STUDENT PERCEPTIONS OF VARIATION IN A SAMPLING SITUATION

Chris Reading Centre for Cognition Research in Learning and Teaching, University of New England, Australia	Mike Shaughnessy Department of Mathematical Sciences, Portland State University

Variation is essential to the study of statistics, but may be a neglected topic in school mathematics. Although students may recognize that variation will occur in a sampling situation they have difficulty in discussing reasons for this variation. Three forms of an item were trialed in clinical interviews with students in grades 4 - 12. The interviews probed for students' understanding of centres and spreads which can occur in the resulting sampling distribution for repeated trials of an experiment. This paper contains four case studies of students' reasoning during the interviews.

Introduction

Despite the relative importance of variation to the study of statistics (Wild & Pfannkuch, 1999), there does not appear to have been as much research into students' understandings of measures of dispersion as into their understandings of measures of central tendency. One reason suggested by Shaughnessy (1997) for this deficiency is that research often mirrors the emphasis in curricula material, which tend to focus on measures of central tendency and neglect a careful development of measures of spread. In addition, teachers often avoid teaching spread because they do not wish to introduce the procedurally messy notion of standard deviation.

Given that more students tend to use measures of central tendency than measures of dispersion to reduce and describe sets of data (Reading, 1996), more needs to be uncovered about students' understanding of variability. Despite a lack of questions involving variation in the 1996 National Assessment of Educational Progress (NAEP) in America, one extended response item on sampling that investigated students' reasoning about centres did provide an opportunity for students to discuss spread (Shaughnessy, 1999, p. 2). However, only one student in a sample of 250 responses even raised the issue of spread. This led us to redesign the task to trigger responses that would help find out more about students' understanding of variability.

Research Design

A written task (see Figure 1) similar to the NAEP item, based on sampling from a bowl containing 100 lollies, was piloted with students in schools in Australia and America (Shaughnessy et al., 1999). The results reinforced many phenomena noticed in the NAEP responses but interviews were considered necessary to help explain students' responses. Initially, twelve students were interviewed in Australia, six primary and six secondary. This paper reports on 4 of the interviews, one from each of the grades 4 (Millie), 6 (Jess), 9 (Jane) and 12 (Max).

Students were asked to respond to two different sampling situations: a mixture with 50 red, 30 blue and 20 yellow and another with 70 red, 10 blue and 20 yellow. An actual bowl, containing the correct proportions of wrapped lollies (in some cases coloured plastic bears), was placed in full view. Students were told the lollies were well mixed up. In each case, students were asked how many red lollies could be expected in a handful of 10 lollies. They were then asked to report on the number of reds which would be drawn by six people in a handful of 10 lollies. The lollies were returned to the bowl after each draw and thoroughly re-mixed.

Student Response Form

1A) Suppose we have a bowl with 100 lollies in it. 20 are yellow, 50 are red, and 30 are blue.
Suppose you pick out 10 lollies. How many reds do you expect to get? ___ Would this happen every time?
1B) Altogether six of you do this experiment.
What do you think is likely to occur for the numbers of red lollies that are written down? Please write them here. ______, ______, ______, ______, ______, ______
Why are likely numbers for the reds?
1C) Look at these possibilities that some students have written down for the numbers they thought likely.
Which one of these lists do you think best describes what might happen? Circle it.
a) 5,9,7,6,8,7
b) 3,7,5,8,5,4
c) 5,5,5,5,5,5
d) 2,3,4,3,4,4
e) 7,7,7,7,7,7
f) 3,0,9,2,8,5
g) 10,10,10,10,10,10
Why do you think the list you chose best describes what might happen?

1D) Suppose that 6 students did the experiment— pulled out ten lollies from this bowl, wrote down the number of reds, put them back, mixed them up.
What do you think the numbers will most likely go from? From ______ (low) to ______ (high) number of reds.
Why do you think this?
**(After doing the experiment) Would you make any changes to your answers in 1B-1D? If so, write the changes here.
1E) Suppose that 6 students each pulled out 50 lollies from this bowl, wrote down the number of reds, put them back, mixed them up.
What do you think the numbers will most likely go from this time?
From ________ (low) to ________ (high) number of reds.
Why do you think this?
1F) Suppose that 40 students pulled out 10 lollies from the bowl, wrote down the number of reds, put them back, mixed them up. Can you describe what the numbers would be, what they'd look like?
Why do you think this?
.... These questions were then repeated for the 70 red situation.

