Strategies Used by Students to Compare Two Data Sets

Robyn Reaburn
University of Tasmania
<Robyn.Reaburn@utas.edu.au>

One of the common tasks of inferential statistics is to compare two data sets. Long before formal statistical procedures, however, students can be encouraged to make comparisons between data sets and therefore build up intuitive statistical reasoning. Such tasks also give meaning to the data collection students may do. This study describes the answers given by beginning university students to tasks involving comparing data sets in graphical form, originally designed for students between Grades 3 to 9. The results show that whereas all the students had successfully completed either pre-tertiary mathematics or a bridging mathematics course many had similar difficulties to students of a younger age. In particular, they did not use a measure of centre or proportional reasoning when appropriate.

One of the common tasks in inferential statistics is to compare two data sets. For example, is one group faster than the other group? Does the new drug work better? In the formal procedures of inferential statistics, questions similar to these are often answered by comparing the values of the arithmetic mean of each group while taking into account the value of the standard deviation of each group.

Using less formal means of making comparisons, however, students can compare two data sets by using a measure of centre such as the arithmetic mean or by using proportional reasoning. For students to use a measure of centre they need to know that this statistic is somehow representative of a group (Gal, Rothschild, & Wagner, 1990). Despite the wide spread use of the arithmetic mean (the *average*) in everyday applications, previous research has shown that students often only perceive the arithmetic mean as the learned algorithm. Because these students do not regard the arithmetic mean as a representative number they are generally unsuccessful in using it to make decisions about data (Mokros & Russell, 1995).

Gal, Rothschild, and Wagner (1989) investigated how primary students (Grades 3 and 6) compare two data sets. They found that most of the students in Grade 6 did not use the arithmetic mean in their solutions, even though they were familiar with its calculation. Many of the students used totals even when the data sets were not of equal size. They also found that many of the students in Grade 6 had difficulty in using proportional reasoning. In a later study Gal, Rothschild, and Wagner (1990) found that as students became older their understanding of the characteristics of the arithmetic mean improved but there was still a reluctance to use it as a tool to distinguish between two data sets. Whereas the formula for calculating the arithmetic mean was familiar to 2% of Grade 3, 61% of Grade 6, and 91% of Grade 9 students, the algorithm was applied by only 4%, 14% and 48% of the students respectively. They also did not generally use proportional reasoning or visual comparisons of the given graphical displays to reach their conclusions. Watson and Moritz (1998) also investigated students' thinking in comparing two data sets. In their study, 88 students from Grades 3 to 9 were given a series of tasks that required them to make comparisons between two data sets given in graphical form. Many of the students did not use the arithmetic mean in their conclusions, and those who did (10% of the Grade 6 students and 54% of the Grade 9 students) did not always do so successfully.

Another strategy in such tasks is to use proportional reasoning, which is valid when the groups are not of equal size. Proportional reasoning involves multiplicative reasoning instead of additive reasoning. For example, in answer to the question, "If green paint is

In J. Dindyal, L. P. Cheng & S. F. Ng (Eds.), Mathematics education: Expanding horizons (Proceedings of the 35th annual conference of the Mathematics Education Research Group of Australasia). Singapore: MERGA.

made from one blue and three yellow and I added two more blues, how many more yellow would I need?” many primary students answer that as two have been added to the blue, two will have to be added to the yellow (an additive strategy) instead of realising that for every one blue, three yellows are required (a multiplicative strategy) (Parish, 2010). It has been argued that children are not able to use proportional reasoning under the age of 12, but research has shown that young children are capable of using proportional reasoning if the tasks are selected appropriately (Boyer, Levine & Huttenlocher, 2008). For example, proportional reasoning problems can be as simple as deciding which group has “more” girls if there are two girls in a group of four, or two girls in a group of five (Van de Walle, Karp, & Bay-Williams, 2010).

An ability to carry out an algorithm is not necessarily accompanied by an understanding of the significance of the answer, and previous research has shown that students may mask a lack of understanding by following the required algorithm (Garfield & Ahlgren, 1988). If statistics instructors assume that students already have a familiarity with fundamental ideas such as the arithmetic mean when they do not, students’ future understanding may be compromised.

Method

Participants

The study described here was part of a wider project which, in part, assessed students’ familiarity with statistical reasoning on entry to a tertiary institution (Reaburn, 2011). The sample consisted of 75 tertiary students on entry to a first year introductory statistics unit. All the students had either successfully completed a pre-tertiary mathematics course, or successfully completed an introductory calculus bridging unit before admission to the unit. The students were asked to fill out a questionnaire in the first week of the unit, part of which was based on the tasks in the study by Watson and Moritz (1999). There were four tasks in the section described here, and these were ordered in what was considered to be of increasing difficulty.

