Assessing the Statistics Knowledge of Preservice Teachers

Tim Burgess
Massey University
<t.a.burgess@massey.ac.nz>

This paper reports on aspects of preservice teachers' statistical knowledge as assessed through an open-ended problem-solving task. This type of task is considered to be a useful and valuable way of assessing the development of data sense. The level of a teacher's knowledge is known to be critical in relation to effective teaching. This study indicates that the level of statistics thinking of preservice students presents some challenges for teacher educators to address within their courses and programmes.

Teacher educators are constantly grappling with issues related to the level of mathematical content knowledge of their students who are involved in preservice teacher education. There is wide recognition that an adequate knowledge of mathematics is a necessary (but not sufficient) component of being able to teach well. If teachers have an inadequate knowledge of mathematics itself, they will leave teaching opportunities unexplored because of their own lack of mathematical sensitivity and insight (Ball & Bass, 2000). In a previous study reported by Burgess (2000), the probability understanding of preservice teacher education students was investigated and compared with the level of understanding of a group of Year 7 and 8 students. That study showed that the preservice teacher education students demonstrated many of the same types of probability misconceptions as the younger school students in spite of the extra real-life experiences involving probability that the older students would have had.

Within the area of statistics, it is widely acknowledged that students may be able to apply certain skills (such as calculating a mean when given a set of numbers) but the students' level of understanding may not match their skill level. Students must be given opportunities to present statistics as part of a problem-solving process (Garfield, 1993). The development of "data sense" (Friel, Bright, Frierson, & Kader, 1997) includes being able to generate information on which graphs and statistics are constructed and is considered to be an important part of becoming statistically literate. Friel et al also describe a statistical investigation as being a process consisting of the following components: posing a question, collecting data, analyzing data, interpreting results, and communicating the results. Aspects of the investigation process related to using graphs include what Curcio (1997) referred to as reading the data (extracting information explicitly and directly from a graph), reading between the data (combining and comparing data), and reading beyond the data (extrapolating and predicting from the data). Many of these skills and others within the statistics realm have also been classified by Wild and Pfannkuch (1999) into different types of statistical thinking. Some of the types of statistical thinking include 'transnumeration' – a dynamic process of changing data representations in a system to arrive at a better understanding of that system (such as through classification, graphical representations, or trying statistical models); acknowledgment of the existence of and decision-making in relation to variation; and integration of context knowledge, statistical knowledge and information in the data, in order to produce implications, insights and conjectures.

Students’ understanding of statistics have been investigated by a number of researchers. In one example, Chick (2000) studied a group of first-year university students who were taking a mathematics service course as part of their degree. She classified the statistical techniques used by the students and concluded that more needs to be done to ensure that students have the ability to integrate the appropriate skills and understanding in context.

The aim of this current study was to assess the statistics skills of a group of students at the commencement of a statistics module in a mathematics elective course. How would they handle an open-ended problem which did not specify particular procedures to be used? Do the preservice students have an adequate understanding of statistics so that teacher education courses can concentrate on pedagogical knowledge as it relates to statistics?

Method

In order to assess the statistics skills and understanding of group of first year preservice teacher education students, a method was needed which would give the students the opportunity to ‘get inside’ some data and investigate it. Such a problem-solving approach would not limit the students to certain, pre-determined procedures. It was decided that the ‘data cards protocol’ as developed by Watson, Collis, Callingham, and Moritz (1995) would be suitable. A number of studies have used the ‘data cards protocol’ in its original form of giving data about 16 children to the participants in the form of a set of 16 cards, with six data about each child on one card (see an example of a card in the Appendix), while other studies have provided the data in table form (such as Chick (2000, 1999)). For this study, it was decided that the original form (using 16 cards) would be more suitable for assessing a wider range of the students’ skills than if the data was already tabulated. Tabulation of the data by the students can be an important step in the process of making sense of the data.

