High School Students' Understanding of Sampling Variability: Implications for Research

Sashi Sharma
School of Education
University of Waikato

ABSTRACT: Concerns about students' difficulties in statistical reasoning led to a study which explored year 11 (14 to 16 year-olds) students' ideas in this area. The study focussed on sampling variability, probability, descriptive statistics and graphical representations.

This paper presents and discusses the ways in which students made sense of sampling variability constructs obtained from the individual interviews. The findings revealed that many of the students used strategies based on beliefs, prior experiences and intuitive strategies.

While students showed competence with probability, descriptive statistics and graphical representations tasks, they were less competent on tasks involving sampling variability. This could be due to instructional neglect of this concept or linguistic problems.

The paper concludes by suggesting some implications for further research.

Background

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 need of statistical thinkers to acknowledge the omnipresence of variation and to consider appropriate ways to quantify, explain and model the variability in data. Moore also highlights the key role that variation plays in students' understanding of chance. He writes that it is important for students to understand the idea that "chance variation, rather than deterministic causation, explains many aspects of the world" (p. 99). 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. It must be noted that although some authors, such as Ben-Zvi and Garfield (2004), distinguish between the terms variation and variability, in this paper these terms have been used interchangeably to represent how data vary.

Although it has been argued that variability plays a fundamental role in students' understanding and application of statistics and chance (Metz, 1997; Ministry of Education, 2004: Moore, 1997; Watson & Moritz, 2000), little research attention has been given to these concepts (Ben-Zvi & Garfield, 2004; Shaughnessy, 1997; Shaughnessy, Watson, Moritz, and Reading, 1999). For instance, Ben-zvi and Garfield (2004) state:

> Despite the attention paid by statisticians and statistics instructors to this important topic, to date little has been published about how people, particularly novices and statistics students, actually reason about variability, or how this type of reasoning develops in the course of statistics instruction (p. 4)

Shaughnessy (1997) suggests some possible reasons for the lack of attention to research on variability. One is that although averages play a very important role in the school mathematics curriculum, variation does not. Since research often reflects the emphasis in curricula materials, there have been investigations into students' ideas about averages and a variety of models for understanding measures of central tendency have been suggested. Shaughnessy points out that there does not appear to be a similar trend into exploring students' ideas about variability. The neglect of variation is noted in the *National Assessment of Educational Progress* (Zawojewski & Heckman, 1997) studies which tested student achievement in Grades 4, 8, and 12. The statistics assessment test items addressed concepts of descriptive statistics, mainly measures of central tendency and range as a measure of variability. However, it was a low level computational task. Given that variation plays a central role in the study of statistics and a lack of research in this area, more research needs to be undertaken to better understand how students explain variation.

Previous Research

A number of research studies from different theoretical perspectives seem to show that students tend to have intuitions which impede their learning of probability concepts. Some prevalent ways of thinking which inhibit the learning of probability are: representativeness, equiprobability bias, beliefs, and human control.

- Representativeness: According to this strategy, students make decisions about the likelihood of an event based upon how similar the event is to the population from which it is drawn or how similar the event is to the process by which the outcome is generated (Tversky & Kahneman, 1974). For instance, a long string of heads does not appear to be representative of the random process of flipping a coin, and so those who are employing representativeness would expect tails to be more likely on subsequent tosses until things evened out.

- Equiprobability bias: Students who use this bias tend to assume that random events are equiprobable by nature. Hence, the chances of getting different outcomes, for instance, three fives or one five on three rolls of a die are viewed as equally likely events (Lecoutre, 1992).
- Beliefs: Research shows that a number of children think that their results depend on a force, beyond their control, which determines the eventual outcome of an event. Sometimes this force is God or some other force such as wind, other times wishing or pleasing (Truran, 1994).
- Human Control: Research designed to explore children's ability to generalise the behaviour of random generators such as dice and spinners show that a number of children think that their results depend on how one throws or handles these different devices.

Representativeness: To illustrate the undue confidence that people put in the reliability of small samples, take Tversky and Kahneman's (1974) problem given 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. Subjects who relied on the representative heuristic tended 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 five 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 indicate 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 this context. There is no sample in this question, there is just the sample space, 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.

To investigate the influence of the concept of sample variability on students' thinking as opposed to sample representativeness, Rubin, Bruce and Tenney (1991) asked 12 high school students to evaluate two different ways of dividing up 400 runners: 200 fast and 200 slow into blue and red teams. One was the running ability of each runner and the other was to assign runners to each team randomly (by choosing names out of a hat and assigning alternate runners to each team). The students were asked to provide an assessment of the fairness of the hat method: how likely it was to produce teams that were balanced in terms of fast and slow runners. Many students reported that unequal teams were possible with the hat method: teams with 150 fast runners and 50 slow ones were possible outcomes. Rubin *et al.* point out that one possibility for this problem could be the availability heuristic as indicated by the following response:

> Well, usually, we do that in gym so you know, you go there, and they come out pretty uneven. One team is much better than the other team. It doesn't really work out very well (ref).