Figure 1 - Student Response Form (Condensed)

Responses were sought in three different forms, LIST (write the number of reds in each handful), CHOICE (choose one of seven different multiple choice options) and RANGE (give the lowest and highest number of reds). Students were also asked to explain their responses and then given the chance to alter any of their responses after having actually drawn six samples of 10 from the bowl. The extended questions, parts E and F of the response form (Figure 1), were only presented to Year 12.

Responses were coded on two dimensions, centre and spread. Centres were coded as LOW, MEAN-CENTRED or HIGH depending on the central tendency. Spreads were coded as NARROW, REASONABLE or WIDE according to the dispersion. This coding (Shaughnessy et al., 1999) was based on information obtained from student responses to a written version of the task and was used to build a profile of each student's perception of the sampling situation. For example, a response 5 7 9 8 6 5 would be coded as HIGH-REASONABLE for the 50-red mixture, since the centre of the student's response was high compared to the expected mean (5 in this case) and the range of the numbers of reds pulled was neither really narrow nor really wide. Similarly, 5 5 5 6 5 5 would be coded MEAN-CENTRED but WIDE and 1 3 5 7 9 10 would be coded MEAN-CENTRED and WIDE.

The interviews were designed to compare students' responses across the three versions of the tasks, and to investigate students' perceptions of variation. More specifically, the following questions we investigated: Would one form of the question LIST, CHOICE or RANGE give most information about students' conceptions of variation? How do students' describe variation in a sampling situation? What reasons will students give for their responses? Does the proportion of the colours influence the response? How does actually conducting the sampling experiment alter the responses? Are the students' responses consistent across versions of the task? What is the students' overall perception of the sampling distribution for this task? Can they attend to a range of likely outcomes, as opposed to focusing on single outcomes, that is, will they consider spread as well as centre in their responses? The four case studies will be considered, and then we will summarize student thinking on these questions. Directly quoted comments from students are presented in italics.

Millie

Millie suggested one handful, for the 50-red situation, would yield 3 or 4 red and appreciated that it should be different each time. Her LIST (5 2 3 6 4 3) and CHOICE (2 3 4 3 4 4) responses were also low. Her explanation lots of other colours as well, suggests that she was not considering that there were 50 red, only that she would expect to get some of each of the 3 colours. It is possible that

Millie was thinking according to an 'equal probability' bias, or that she thought each colour would come out with the same 'fair' chance. Millie did not want to choose the option where all numbers were the same, indicating an appreciation of the variation of outcomes. Millie gave responses with reasonable spread. Her RANGE (3 to 6) suggested a mean-centred response but when details of the spread are considered across the three forms of the response she was responding quite low. Millie decided not to change her responses after doing the experiment (1 3 5 4 6 4) because, although the experiment yielded a 1, she decided against including a 1 as she may not get a 1 next time.

Although Millie still chose 4 as the number of reds for the 70-red situation, she did LIST (3 6 5 4 4 7) some higher numbers, because there were more reds than before. She was still not able to focus in on the number of reds, just comparing it to the previous 50-red situation. Her CHOICE (3 7 5 8 5 4) and RANGE (3 to 7) responses were reasonable in spread but too low. As before, Millie did not want all numbers to be the same. After completing the experiment (7 5 5 7 7 7) Millie changed one 4 to a 6 and the other to a 5. The experiment has suggested to her that her response was too low, however she still was not concerned about the 3 she gave.

Generally speaking Millie's responses are too low. Millie did not explicitly focus on the number of reds in the mixture, and never considered the relative ratios of the colours. Although her ranges are reasonable she appears to consider that variation will occur evenly over her preferred range. Millie generally had difficulty in explaining why she chose the responses that she did.

