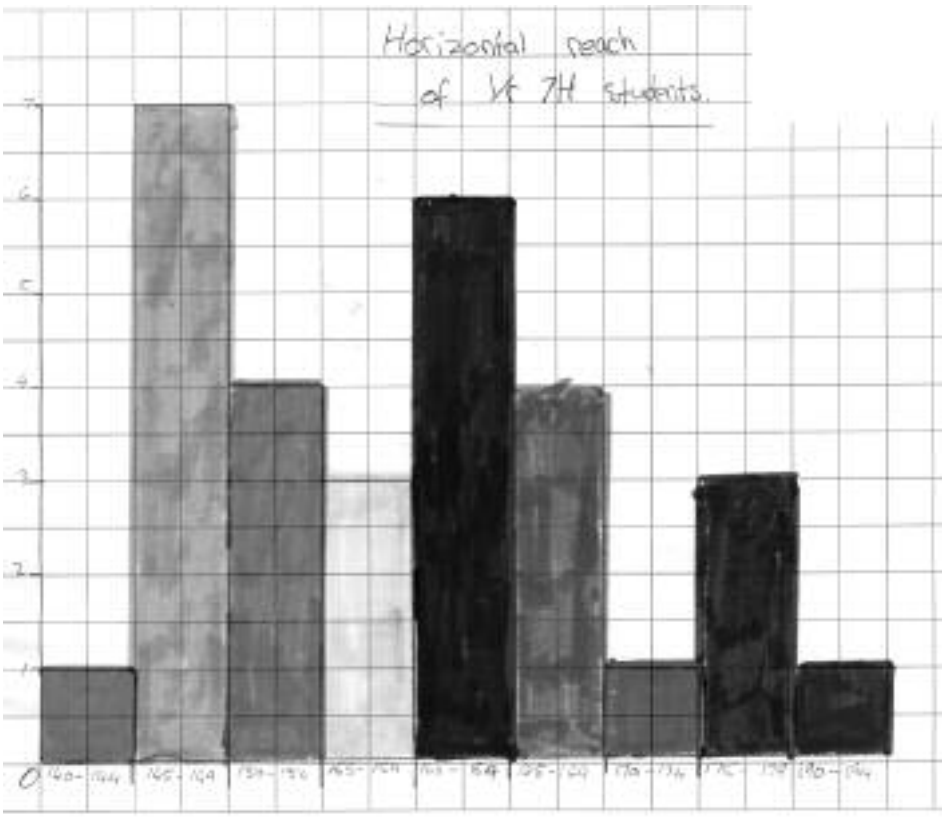

Figure 3. Histogram of horizontal reach with 5cm interval sizes.

same period. It appears that the teaching experiment of three lessons was successful in facilitating the development of skills necessary for organising and representing a large set of numerical data in about three quarters of these Year 7 students.

What did we learn?

For the majority of students, there is no real problem in organising data although some prefer to order the data rather than form intervals. What they find difficult is transforming the grouped data into a valid graphical representation. Although organisation is a necessary requirement for being able to represent the data, it is not sufficient for the task. What appears to be the stumbling block for many students is the realisation that the frequencies for each class/group generated in the data-organisation step become the new variable to be plotted on the graph and represented on the vertical axis.

Instruction did play an important role in assisting students to represent the grouped data. After the series of just three lessons, 74% of the students were able to produce valid histograms. One could imagine that with further experience over an extended period of time, most of the remaining students would be able to master these skills. Coverage of the topic by teachers in Year 7 therefore seems to be the major issue in developing such skills. It cannot be said that the topic is too difficult for Year 7 students, nor that it is inappropriate to be included in the mathematics curriculum at the Year 7 level.

When a sample of students of different ability levels was interviewed, the majority believed that the topic was not too difficult, and that it was interesting and enjoyable. By approaching the topic with a student focus (through the collection of data about the students themselves), students' interest and involvement can be gained and maintained in this topic.

References

Bright, G. & Friel, S. (1998). Graphical representations: Helping students represent data. In S. Lajoie (Ed.), *Reflections on Statistics: Learning, Teaching and Assessment in Grades K–12*. Mahwah, NJ: Erlbaum.

Department of Education, Queensland (1990). *Sourcebook for Year 7, Years 1 to 10 Syllabus in Mathematics*. Brisbane: QLD: Author.

Jones, G., Langrall, C., Thornton, C., Mooney, E., Wares, A., Perry, B., Putt, I. & Nisbet, S. (2001). Using statistical thinking to inform instruction. *Journal of Mathematical Behavior, 20*, 109–144.

Konold, C., & Higgins, T. (2000). Working with data: Highlights of related research. In S. Russell, D. Schifter & V. Bastable, Working with data. A module of the K–6 professional development curriculum, *Developing Mathematical Ideas* (In press).

Moore, D. S. (1992). *Statistics: Decisions through Data*. Videotape 1, Unit 3: Histograms & Distributions. Lexington, MA: Consortium for Mathematics and its Applications (COMAP) Inc., distributed by AAMT, Adelaide.

Nisbet, S. (2003). Getting organised: The role of data organisation in students' representation of numerical data. Paper presented to the 27th Annual Conference of the International Group for the Psychology of Mathematics Education. PME 267, University of Hawaii, USA, July, 2003.

Nisbet, S., Jones, G., Thornton, C., Langrall, C. & Mooney, E. (in press, 2003). Children's Representation and Organisation of Data. Manuscript accepted for *Mathematics Education Research Journal 15* (1).

Steven Nisbet is a Senior Lecturer in Mathematics Education at Griffith University, and a long time member of QAMT & AAMT. He has a special interest in students' statistical thinking.

Ian Putt is a Senior Lecturer in Mathematics Education at James Cook University and is a long-time member of QAMT and AAMT. Two of his special interests are teacher professional development in maths and primary students' statistical thinking.