The 30 students involved in the study included 21 who had studied mathematics through to Year 13 of secondary school and as such would have encountered more sophisticated statistical procedures than would be needed to adequately deal with the problem. The task was given as an individual one with a time limit of one hour; each student was given a set of the 16 data cards along with the written instructions for the task: “Examine the data and produce a report which highlights all aspects of the data which you think are interesting. You are free to choose your approach to this problem and what to include in your report. If you make any claims about the data, be sure to back them up with suitable evidence.”

The students’ responses were to be analysed for evidence of transnumeration. However since this covers a large number of statistical procedures, it was decided that it would be necessary to examine in more detail particular aspects of transnumeration. Using similar categories to those investigated by Chick (2000), the responses were analysed for use of the following techniques: listing of the data (both unordered and ordered), tabulation (including quasi-tables – those without labelled columns), types of graphs, statistical summaries (including mean, median, mode, range, maximum, minimum, quartiles), written statements involving frequencies (or relative frequencies or ratios), and summaries/conclusions.

Results and Discussion

Since the data was given to the students in the form of 16 data cards, the students used various ways of making sense of the data from the outset. Many arranged the cards according to various categories of interest whereas others, having laid the cards out in no particular order, started some form of written summaries such as lists and tables.

One student wrote out unordered lists of four different categories of data (namely favourite activity, age, weight, and number of fast food meals eaten per week). The student went on to create some quasi-tables of favourite activity (giving both a frequency and a percentage for each) and eye colour (frequencies), but for age and weight, calculated the mean for each from the unordered data listing. Two bar graphs were drawn, one for the frequencies of favourite activity and the other for frequencies of eye colours.

Interestingly, 16 of the students created a total of 40 quasi-tables whereas only 8 students created a total of 11 well-formed tables (with column headings and frequencies or summary data). The quasi-tables included those that used tallies but without frequencies. In fact, 14 of the tables or quasi-tables used only tallies. Only 8 students created contingency tables (some of which used tallies rather than frequencies). Two such examples were the frequencies of the weights of students categorised by gender, or the total number of fast food meals consumed by students when classified by their favourite activities. One student included a table of the total number of fast food meals consumed by students when classified by age and gender. Even though this was the most sophisticated table formed, a problem would have arisen if the student had made comparisons between the data shown in the table as there were a different number of observations (children) that contributed to each total. It would have been more appropriate to use means rather than totals; however the problem did not arise as the student did not make any concluding statement based on the table.

Of the 30 students, 21 wrote summaries or conclusions of their investigations. Although the instructions for the activity requested the students to "produce a report which highlights all aspects of the data which you think are interesting", nine students wrote no summary. Of the nine, one student drew one bar graph; another had calculated the mean age within each favourite activity and had followed this up with a bar graph of the same data. They both had performed some statistical procedures but for no apparent purpose other than drawing a graph or calculating means.

The simplest level of summaries were those which gave a written description of frequencies in the data, a written form of the mode, or something that was shown directly in a graph. For example:

- Out of the 16 children, blue was the most popular eye colour;
- The age of the students ranged from 8 years old up to 18 years old. As shown in the graph, there were 2 people aged 8, 9, 12, 17, 18 years old and only 1 person aged 10, 11, 13, 14, 15, 16. All other age groups were not represented.
- Fast food is eaten less than 5 times a week by 74% of the children. [The student's pie graph showed this directly.]

Other summaries included grouping certain data into categories and reporting frequencies or tendencies as evidenced by measures such as means or totals. For example:

- This shows that the males were heavier than the females on average. The average age was pretty close to each other while the males averaged 4kg more than the females and their top weight was also heavier than the females' top weight.
- The data [average number of take-out meals for males and for females; and number of inactive males is 5 and number of inactive females is 5] shows that people who have an active hobby eat less take-out meals, and more people are inactive than active in their hobby.
- Females, with an average age of 12.88, seem to have more of their favourite activities ones which require more stimulation of the brain shown by the keen interest in board games, reading, and some mentally tough sports, where the males seem to just laze¾watching TV being a high percentage with the exception of a board gamer and rugby players and their average age is 13 also.