The student does not recognise the effect of sample size: she generalises inappropriately from her gym class experience with small samples to the runners problem with a sample of size 200.

Equiprobability: Lecoutre (1992) used the following question in an experimental study of 1000 students with various backgrounds in probability:

> Two dice are simultaneously thrown, and the following two results are obtained:
>
> R1: 5 and 6 are obtained
>
> R2: a 6 is obtained twice
>
> Do you think the chance of obtaining each of these results is equal?
>
> Or is there more chance of obtaining one of them, and if so, which, R1 or R2?
>
> Or is it impossible for you to give an answer, and if so, why?

Lecoutre reports that most of the subjects answered incorrectly, that the two events have the same probability. From a systematic analysis of the justifications provided by students, it appeared that the most frequent cognitive model (65%) was based on the following type of argument: The two results are equiprobable because it is a matter of chance. Lecoutre also found that the equiprobability bias was highly resistant to change. For instance, it was found that increasing age, greater experience and experimental context (combinatorial information, frequency information) had little or no impact on this bias.

Beliefs: Amir and Williams (1994) propose that beliefs appear to be the elements of culture with the most influence on probabilistic thinking. They interviewed thirty-eight 11 to 12 year old children about their concepts of chance and luck, their beliefs and attributions, their relevant experiences and their probabilistic thinking. Some pupils thought God controls everything that happens in the world, while others thought God chooses to control, or does not control anything in the world.

Several pupils believed in superstitions, such as walking under a ladder, breaking a mirror, and lucky and unlucky numbers. There were also beliefs directly related to coins and dice; for instance, when throwing a coin, tails is luckier. A majority of children in the Amir and Williams study concluded that it is harder to get a 6 than other numbers (17 out of 21 interviewees). The children remembered from their experience of beginning board games waiting a long time for a 6 on the die.

Human control: Fischbein and Schnarch (1997) asked 139 junior high school students (prior to instruction) to compare the probability of obtaining three fives by rolling one die three times versus rolling three dice simultaneously. Two main types of unequal probabilities were mentioned by about forty percent of the students. Of these students, about three-fifths considered that, by successively throwing the die, they had a higher chance of obtaining the expected result, and about two-fifths considered that by throwing three dice simultaneously they would have a higher chance of obtaining the expected results. Some of the justifications provided by the students for successive trials were:

> By rolling one die at a time, one may use the same type of rolling because the coins do not knock against each other and therefore do not follow diverse paths.

The opposite solution, that is, there was a greater chance of getting three heads by tossing the coins simultaneously also had its proponents. Reasons given included:

> Because the same force is imparted.
>
> One can launch in the same way.

The explanations indicate a belief that the outcomes can be controlled by the individual. In the Truran study (1994), many different methods of tossing coins and dice were described by children in order to get the result they wanted. For example, for tossing three dice together, some children thought that it is better to throw the dice one at a time because (when tossed together) dice can bump into each other and change the numbers which would otherwise have come up. Even if their carefully explained and demonstrated method did not work, children were still convinced that if they did everything right, it would work the next time. With respect to the task:

> Are you more likely to get 5 on each of these three dice, by rolling one dice three times or rolling all three dice together? Can you say why?

The responses suggest that the children's perceptions of the behaviour of random generators with similar structure were based on the physical attributes of the random generators.

Shaughnessy and Zawojewski (1999) documented similar statements among twelfth graders on probability tasks. Rather than using the sample space or multiplication principle, the students attributed the results to some sort of physical property of the spinners such as the rate at which the spinners were spun, the initial position of the arrows, or even some external influence, for instance, the wind.

Mathematics educators should realise that the students they teach are not new slates waiting to have the formal theories of probability written upon them. The students already have their own built-in heuristics, biases and beliefs about probability and these cannot, as it were, be simply wiped away. If student conceptions are to be addressed in the process of instruction, then it is important for teachers of probability to become familiar with the alternative conceptions that students bring to classes. Such information may help teachers plan learning activities and students overcome their difficulties. In the current interview-based study, open-ended tasks 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 secondary school selected for the research was a typical high school in Fiji. The class consisted of 29 students aged 14 to 16 years, of which 19 were girls and 10 were boys. According to the teacher, none of the students in the sample had received any in-depth instruction on statistics and probability prior to the interviews. The whole class participated in the first phase of interviews, and due to time constraints 14 students participated in the second phase during which the class teacher taught a unit on statistics and probability. This group of 14 was representative of the larger group in terms of abilities and gender. The data reported in this paper comes from the second phase interviews.