Jess

Jess expected more red as one can see more red and despite claiming that the exact number of reds did not matter, she predicted 5 in one handful for the 50-red situation. Her appreciation for variation was evident when she said it would not be the same every time. Jess gave mean-centred responses for CHOICE (3 7 5 8 5 4) and RANGE (3 to 8) but her LIST (5 4 6 3 4 8) was low. All three responses were reasonable in width. Jess chose her CHOICE response because it was more spread out, meaning that she didn't want the same number each time, rather than wanting to include all of the numbers (as one would expect 'more spread out' to mean). In fact, Jess stated that the size of the number didn't matter. Then, when explaining that extremes, such as 1 or 2, were not possible her reason was that there are 50 red. After performing the experiment (5 6 4 5 8 3) Jess chose to change the 2 to an 8 in her LIST, making the response mean-centred rather than low. Her reason was that they were all mixed up, but she did go on to say that it was hard to explain.

Jess suggested that 9 may come out in the handful for the 70-red situation, although she did suggest 7, 8 or 6 for next time. Jess gave mean-centred and reasonable width responses for the LIST (9 7 5 6 8 9), CHOICE (5 9 7 6 8 7) and RANGE (5 to 9). Her explanations basically wanted higher numbers and, although mentioning the 70 red, she just said that numbers should be picked in the higher half of 5 (meaning 5 to 10). Jess did not want to perform the experiment this time.

Jess had a better feel for the sampling distribution than Millie, having both mean-centred and reasonable spread responses in both sampling tasks. However, Jess was still not able to give explicit reasons for her responses in terms of the colour proportions and any variation she indicated was fairly uniform in manner.

Jane

Jane suggested 5 in a handful for the 50-red situation and then said that 5 could happen again. However, she also stated that it could be something else, even allowing that probably could get all reds. Although her LIST (5 7 6 2 4 3) was mean-centred, her RANGE (4 to 10) and CHOICE (5 9 7 6 8 7) responses were high. All three had reasonable spread. She could not clearly explain why she gave the LIST and RANGE responses but made the CHOICE response because she wanted some mixed up and two doubles. Jane felt that someone might get doubles. This indicates that Jane was not just comfortable with the idea of the same number appearing more than once, but seemed to think that it was necessary. The response probably not enough for all six of them to pull out 5 red, was confusing as the actual numbers Jane gave were higher which would require even more total red. She may have forgotten that the lollies were replaced each time. Jane's responses are representative of many students (on the written version) who responded with the total number of reds drawn for all six handfuls, rather than one handful at a time. After performing the experiment (7 5 6 6 3 6) Jane made no changes and justified her actions by matching up, number for number, the experimental results with her responses.

Jane suggested that 10 would be drawn out for the 70-red situation because there's heaps of reds in there and in fact when writing her LIST (10 7 8 10 9 10) she included 10 many times. The RANGE (9 to 10) also focused on the very high values. However, with the CHOICE (5 9 7 6 8 7) Jane gave a more mean-centred response, even though it was not consistent with the other responses. She had avoided the all 10 choice though because she felt that the numbers could not all be the same. Having completed the experiment (8 9 6 6 5 8) Jane changed the RANGE to 5 to 10.

Jane seems to fixate on the large number of reds in the bowl, especially in the second version of the task. She has a tendency to give high, though reasonable spread, responses. Choosing 5 for the 50-red situation suggests that she had some appreciation of the 50% red mix even if she couldn't explain her response. However, her ratio concept is not solid, since the fact that she estimates 10 for the 70 red situation suggests that she views this as a 'more than 50%' situation rather than taking note of the 70% mix. Jane also appears to better represent the spread of the values for the 50 rather than the 70 red situation, suggesting again that she is more comfortable with the 50% mix.

Max

Max suggested getting about half red in the 50-red situation but gave the complicated ratio of 2.5 to 1 as the ratio to yellow when explaining his response. However, he kept his options open by adding but you never know you could have a bit of luck one time. His LIST (3 4 5 5 6 4) together with his reason I tend to think that it could be averaged around half each time - so around there somewhere, but I think it would go over a number occasionally just because of luck basically indicated that he both appreciated, and could explain, the influence of half the lollies being red. The RANGE (3 to 7) and its justification, the average, around in the middle, about half way also demonstrated this. Both responses are mean-centred and reasonable width and indicate a good appreciation of the 50% mix and variation amongst the outcomes.

