TEACHERS' REASONING ABOUT STATISTICAL VARIATION: IMPLICATIONS FOR TEACHING AND RESEARCH

Sashi Sharma

University of Waikato, Hamilton, New Zealand

Many statistics educators claim that variability plays a central role in statistical thinking. For instance, Moore (1990) puts variability at the heart of the process of statistical thinking and describes the needs of statistical thinkers to acknowledge the omnipresence of variation and to consider appropriate ways to quantify, explain and model the variability in data. Ministry of Education (2004) states that since the idea of probability as long-run relative frequency needs to be addressed with students, variation can no longer be avoided. Although it has been argued that variability plays a fundamental role in students' understanding and application of statistics and chance little research attention has been given to these concepts (Ben-zvi and Garfield, 2004; Shaughnessy, 1997; Shaughnessy, Watson, Moritz, and Reading, 1999).

Additionally, since the success of any curriculum innovation ultimately depends upon teachers, they need a deep and meaningful understanding of any mathematical topic they teach. Heaton and Mickelson (2002) argue that if statistics education is to be addressed seriously in elementary education, specific focus needs to be placed on the learning of teachers. They add that we cannot attend to children's understanding of statistics without simultaneously attending to teachers' understandings. Teacher education programmes in New Zealand do not require a course in statistics for primary education majors. Furthermore, whatever probability and statistics knowledge teachers have acquired at secondary or university was not usually taught in a way designed to develop understanding or critical thinking. While the teachers may be able to use statistical techniques to solve statistical problems, they may not possess the knowledge and the abilities for developing adequate statistical thinking in their students. There appears a need to collect data from teachers at both the pre-service and in-service levels regarding their conceptions about statistics. At the pre-service level, this information will help teacher educators develop courses which confront statistical misconceptions and beliefs and sensitise the future teachers to the alternative conceptions they can expect to encounter in their students.

RESEARCH ON STATISTICAL VARIATION

To illustrate the undue confidence that people put in the reliability of small samples, Tversky and Kahneman (1974) gave the following problem to tertiary students:

Assume that the chance of having a boy or girl baby is the same. Over the course of a year, in which type of hospital would you expect there to be more days on which at least 60% of the babies born were boys?

(a) In a large hospital

(b) In a small hospital

(c) It makes no difference

Most subjects in Tversky and Kahneman's study (1974) judged the probability of obtaining more than 60% boys to be the same in the small and in the large hospital. According to Tversky and Kahneman (1974) the representativeness heuristic underlies this misconception. People who rely on the representative heuristic tend to estimate the likelihood of events by neglecting the sample size or by placing undue confidence in the reliability of small samples. However, the sampling theory entails that the expected number of days on which more than 60% of the babies are boys is much more likely to occur in a small hospital because a large sample is less likely to stray from 50%. Shaughnessy (1997) provides evidence that students may actually superimpose a sampling setting on a question where none is there to begin with, in order to establish a centre from which to predict. For instance, consider the following task given to a sample of tertiary students at the beginning of a class in statistics:

A fair coin is flipped 5 times in succession. Which do you feel is more likely to occur for the five flips?

(a) HTTHT (b) HHHHH (c) they have the same chance of happening.

The responses indicated a great variety of conceptions, and interpretations of the problem. The notion of a representative sample that is so helpful in the Tversky and Kahneman (1974) survey can cause problems when applied in the above context. There is no sample in the above question, there is just the sample space and yet some of the students appeared to superimpose a sampling context on the original question in order to employ the representativeness strategy in their responses.

Watson and Kelly (2003) considered students predictions and explanations for outcomes when a normal six-sided die is tossed 60 times. Since the task was part of a larger study, they were able to consider differences across grades 3 to 9 students’ change in performance after some classroom chance and data experiences that were devised to enhance appreciation of variation. The researchers used a five code hierarchy to analyse the responses: pre-structural, uni-structural, transitional, multi-structural and relational. The students using the relational level responses used appropriate variation and explanations reflecting the random nature of the process. Only 7% of students across grades 5 and 7 responded appropriately. A decrease was evident in grade 9. The researchers suggest that teachers themselves may be a useful focus of research in terms of their own understanding of expectation and variation.

OVERVIEW OF THE STUDY

The research setting was a graduate mathematics education course situated in the second semester for prospective primary teachers at a university. A group of 24 pre-service teacher education students completed a questionnaire during one of the tutorials. All these students were in their final year of education.