To explore the full range of students' thinking, open-ended questions to do with probability and statistics constructs were selected and adapted from those used by other researchers. The interview question related to sampling variability was:

> The coin problem
> 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.

Each student was interviewed individually by myself in a room away from the rest of the class. The interviews were tape recorded for analysis. Each interview lasted between 40 to 50 minutes. Paper, a pencil and a calculator were provided for the student if he or she needed it.

Results

Analysis of responses indicated that the students used a variety of intuitive strategies and previous experiences for solving the problem. I created a simple four stage based model that could be helpful for describing research results relating to students' statistical conceptions, planning instruction and dissemination of findings to mathematics educators. The four categories in the model are: non response, non-statistical, partial-statistical and statistical.

The main focus in this paper is on the non-statistical responses (in which students made inappropriate connections with everyday experiences or learning in other areas) and the partial-statistical responses (in which students used intuitive strategies, applied rules and procedures inappropriately, referred to some statistical points without generalising to all information). In this section, the types of responses are summarised and the ways in which the students have explained their thinking is described. Extracts from typical individual interviews are used for illustrative purposes. Throughout the discussion, I is used for the interviewer and Sn for the nth student. Results of student responses to Item 1 are summarised in Table 1.

Table 1:

Response types for task involving sample variability (n = 14)

Response type	Number of students using response
Non-response	1
Non-statistical	6
Partial-statistical	7
Statistical	-

As Table 1 shows, no student managed to respond to this problem in a statistical manner. The responses of the other 13 students were roughly evenly divided between non and partial-statistical.

Non-statistical responses: 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, the results of this study indicate that six students based their reasoning on their cultural beliefs and experiences. Two students judged that the probability of obtaining more than 80% heads was more likely to occur with 50 flips of a coin than with 10 flips. The students did not attend to the effect of sample size on variability when making estimates of the likelihood of outcomes. Thus, the base rate data of 80% variability was neglected because it did not have any causal implications. The explanations given by Student 17 are indicative of the causality perspective.

> *Yeah, Oh ... would be Shelley. Because Shelley's amount comes to 50; Anita does it only 10 times. Oh ... Shelley because she does more flips. She got more chance to get 80%.*

The other four students with responses in this category thought that the flipping of coins depended on luck or how one tosses the coin. For example, Student 20 explained,

> *Eh ...as I said before... because when you throw each time the coin comes head or tail or tail or head. She will throw the coin in one direction so she will get HHH,* when *she will change in another direction she will get all tails... So it will depend on how fast you throw and how fast the coin swings.*

Student 6 offered the following response:

> *This flipping of coins depends on luck; if a person ... is a lucky person, then he will be able to have heads.*

Partial-statistical responses: Of the seven students with partial-statistical responses on this task, one applied rules inappropriately, whereas five based their reasoning on intuitions such as equiprobability and unpredictability.

The particular rule applied inappropriately by these students was the percentage rule. For example, Student 5 responded that Shelly is more likely to get 80% *because she gets 40.* She simply calculated 80% of 50, getting an answer of 40. Maybe the student did not understand the question and readily performed arithmetic operations on the numbers given in the problem.

Students who base their explanations on the unpredictability bias tended to assume random events to be unpredictable by nature. One of the keys to understanding a probability model is balancing the ideas of theoretical and empirical probability. It is true that individual events are unpredictable, however, long run frequencies are predictable. Over-reliance on theoretical probability is likely to lead to the belief that all outcomes are unpredictable. From the students' explanations, it is clear that their understanding of variation in small samples was minimal in this context. Student 25 believed that both Shelly and Anita were likely to get 80% or more heads because *I don't know what will come. It can be tail or head.* Student 29 thought likewise:

> *It can be anyone because she tossed the coin 50 times, she can get more heads, and even this one too [meaning Anita] she can get less tails too eh; less heads. Both have 50% chances of getting heads.*

Three students thought that neither of them could get 80% heads. Part of the explanations provided by the students seem to indicate a view that chance is naturally equiprobable. Even repeated probing by myself did not induce any statistical thinking. Since there are two outcomes in the sample space, these students believe that each should occur about a half of the time, regardless of the sample size. Two students even altered their data in this problem to align it with their personal preferences. When asked which person is more likely to get 80% or more heads, student 2 responded that both of them would get the same.

S2: Because a coin has two sides only heads or tails. One tosses it 10 times or 100 times, it makes no difference.

I: So you think that both will get 80% or more heads?

S2: 80% or more heads, no [laughs]. Not 80%. They will be getting 50% or more heads.

I: Can you say why?

S2: Because the chances of heads and tails is half. So it will be 50%.

