Exploring Pre-service Teachers' Reasoning about Variability: Implications for Research

Sashi Sharma
University of Waikato
sashi@waikato.ac.nz

Concerns about the importance of variation in statistics education and a lack of research in this topic led to a preliminary study which explored pre-service teachers' ideas in this area. The teachers completed a written questionnaire. Responses were categorised in relation to a framework that identifies levels of statistical thinking. Although the pre-service teachers have had more real-life experiences involving probability and have been involved in the study of probability concepts at secondary school level, they still demonstrated the same misconceptions as younger students. Additionally, while more students showed competence with the birth problem, they were less competent on the die toss question. This could be due to task format or contextual issues. The paper concludes by suggesting some implications for further research.

Although it has been stated that statistical variation plays a fundamental role in students' understanding and application of statistics and chance (Ben-zvi & Garfield, 2004; Metz, 1997; Ministry of Education, 2004: Moore, 1997), little research attention has been given to variability and related concepts (Shaughnessy, 1997; Shaughnessy, Watson, Moritz, and Reading, 1999). 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. 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 the students. There appears a need to collect data from teachers at both the pre-service and in-service levels regarding their conceptions about statistics and probability. 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. Prior to discussing the details of my own research, I will briefly discuss the theoretical framework and some existing literature on statistical variation.

Theoretical Framework

Much recent research suggests that socio-cultural theories combined with elements of constructivist theory provide a useful model of how students learn mathematics. Constructivist theory in its various forms is based on a generally agreed principle that learners actively construct ways of knowing as they strive to reconcile present experiences with already existing knowledge (von Glasersfeld, 1993). Students are no longer viewed as passive absorbers of mathematical knowledge conveyed by adults; rather they are considered to construct their own meanings actively by reformulating the new information or restructuring their prior knowledge (Cobb, 1994). However, this active construction process may result in alternative views as well as the student learning the concepts intended by the teacher. Another notion of constructivism derives its origins from the work of socio-cultural theorists such as Vygotsky (1978) and Lave (1988) who suggest that learning should be thought of more as the product of a social process and less as an individual activity. There is strong emphasis on social interactions, language, experience, collaborative learning environments, catering for cultural diversity and contexts for learning in the learning process rather than cognitive ability only. Mevarech and Kramarsky (1997) claim that the extensive exposure of our students to statistics in out-of-school contexts may create a unique situation where students enter the mathematics class with considerable amount of statistical knowledge. This means that during the teaching and learning process, students draw inferences about the new information presented to them by relating to some aspect of this prior knowledge to develop a deeper meaning for statistical concepts. This research was therefore designed to identify students' alternative ideas about variability, and to examine how they construct them.

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 states 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%.

Conversely, 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.

Shaughnessy et al. (1999) surveyed 324 students in grades 4-6, 9 and 12 in Australia and the United States using a variation of an item on the National Assessment of Educational Progress (Zawojewski & Shaughnessy, 2000). Three different versions of the task were presented in a Before and in a Before and After setting. In the latter setting students did the task both before and after carrying out a simulation of the task. Responses were categorised according to their centre and spreads. While there was a steady improvement across grades on the centre criteria, there was no clear corresponding improvement on the spread criteria. There was considerable improvement on the task among the students who repeated it after the simulation. The researchers conjectured that the lack of clear growth on spreads and variability and the inability of many students to integrate the two concepts (centres and variation) on the task may be due to instructional neglect of variability concepts.

As part of a larger study, I (Sharma, 1997) used the following item to explore high school students' understanding of sampling variation.

> Shelly is going to flip a coin 50 times and record the percentage of heads she gets. Her friend Anita is going to flip a coin 10 times and record the percentage of heads she gets.
> Which person is more likely to get 80% or more heads? Explain your answer.

The students were interviewed by myself and interviews were tape recorded and transcribed for analysis. From a statistical point of view, more than 80% heads is more likely to occur in the small sample because the large sample is less likely to stray from 50%. However, none of the students in my study were considered statistical on this item, students based their reasoning on their cultural beliefs, everyday experiences and intuitive strategies.