Some summaries involved the students using the graph to make some interpretation that was not directly shown in the graph. For example:

- When the mean (45.75kg) and the median (47.5kg) are looked at there is not much difference between the two, but when you look at the bar graph these numbers don't mean much as the 41-50 bar is not the biggest. This is because of the extremes in the weights, with one person weighing 74kg taking the mean up and 4 people that weigh below 30kg dragging it down. The mean and the median are not very accurate because of this.

The summaries and conclusions indicate a variety of levels of understanding the data; some students have only considered one variable at a time and reported on the frequencies, while others have considered as many as three different variables and related their summaries or conclusions to the frequencies or other statistical measures within each of the groupings.

The overall level of competency with graphing the data was disappointing. Even though school exercises and activities involving graphing highlight the necessity for titles and adequate labelling of axes, many of the graphs produced by the students lacked some of these requirements. One pie graph was drawn (without a title although it did include a key). One box and whisker plot was drawn of the weights of the children but neither of that student's two summary statements referred to the graph or even to weight in general. Fifteen students drew bar graphs or histograms. Most of the histograms drawn were for data that should have been represented in a bar graph (such as the frequencies of each eye colour or the frequencies of each favourite activity).

One of the histograms (see Figure 1) had unlabelled axes but it could be determined that each rectangle in the graph represented the number of fast foods consumed weekly by one child (therefore the 16 bars or spaces); the numbers of fast food meals had been ordered and were represented on the graph as ordered data (hence the staircase effect of the graph). This student's summary included a statement about the mean number of fast food meals consumed per week, but did not refer to any aspect of the graph.

Another graph showing the total number of fast food meals as grouped by gender and age (see Figure 2) was used as the basis for the following conclusion:

- It can be seen that boys eat more takeaways in a week than girls. But it also cannot be prejudged that it is by choice, parents may be main providers.

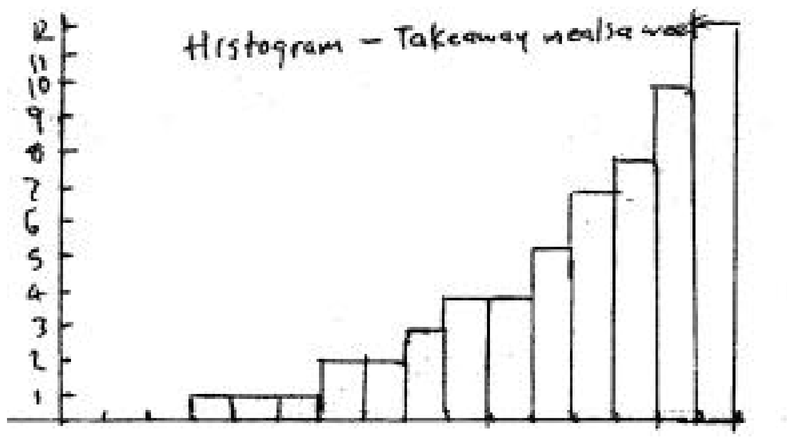

Figure 1. Example of a student's 'histogram'.

This conclusion requires totals to be formed from the boys' bars and the girls' bars and is only valid because there are the same number of boys as girls (even though the student does not state this in the 'report'). This graph is interesting as there is one bar for each child which shows the gender of the child (through the colouring of the bar) and the number of takeaways eaten (as shown by the height of the bar). How this student would have handled the situation if there had been two children of the same gender and age is not known, although it could be surmised that another bar would have been drawn. The graph does not summarise the data as such, it merely presents the individual data in graphical form.