The Tasks

For these tasks the students were to compare the scores of four pairs of groups based on data presented in graphical form (Figure 1). For the first pair of groups (Task 1) the groups were of equal size and all of one group had higher scores than the other. For the second pair (Task 2) the two groups again were of equal size and although there was some overlap in the scores, one group clearly had a higher mean score than the other. In the third pair (Task 3) there were equal numbers in each group and the means, medians and modes were equal, but one group had a wider spread than the other group. For the final pair (Task 4) the group numbers were not equal, and it was expected that students would have to make a judgement using the value of the arithmetic mean or median or by using proportional reasoning. The introduction to these four tasks was.

> A tertiary institution is comparing the scores of some tutorial groups on a test of basic statistics facts. The test had nine questions. The scores for two of these tutorial groups are shown in the charts below. Each circle represents one person. Therefore for Group A four people answered two questions correctly, and two people answered three questions correctly.

After each pair of graphs the question was asked.

> Did the two groups perform equally well, or did one group perform better? Please give reasons for your answer.

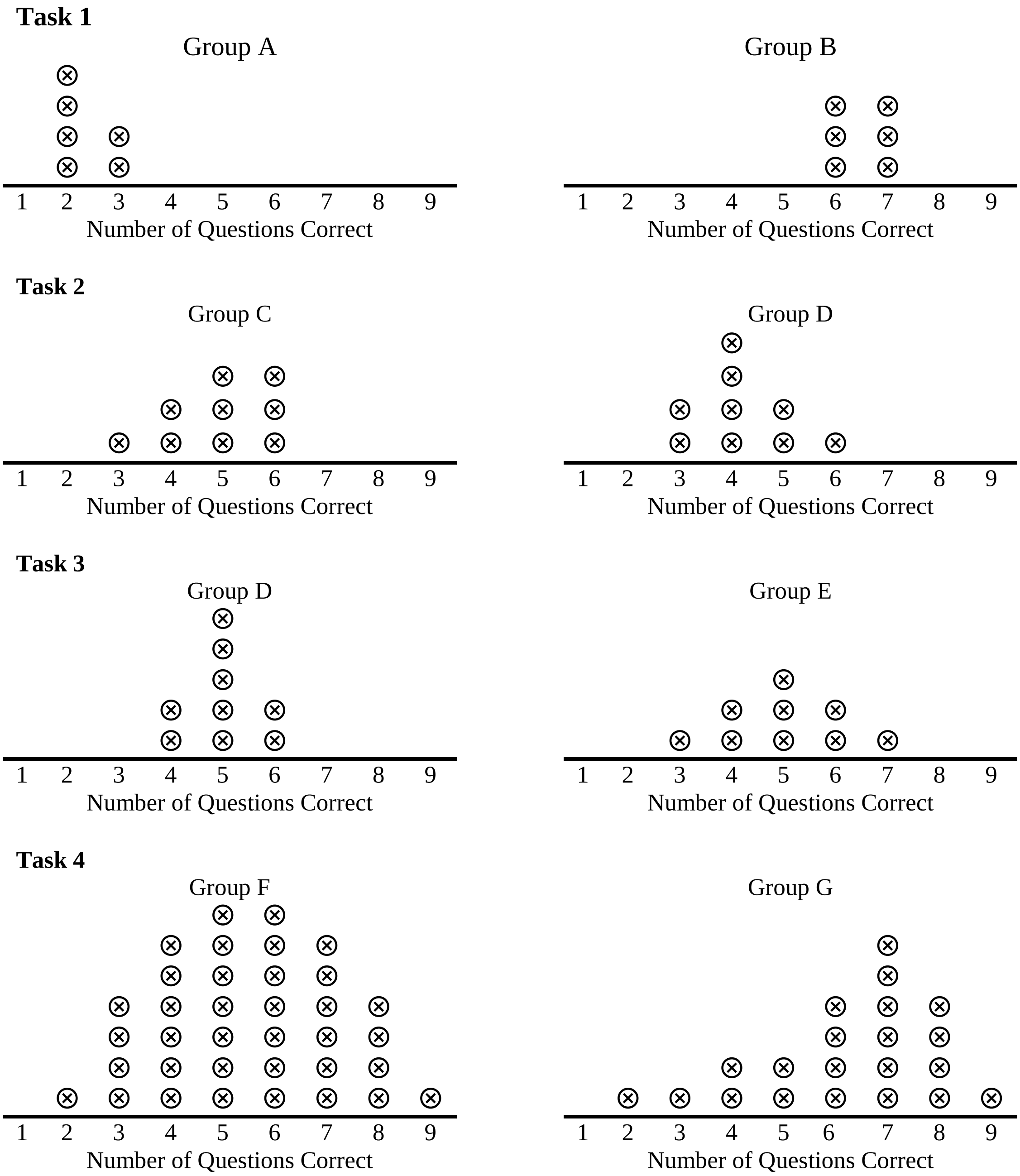

Figure 1. The tasks given to the students. For each pair of graphs, the students were required to state which group “performed better” and why.