The birth problem (Item 1) attempted to explore students' understanding of variation in everyday setting. The students had to select the appropriate option and provide appropriate reasoning. The die question (Item 2) was used to elicit students' ideas about variation embedded in a chance generating device. Responses demanded both numerical and qualitative descriptions.

Item 1

Half of all newborns are girls and half are boys. Hospital A records an average of fifty births a day. Hospital B records an average of ten births a day. On a particular day, which hospital is more likely to record 80 percent or more female births?

(a) Hospital A (with fifty births a day)

(b) Hospital B (with ten births a day)

(c) The two hospitals are equally likely to record such an event.

Please explain your answer.

Item 2

(a) Imagine you threw a die 60 times. Fill in the table below to show how many times each number might come up.

Number on Die	How many times it might come up?
1	
2	
3	
4	
5	
6	
TOTAL	60

(b) Why do you think these numbers are reasonable?

RESULTS

Students' responses to Item 1 were categorised both on the basis of their appreciation (option b) and non-consideration (option c) for variation. Students' numerical responses on Item 2 were coded on two scales (Shaughnessy et al., 1999), a centering scale (10, 10. 10. 10. 10. 10) and a scale for variation (low, appropriate, high). The criteria for determining the appropriateness of variation displayed in the numerical answers was the same as that of Watson and Kelly (2003). Appropriate variation was demonstrated if the standard deviation in the responses fell between 1.2 and 4.7. I created a simple three category rubric that could be helpful for describing research results. The three categories in the model are: non-statistical, partial-statistical and statistical.

Statistical Responses. To be considered statistical on Item 2, students had to display appropriate variation and also provide explanations reflecting the random nature of the process. While seven students managed to respond in a statistical manner on Item 1, only two did so on Item 2. The following responses come from this category.

b. Because 10 births a day is not a sufficient number to produce a reliable result. Because the sample is smaller it has more variability. (S16)

b. Short frequencies are more likely to deviate from the true probability. (S22)

Because each number should come up roughly 10 times, give or take a few. The more times the dice is thrown the better. (S16)

Partial-statistical responses: There were two types of partial-statistical responses. One type realised conflict of probability and variation (Level 2). The other type produced responses based on the equiprobability bias (Level 1). Students who based their explanations on the equiprobability bias tended to assume random events to be equiprobable by nature Of the 15 students with partial-statistical responses on Item 1, seven used level 2 type of responses whereas the rest based their reasoning on intuitions such as equiprobability (Level 1). The following are indicative of partial-statistical responses on Item 1.

b. Because the sample is smaller so 8/10 is more likely than 40/50 girls. (L2, S23)

c. There is always a chance that both hospitals might record 80% female births

because probability is to do with equally likely outcomes. (L1, S17)

Of the 18 students considered partial-statistical on Item 2, 15 responded with no or high variation in their predictions and based their reasoning on equal probability or were part-way to providing an appropriate explanation but needed more detail and precision. These responses are equivalent to Level 1 type of responses.

10, 10, 10, 10, 10, 10. Because each number has one in six chances of being thrown. (S10)

There are 6 numbers and they all have an equal chance of coming up i.e. 60/6=10 each. (S20)

Because assuming the die is weighted evenly you are equally likely to throw either number. The sample is big enough to make it reasonable to assume an even chance. (S21)

Three students provided Level 2 type of responses. Although the students responded with reasonable variation, they did not provide adequate explanations. The following explanations are indicative of this category.

9, 10, 10, 11, 12, 8. Because it is unlikely each number will come up an equal number of times, even though the probability is 6/60 for each number. (S23)

8, 10, 12, 16, 7, 7. Because any set of numbers is possible as long as they sum to 60. (S19)

Non-statistical responses: Two students judged that the probability of obtaining more than 80% females was the same for both hospitals because the chance was the important factor not the number of births. Thus the base rate data of 80% variability was completely ignored because it did not have any implications. The four students with responses in this category for Item 2 used the centre criteria for prediction. Three of these students did not give any explanations or used terms such as random for their predictions whereas one applied rules inappropriately.

DISCUSSION

The survey results indicate that variability concepts of pre-service teachers is not significantly more sophisticated than that of students. The findings are consistent with the findings of Watson and Kelly (2003). For instance, in Watson and Kelly study, 7% of students across grades 5 and 7 responded appropriately to Item 2. In the present survey, eight percent of the teachers responded appropriately. One explanation for this could be classroom emphasis on classicist probabilities rather than frequentist approach. Students appreciate equally likely outcomes but fail to conceptualise the variation that can emerge across a number of repetitions of the event. In short, they are unable to integrate expectation and variation (uncertainty) into the sampling construct. This indicates that textbook-type of exercises to do with theoretical probability is insufficient to help students develop a complete understanding of chance events. I agree with Watson and Kelly in recommending that more explicit and repeated recognition of both variation and expectation is needed if a genuine appreciation of variation is to be achieved.