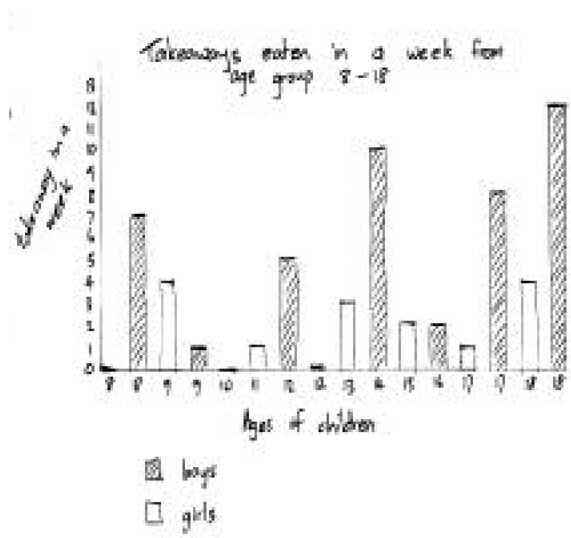

Figure 2. Example of a bar graph showing 3 variables.

One stem and leaf plot was drawn, in spite of the advantages that such a graphical display has over histograms and the fact that this type of graph is dealt with in school. Many of the 'mature' preservice students however had not encountered this type of graph at the stage that this study was conducted. Also, there were two examples of line graphs, both inappropriate for the data being represented. An example of one of these is shown in Figure 3.

Statistical summaries, mainly means, were calculated by 17 students (with many students calculating a number of such measures for various groupings or categories of data. In some reports, even though a mode was not explicitly stated, a summary included a phrase such as "the most common eye colour …". The term 'range' continues to be problematic; the common usage of the term (such as the range of the weights is from 24kg to 74kg) was given 8 times by students, and only twice was the range given as a single number.

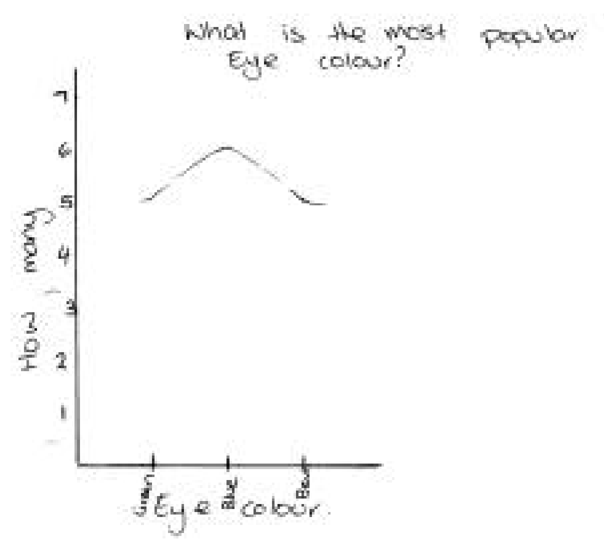

Figure 3. One example of an inappropriate line graph.

One student had headed the report with the question, "Is NZ becoming a race of fatties?" From this start, however, all that was done was a tabulation of the number of fast food meals consumed per week, a calculation of the mean number of fast food meals consumed per week, and a conclusion that focused on whether the fast food intake of young New Zealanders was too high. The conclusion finished with a 'challenge' to parents to decide whether they were putting their children's health at risk of heart disease or other weight-related diseases.

There were 12 students who completed summaries and/or conclusions based on a mixture of statistical procedures. A number focused on the relationships between two or three of the variables of weight, age, gender, number of fast food meals consumed, and favourite activity. Those who had examined 'favourite activity' generally had broken this data into two groups, namely sport and non-sport (or variations of this). They had identified a feature of the data that had not been explicitly stated, and thus showed a higher level of data sense than those who had merely used the given categories of data.

Conclusions

This study shows that there is considerable variation in preservice students' competency levels with basic statistical procedures, although with the exception of tabulation and graphing, the procedures utilized have been done so correctly (although not necessarily appropriately). Tabulation and graphing receive considerable prominence in the school curriculum, yet the students in this study displayed some unusual approaches to these procedures.