Initially, the CHOICE (5 5 5 5 5 5) response was based on I think there would be more of an average, 5 each time but then Max decided to change to 3 7 5 8 5 4 because he felt that the chances of getting 5 every time are not high but the average would need to be around 5. Max is torn between wanting the long term average to be 5 and also wanting to demonstrate variability. This demonstrates interference from probability instruction, in which the focus is on the chance of single occurrences of an outcome (probability of an event or most likely event), rather than on a reasonable range of outcomes (the notion of a distribution). After completing the experiment (4 5 3 4 6 5), Max was more convinced that his CHOICE should be changed and wanted something more spread than all 5s, although he worried that 8 was too high. Max was also presented students pulling out 50 lollies in each handful. He had trouble giving a RANGE response, changing from 12 to 17 to 17 to 20 and then 20 to 26. At first guessing, he eventually arrived at 20 to 26 after realizing that half of the lollies were being pulled out. However, Max was not able to articulate that half of 50 was 25.

When asked how many in one handful for the 70-red situation, Max replied at least 6 but chance it could go to 10, already indicating that variation was possible when asked for a single response. However, his reason only mentioned many red, not the proportion. He did include the need to be mixed really well, though. All three responses, LIST (5 8 7 6 6 8), CHOICE (5 9 7 6 8 7) and RANGE (5 to 9) were mean-centred and reasonable in width. However, his reasons generally reflected a need for the numbers to be at least 5. It was not clear why 5 was chosen. After the experiment (8 4 6 7 7 9) Max considered changing a 5 to a 4 but then decided that the 4 was just luck anyway and if the experiment was repeated it may be a 5 and then he would have to change again. Finally, he decided that if there was a chance of getting 4 then it should be included. Now, when presented with selections of 50 lollies Max gave 25 to 40 as the RANGE with the reason that definitely have to go over half because you have over a half of the reds. Then, when considering that there are in fact 70% red decided 50 red was possible.

When confronted with the extended problem of 40 students pulling out 10 lollies each, Max described the results as follow the same pattern. He decided it would be more spread, around 4 to 9 with the same number of 5, 6, 7, 8 and 9 appearing and the average around 7 or something. He even chose to allow for the fact that he may get 2 or 3. Interestingly, Max allowed for the fact that there could be more variation but was still keen to allow equal occurrence of numbers.

Unlike the younger students, Max was able to articulate the effect that the proportion of reds has on the outcomes and he demonstrated a good appreciation of the variation. On the other hand, Max exhibited some interference with his probability concepts in this task. Max at first chose 5 5 5 5 5 5 as the result for the sampling task, and also said that it would approach 5 as an average in the long run. These are answers to different questions than the ones being asking in the task.

Discussion

There appears to be a steady growth throughout grades 4 to 12 in the ability of students to explicitly describe the sampling situation on the centring scale, with language that starts referring to the 'number of red' from Jess (Grade 6) to some explicit use of ratio by Jane (Grade 9) and finally to Max's (Grade 12) explicit use of the 50% and 70% ratios in the two tasks. However, on the spread scale performance is somewhat oscillatory, and responses may not be consistent across versions of the task (see Jane, for example). Particularly obvious was the Year 12 student's initial tendency to give narrow responses. As mentioned above, this could be due to some interference fom instruction in probability, and to lack of instruction in sampling distributions. This tendency has been confirmed in the analysis of the written version of the tasks (Shaughnessy et al., 1999). Max did appear to 'learn' as the interview proceeded, and his later responses indicated a more 'reasonable' spread. Students appear to be more capable of giving explicit reasons for their responses when describing the central tendency of the results of the sampling situation than they are for describing their reasons for dispersion. Mostly, students would say 'well, they won't all be the same'. Except for Max, there was no discussion that the results should cluster around the expected mean number of reds in the mixture.

These students tended to be consistent in their responses across the three forms of the question. However, the LIST form of the question appears to give more information about variability in the sampling situation than the CHOICE or RANGE versions. Generating their own set of outcomes, allows students to demonstrate more about their implicit concept of a sampling distribution for a small number of trials, than when possibilities are suggested to them (as in the CHOICE version). Generally, these students seemed to prefer the RANGE version of the task, as they said it was easier. However, this simplicity meant that it did not give as much information about students' expectations of mean-centredness and spread, as either of the other two forms. The CHOICE form of the question was not so useful as a measure for two reasons. First, it did not allow enough options for those who had a tendency to choose wide. This could be remedied by adding more options but then the question would become unmanageable. Second, students would not have been able to demonstrate their preference for more even spread of the possibilities over their preferred range if they had been restricted to just answering a CHOICE question.