According to Tversky and Kahneman, (1974) and Shaughnessy (1997) the representativeness strategy underlies the sample variability misconception. The results of this survey provide evidence that students did not rely on the representativeness strategy but based their thinking on the equiprobability bias. One possible explanation for this could be that the contexts for the tasks were quite different and the wording of the questions was different. For instance, the word "fair" in Shaughnessy's study (1997) indicates a purposeful construction of the situation - a word that is missing from Item 2 and students may have responded differently to these situations.

The results show that students did not explicitly use words dealing with variation (spread, deviate). These findings are similar to those reported by Shaughnessy et al. (1999). Moreover, many teachers gave answers that were partially correct but did not contain enough detail and did not say precisely what they meant. Additionally, while more students showed competence on Item 1, they were less competent on Item 2. This could be due to task format or contextual issues.

This preliminary survey was just a first phase towards exploring pre-service teachers' conceptions of variability. It suffers from all limitations that accompany a written questionnaire. Moreover, some of the issues addressed in this paper may actually be due to misinterpretation of the questions. Although the study provides some valuable insights into the kind of thinking that students use, the conclusions cannot claim generality because of a small sample. Some directions for future research are implied by the limitations of this study.

IMPLICATIONS FOR TEACHING AND RESEARCH

One direction for further research could be to replicate the present study and include a larger sample of students from different educational backgrounds to claim generality. Probably there is a need to conduct individual interviews with teachers in order to probe their conceptions of variability at a greater depth. A sample of these students could also be interviewed while they gather actual data on the die question to see if the variation in results of trials influence their predictions.

Another implication relates to meaningful contexts. The picture of students' thinking in regards to sampling variation is somehow limited because students responded to only two items. There is a need to include more items using different chance contexts such as drawing objects from containers and using various statistical representations in order to explore students' conceptions of variation and related contexts in much more depth. It is also important for future research to employ a variety of task formats. Perhaps extending the question to include range and choice versions (Shaughnessy et al., 1999) graphical representation would be useful.

It appears that variability concepts of pre-service teachers is not significantly more sophisticated than that of students they going to teach. This issue needs to be addressed in teacher education mathematics courses to ensure that the content knowledge that teachers take to the classroom is appropriate for effective teaching. A variety of suitable activities for overcoming these alternative conceptions need to be found or designed.

Finally, like the primary graduates, primary undergraduate mathematics education students and in-service teachers are likely to resort to partial-statistical or deterministic explanations. Research efforts at this level are crucial in order to inform teacher educators and curriculum writers.

References

Ben-zvi D. & Garfield, J. (2004). Research on reasoning about variability: A forward. *Statistics Education Research Journal, 3*(2), 4-6.

Heaton, R. M. & Mickelson, W. T. (2002). The learning and teaching of statistical investigation in teaching and teacher education. *Journal of Mathematics Teacher Education, 5*, 35-59.

Ministry of Education. (2004). *Numeracy professional development projects supplement*.(Book 9) Wellington: Ministry of Education.

Moore, D. (1997). New pedagogy and new content: the case of statistics. *International Statistical Review, 65*(2), 123-165.

Shaughnessy, J. M. (1997). Missed opportunities in research on the teaching and learning of data and chance. In F. Biddulph and K. Carr (Eds.), *People in mathematics education. Mathematics Education Beyond 2000*. *Proceedings of the 23[d] Annual Conference of the Mathematics Education Research Group of Australasia* (pp. 6-22). Hamilton: MERGA INC.

Shaughnessy, J. M., Watson, J. Moritiz, J. & Reading, C. (1999). School mathematics students' acknowledgement of statistical variation. Paper presented at the 77th Annual National Council of Teachers of Mathematics Conference, San Francisco, CA,

Tversky, A. & Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. Science, 185, 1124-1131.

Watson, J. M. & Kelly, B. A. (2003). Predicting dice outcomes: The dilemma of expectation versus variation. Mathematics Education Research: Innovation, Networking, Opportunity *Proceedings of the 26[rd]Annual Conference of the Mathematics Education Research Group of Australasia* (pp. 728-735). Deakin University, Melbourne: MERGA INC.