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. In the current study, two open-ended items were used to determine specific student conceptions and the factors that contribute to these constructs. An overview of the research design follows, after which I will discuss the results of my study.

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 an 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. In both these questions, the students had to consider variability assumption related to the events, hence this is the central notion to which I refer in both items.

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?

Analysis

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 framework that could be helpful for describing research results and planning instruction. The three categories in the model are: non-statistical, partial-statistical and statistical. The term statistical is used in this paper for the correct responses. However, I am aware that such a term is not an ideal one. Student teachers possess interpretations and representations which may be situation specific and hence these ideas have to be considered in their own right. Statistical simply means what is usually accepted in standard mathematics text-books and research literature.

Results

In this section the types of responses are summarised and the ways in which the students have explained their thinking is described. Typical responses are used for illustrative purposes. Throughout the discussion, Sn is used for the *nth* student.

Statistical Responses

From a statistical point of view (Item 1), more than 80% female births is more likely to occur in Hospital B because the large sample is less likely to vary from 50%. 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 (Lecoutre, 1992), 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 only 8 of 10 have to be girls. In (a) 40 of the 50 have to be girls. (L2, S11)
>
> 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 chance of being thrown. (S10)
>
> There are 6 numbers and they all have an equal chance of coming up ie 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. Student 17 said: 6/60 or 1/10 of a chance of each number being rolled.

Discussion

The thinking of most of the students in this survey was heavily influenced by equally likely and expectation conceptions rather than a consideration of variability. Although some students do appear to possess notions of variability, they were often unable to integrate expectation and variability into their explanations. After discussing the findings in a broader context, this section suggests some directions for further research.

Sampling Variability: A Broader Context

The survey results indicate that variability concepts of pre-service teachers are 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 exercises to do with theoretical probability are 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 were 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. Given the subtleties of interpretations, it is unlikely that the items used in the survey described in this paper would have discriminated finely enough. 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.

Secondly, this small scale investigation into identifying and describing students' reasoning has opened up possibilities to do further research at a macro-level on students' thinking and to develop more explicit categories for each level of the framework. Such research would validate the framework of response levels described in the current study and raise more awareness of the levels of thinking that need to be considered when planning instruction and developing students' statistical thinking.

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) and Green's (1983) graphical representation would be useful.

It appears that variability concepts of pre-service teachers are 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.

Cobb, P. (1994). Where is the mind? Constructivist and sociocultural perspectives on mathematical development. *Educational Researcher*, 23(7), 13-20.

Green, D. (1983). Shaking a six. *Mathematics in Schools*, *12*(5), 29-32.

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.

Lave, J. (1991). *Cognition in practice: Mind, mathematics and culture in everyday life*. New York: Cambridge University Press.

Mevarech, Z. R., & Kramarsky, B. (1997). From verbal descriptions to graphic representations: Stability and change in students' alternative conceptions. *Educational Studies in Mathematics, 32*, 229-263.

Metz, K. E. (1997). Dimensions in the assessment of students' understanding and application of chance. In I. Gal & J. B. Garfield (Eds.), *The assessment challenge in statistics education* (pp. 223-238). Amsterdam The Netherlands: IOS Press.

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.

Sharma, S. (1997). *Statistical ideas of high school students: Some findings from Fiji.* Unpublished doctoral thesis. Waikato University, Hamilton, New Zealand

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

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.

Von Glasersfeld, E. (1993). Questions and answers about radical constructivism. In K. Tobin (Ed.), *The practice of constructivism in science education* (pp. 24-38). Washington: American Association for the Advancement of Science.

Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes*. Cambridge: Harvard University Press

Watson, J. M. & Kelly, B. A. (2003). Predicting dice outcomes: The dilemma of expectation versus variation. In L.Bragg, C.Campbell, G.Herbert & J.Mousley (Eds.), *Mathematics education research: Innovation, networking, opportunity* (Proceedings of the 26rdAnnual Conference of the Mathematics Education Research Group of Australasia, pp. 728-735). Sydney: MERGA.

.