It may be hard for students to describe variation with only six handfuls. For example, if the student decided that a suitable range for the 50-red situation is 3 to 7, then he or she tended to try and cover the whole range when giving the six numbers to show that each is possible. This left little scope for indicating variation, usually giving the impression that the student expects an even distribution of outcomes. Hence the 40-student question, added for Max, may give more information. However, even with the scope of 40 numbers, Max still seemed to expect an even distribution. The three younger students all justified the range chosen by giving reasons why extreme values should not be included. It was only Max who discussed average when justifying the range given, wanting numbers around in the middle. In a revised version of the sampling task now being used with secondary students, we ask them to imagine repeating the task 100 times, and then to draw a histogram for the frequency of the number of reds pulled (labeled axes given). It would perhaps be even better to then include a computer simulation of the sampling task, so that several distributions of 100 trials could be quickly generated, and ask the students what they think, and if they would change their own graphs.

When indicating variation students do not use specific words such as 'vary', 'deviate', 'fluctuate' or 'variation'. This observation is consistent which the results observed for the written responses (Shaughnessy et al., 1999). Millie and Max, however, make use of the expression 'more spread'. Millie meant she just wanted different numbers, that is, not the same number every time. Similarly, Max gave this as an explanation of why he no longer thought his first choice of the all 5s option was suitable. Although students do give some indication of the variation that they expect by choosing particular numbers, there is very little 'discussion' of that variation which takes place.

The students' justifications for responses are interesting in terms of the reasons they gave for the range. For low range estimates students often indicated a preoccupation with totals, while for high range predictions students were usually concerned with the large number of reds rather than the proportion of reds. When the Year 12 student, Max, chose the all 5 option sampling issues may have become confounded with calculating probabilities.

Justifications also suggest that some students try to 'explain' the 'unexplainable' by finding reasons for the 'variation' within the set of numbers given. If the experimental results are not self-confirming with their predictions, students try to make them so with their explanations, which might include references to 'variables' such as the size of the hand, how well the lollies are mixed and the position selected from in the bowl. There also appears to be a need to 'be right' in predicting the results of the sampling. Students sometimes 'think hard' when making their choice on the LIST version of the task.

The proportion of colours in the bowl does appear to effect a student's ability to predict the outcomes. With so many student experiences involving the notion of a half it is not surprising that the younger students cope better with the 50% mix than the 70% mix. Even though mean-centred, reasonable-width responses can be given, and often explained, for the situation where half the lollies are red, as soon as the proportion becomes unbalanced, with 70% red, students only seem able to deal with situation as being 'greater than 5'. Also, it appears to be more difficult for students to justify their responses with the 70% mix.

Conclusion

On existing evidence it appears that the LIST form for the question is the most useful for getting students to describe the sampling situation. However, it may be that the situations with many more people, such as 40, for the experiment will most likely encourage students to engage in a discussion of possibilities which will include consideration of the variation. Students improve with age in their ability to describe the sampling situation but are not able to articulate well the reasons for their responses. The uniformity of the variation that they expect suggests that students may need to experience more such sampling situations to better appreciate the possible variation. Also, more experiences need to be presented to students which involve proportions other than 50% mixes.

References

Reading, C. (1996). *An Investigation into Students' Understanding of Statistics*. Unpublished Doctoral Dissertation, University of New England, Armidale, Australia.

Shaughnessy, J. (1997). Missed opportunities in research on the teaching and learning of data and chance. In F. Biddulph & K. Carr (Eds.), *People in Mathematics Education*, (Part I, pp. 205-237). Dordrecht, The Netherlands:Kluwer.

Shaughnessy, J., Watson, J., Moritz, J. and Reading, C. (1999). *School Mathematics Students' Acknowledgment of Statistical Variation*. Paper presented at the 77th Annual Conference of the National Council of Teachers of Mathematics, April, San Francisco, California.

Wild, C. and Pfannkuch, M. (1999). Statistical Thinking in Empirical Enquiry. *International Statistical Review*, vol. 76, no. 3, December.