Results

Task 1

Table 1 indicates that the most common response to this task was to state that there were “more” correct answers in Group B, with no further justification. The next most common response was to indicate that the entire group in B had higher scores than the entire group in A. Of the students who used the arithmetic mean and/or median, four calculated the arithmetic mean fully without using an approximation. The other students who answered Group B had performed better stated directly what the scores were, for example "Group A has 2s and 3s and Group B has 6s and 7s." A small number of students stated that Group A had performed better.

Task 2

Table 2 indicates that again the most common response was to state that there were "more" correct answers. Some students used the arithmetic mean and/or the median to justify their answers and others explicitly compared the scores. Several students used totals in their justifications, half of whom stated that the group sizes were equal.

Table 1

Answers Given by the Students to Task 1

Answer	Number n = 75	Percentage (Rounded)
Group B		
There are "more" correct answers in Group B	23	30
Everyone in Group B had higher scores than Group A	15	21
Arithmetic mean and/or median used	10	13
Used the scores, e.g. "Group A has 2s and 3s and Group B has 6s and 7s."	15	21
Totals calculated and stated that group sizes are equal	1	1
Group A	4	5
No response	7	9

Task 3

In this task 47% of the students stated that the two groups performed equally well. Approximately half of these used the mean and/or median in their justifications, and the others used the totals. The most common reason given by those who designated that one of the groups performed better was that "more" had higher scores. This reason was used by those who selected both Group E and Group F as the better performer.

Table 2

Answers Given by the Students to Task 2

Answer	Number n = 75	Percentage (Rounded)
Group C		
There are "more" correct answers in Group C	31	41
Arithmetic mean and/or median used	8	11
Used the scores, e.g. "Group C has one 3, two 4s, three 5s, and three 6s. Group D has two 3s, four 4s, two 5s and one 6."	15	21
Totals calculated and stated that group sizes are equal	3	4
The groups were equal	7	8
Group D	3	4
No response	8	11

Table 3
Answers Given by the Students to Task 3

Answer	Number n = 75	Percentage (Rounded)
Equal		
Arithmetic mean and/or median used	19	25
Totals calculated and stated that group sizes are equal	10	13
No explanation	7	9
Group E		
More consistent	3	4
More got higher scores	13	17
Arithmetic mean and/or median used	2	3
No explanation	6	9
Group F		
More got higher scores	3	4
No explanation	3	4
No response	9	12

Task 4

Table 4 shows that a higher proportion of students used a calculation or estimate of the arithmetic mean and/or median to answer this task than the previous tasks. Of the other students who stated that Group H performed better, 13 explicitly used proportional reasoning, one implied proportional reasoning, and five did not give an explanation. Of the 20 students who stated that Group G performed better seven said that there were more people, therefore giving a “better” score. Seven others did not give an explanation. Six students said the two groups performed equally well and five said that the problem was “not fair” or could not be done.

Table 4
Answers Given by the Students to Task 4

Answer	Number n = 75	Percentage (Rounded)
Group H		
Higher arithmetic mean and/or median	20	27
Proportional reasoning	13	18
“More” got higher scores	1	1
No explanation	5	7
Group G		
More people therefore this group performed better	7	9
More in the higher range	4	5
More balanced	2	3
No explanation	7	9
Equal	6	8
Too hard/cannot be done	4	5
Not fair	1	1
No response	5	7

Discussion

For Task 1 a visual inspection quickly shows that all the scores in Group B are higher than those in Group A. A calculation of the arithmetic mean or median is not required to answer this task and most of the students used strategies that did not require such calculations. In Task 2, where there was some overlap of the scores, a higher proportion of the students used a full calculation or estimates of the arithmetic mean and /or median. A number of students also used totals, which is an acceptable strategy when each group has the same number of scores.