To be fully literate statistically, that is to show a well-developed data sense, the students should be able to explore the data, develop some questions to be answered, apply some appropriate procedures that will assist with the answering of the questions, analyze and interpret the results, and finally communicate their findings. In this task, students generally did not appear to be following such a sequence. They performed various procedures in isolation, as stand-alone tasks that had no further purpose. It is this aspect that raises the greatest concern. Summaries and conclusions were the main focus of the task yet many did not link these directly with the procedures they had performed earlier.

There are implications firstly for teaching statistics in school. It would appear that greater emphasis needs to be placed on the purposes for which particular procedures are performed, and making sense of the information these convey. For instance, the drawing of a graph should be used early in the statistical investigation to make observations about the distribution of the data so that appropriate measures can be chosen (such as, should the mean or the median be used for this data) rather than as an end point as it often currently perceived by students. It may be that students are being given procedures at an age that is too early for them to understand and adequately interpret the results of these procedures. Competence with the mechanical process is not sufficient for fully appreciating and knowing when to utilize the procedure.

In teacher education, there needs to be a focus problem-solving tasks, with an emphasis on the process of using procedures for a particular purpose. The entire investigation process needs to be made more explicit and given more prominence. While the focus of teacher education courses should be on the pedagogical content knowledge relevant to statistics, the preservice teachers can also be exposed to statistical content to enhance their own knowledge. For future teachers to make a difference to the statistical understanding of their students, they must be exposed to exemplary practice as regards statistics and in particular, making sense of data. Although the time available in preservice teacher education courses is limited for dealing with one content area such as statistics, the students' experiences must be realistic for what they will encounter later in their own classrooms.

References

Ball, D., & Bass, H. (2000). Interweaving content and pedgogy in teaching and learning to teach: Knowing and using mathematics. In J. Boaler (Ed.), *Multiple perspectives on mathematics teaching and learning* (pp. 83-104). Westport, CT: Ablex Publishing.

Burgess, T. A. (2000). Are teachers' probability concepts more sophisticated than those of their students? *Mathematics education beyond 2000* (Proceedings of the 23rd annual conference of the Mathematics Education Research Group of Australasia, Fremantle, pp. 126-133). Sydney: MERGA.

Chick, H. (1999). Jumping to conclusions: Data interpretation by young adults. *Making the difference* (Proceedings of the 22nd annual conference of the Mathematics Education Research Group of Australasia, Adelaide, pp. 151-157). Sydney: MERGA.

Chick, H. (2000). Young adults making sense of data. *Mathematics education beyond 2000* (Proceedings of the 23rd annual conference of the Mathematics Education Research Group of Australasia, Fremantle, pp. 157-163). Sydney: MERGA.

Curcio, F. R., & Artzt, 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* (pp. 123-138). Amsterdam: IOS Press.

Friel, S. N., Bright, G. W., Frierson, D., & Kader, G. D. (1997). A framework for assessing knowledge and learning in statistics (K-8). In I. Gal, & J. B. Garfield (Eds.), *The assessment challenge in statistics education* (pp. 55-63). Amsterdam: IOS Press.

Garfield, J. B. (1993). An authentic assessment of students' statistical knowledge. In N. L. Webb, & A. F. Coxford (Eds.), *Assessment in the mathematics classroom* (pp. 187-196). Reston, VA: National Council of Teachers of Mathematics.

Watson, J. M., Collis, K. F., Callingham, R. A., & Moritz, J. B. (1995). A model for assessing higher order thinking in statistics. *Educational Research and Evaluation, 1*, 247-275.

Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review, 67*(3), 223-265.

Appendix

One example of a data card:

Name:	Mary Minski
Age:	13
Favourite activity:	Reading
Eyes:	Green
How many fast food meals do you eat every week?	3