Task 3, judging by the relatively large number of students who gave no explanation, appeared more difficult. The question of what was meant by "better" in this context was deliberately left open, and a small number of students used the criterion of consistency to make their judgments. There were students who used the reasoning of "more with higher scores" to choose both Groups E and F. Of particular concern is the small number of students who incorrectly calculated the arithmetic mean for these groups and therefore came to the conclusion that Group E performed better than Group F. The arithmetic mean, median and mode are all equal in this task, and this should have been apparent using visual inspection. This suggests that these students did not have a conceptual understanding of these statistics. It is also apparent that the students who selected one of the groups as better (apart from when they used the criterion of consistency) were not aware that the arithmetic mean could also be considered as a balancing point, a strategy used successfully by some of the students in Grades 6 and 8 in the study by Mokros and Russell (1995).

Task 4 was deliberately set so that more sophisticated reasoning would be needed in making a judgement than in the previous tasks. An indication of the difficulty is shown by the increased proportion of students who did not give explanations for their answers. Of the 39 students who correctly answered Group H, 20 used the arithmetic mean or the median and 13 used proportional reasoning, either directly or implied. Close to one half of the students did not answer this question correctly. Of particular concern is the number of students who argued that as there were more people in one group then this group performed better. These students did not think to use even the simplest form of proportional reasoning available to young students (Boyer, Levine & Huttenlocher, 2008). All of these students would have been previously exposed to proportional reasoning problems. Also of concern was the number of students who stated that the task could not be done or was not fair. This form of reasoning was used by students of Grade 7 and under in the study by Watson and Moritz (1999) and it is disturbing to find this reasoning used by students who have successfully completed a pre-tertiary mathematics. It is not likely that any of these students were unfamiliar with the algorithm to calculate the arithmetic mean.

This study has found that problems noted in younger students may persist beyond the end of secondary education, even by those who have studied mathematics until the end of their schooling. Some students did not use the mean or median even when this would have helped answer the question (Gal, Rothschild, & Wagner, 1989, Watson &Moritz, 1999). This study also adds to previous research by Mokros and Russell (1995) who found that many students only see the measures of centre as the algorithms used to calculate them, and do not see these statistics as representative numbers. Since this study the researcher has introduced a question early in the introductory statistics unit asking the students to explain the meaning of the statement "The average score for a class test was 74" without explaining how the statistic is calculated. Generally the students have great difficulty in answering.

Some Implications for Instructors

These results show that students can use a variety of strategies to answer problems and can select the quickest strategy in each context. They might not do so, however, if instructors are not open to allowing students to choose and compare different strategies in the classroom. More importantly, this study alerts instructors that students may competently use algorithms but have little or no understanding of the significance of the results of their calculations. Mokros and Russell (1995) have suggested that instructors who rely too heavily on the algorithm in their teaching may, in fact, inhibit students' understanding. Therefore students not only require practice in using algorithms, but also require practice to put the results of these algorithms into a context. The Australian Curriculum (Australian Curriculum Assessment and Reporting Authority [ACARA], 2011) states that students should draw back to back stem-and-leaf plots in Year 9 and draw box plots in Year 10. It is not until Year 10A, however, does the curriculum explicitly state that comparisons between data sets should be made. Perhaps such comparisons should be used in earlier years to help develop statistical reasoning and to help give meaning and interest to the data students are given.

References

Australian Curicculum, Assessment and Reporting Authority. (2011). *The Australian Curriculum: Mathematics.* Version 2.0, 17 October, 2011. Sydney: ACARA.

Boyer, T., Levine, C., & Huttenlocher, J. (2008). Development of proportional reasoning: Where young children go wrong, *Developmental Psychology, 44*(5), 1478-1490.

Gal, I., Rothschild, K., & Wagner, D. (1989). *Which group is better? The development of statistical reasoning in elementary school children.* Paper presented at the Meeting of the Society for Research in Child Development, Kansas City, MO.

Gal, I., Rothschild, K., & Wagner, D. (1990, April). *Statistical concepts and statistical reasoning in school children: Convergence or divergence?* Paper presented at the Meeting of the American Educational Research Association, Boston, MA.

Garfield, J., & Ahlgren, A. (1988). Difficulties in learning basic concepts in probability and statistics: Implications for research. *Journal for Research in Mathematics Education, 19*(1), 44-63.

Mokros, J., & Russell, S. (1995). Children's concepts of average and representativeness. *Journal for Research in Mathematics Education, 26*(1), 20-39.

Reaburn, R. (2011). *Students' understanding of statistical inference: Implications for teaching* (Doctoral dissertation). Australia: University of Tasmania.

Van De Walle, J., Karp, K., & Bay-Williams J. (2010). *Elementary and middle school mathematics: Teaching developmentally* (7^{th} Ed.). Boston, MA: Allyn & Bacon.

Watson, J., & Moritz, J. (1998). The beginning of statistical inference: Comparing two data sets. *Educational Studies in Mathematics, 37*(2), 145-168.