One student gave the correct answer with partially correct reasoning, *Anita, because she does it fewer times.*

Discussion

It must be acknowledged that the limited use of statistical reasoning on this question may be a consequence of a lack of emphasis on variation in the classroom and curricular materials. This could also be a consequence of a classroom mathematics culture that asks questions with a single answer. It takes considerable self-confidence to say something like *I think the smaller sample is more likely to deviate from the centre* or even to say *I don't understand the question.* Gal (1998) states that suggesting to students that a judgement is called for, rather than a precise mathematical response, will make students think more about data and not look straight away for some numbers to crunch. It appears that this strategy might not work for students who lack experience in explaining their answers or have strong beliefs about random generators.

According to Tversky and Kahneman, (1974) and Shaughnessy (1997) the representativeness strategy underlies the sample variability misconception. The results of this study provide evidence that students did not rely on the representativeness strategy, but based their thinking on the unpredictability and equiprobability bias. One explanation for this could be classroom emphasis on classicist probabilities rather than frequentist approach. Students can reason about the unpredictability of a single event but fail to conceptualise the patterns that can emerge across a large number of repetitions of the event (Metz, 1997). In short, they are unable to integrate uncertainty and pattern aspects into the sampling construct. Another possible explanation for this could be that the contexts for the tasks were quite different and the students were different ages with different statistical backgrounds, hence the strategies employed 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 the Item used in this study and students may have responded differently to these situations.

The finding that students base their thinking in statistics on their prior experiences certainly concurs with the views expressed by Rubin *et al.* (1991) and Amir and Williams (1994). The subjects in the Rubin *et al.* study expected the sample to equal the population and if it differed they believed that the experimenter had made a mistake. Moreover, results show that students did not explicitly use words dealing with variation (spread, deviate). They entertained several ideas about the concepts of sampling, most of which were based on what they had absorbed informally from their everyday environments. These findings are similar to those reported by Watson and Moritz (2000) and Shaughnessy *et al.* (1999).

The results show that quite a number of students think that outcomes on random generators such as coins can be controlled by individuals. The general belief is that results depend on how one throws or handles these different devices. The finding concurs with the results of studies by Shaughnessy and Zawojewski (1999), Truran (1994) and Fischbein and Schnarch (1997). For instance, Shaughnessy and Zawojewski (1999) reported that rather than using the sample space or multiplication principle, the students in the NAEP attributed the results to some sort of physical property of the spinners such as the rate at which they were spun. It must be noted that the students using the control strategy in this study were boys. One explanation for this could be that boys are more likely to play sports and chance games that involve flipping coins and rolling dice to start these games.

Although this study provides evidence that reliance upon control assumption can result in biased, non-statistical responses, in some cases this strategy may provide useful information for other purposes. For example, Student 20's knowledge of physics may have been reasonable. The students using this approach have drawn on relevant common sense information. The responses raise further questions. Is there a weakness in the wording of this task in that it is completely open-ended and does not focus the students to draw on other relevant knowledge? Perhaps, including cues such as "fair" in the item would have aided in the interpretation of this question. Are the students aware of the differences in probabilistic reasoning and reasoning in other curriculum areas? Although we consider the flip of a coin and the throw of a die as random, deterministic physical laws govern what happens during these trials. We can imagine throwing a coin in a way that we can predict the outcome (pushing them smoothly from a height of 1 cm). It depends on the situation and the context. Even with a fair coin, the side that it lands on is virtually completely determined by a number of factors such as which way up it started and the degree of spin. If we knew all this, then with sufficient expertise in physics we could write down some equations which are thought to govern the motion of the coin and use these to work out which way up the coin should land.

Students applied algorithmic procedures inappropriately to the question. It appears that the limited nature of school instruction in statistics did not provide the students with opportunities to refine their concepts and resolve ambiguities. The students learn statistics as a set of rules without learning the meaningful contexts in which they should be applied. Teachers assume that students who learn to process data can transfer these skills to interpreting and developing a critical attitude to information. If statistics is embedded in contextual settings and the shared practical activities of people, students need relevant experiences in which they construct and use statistics. The results support claims made by Lave (1991) that learning for students is situation specific and that that connecting students' everyday contexts to academic mathematics is not easy.

In the study described here, background knowledge, that is often invoked to support a student's mathematical understanding, is getting in the way of efficient problem-solving. Given how statistics is often taught through examples drawn from "real life", teachers need to exercise care in ensuring that this intended support apparatus is not counterproductive. This is particularly important in light of current curricula calls for pervasive use of contexts (Meyer, Dekker & Querelle, 2001; Ministry of Education, 1992) and research showing the effects of contexts on student's ability to solve open-ended tasks (Cooper & Dunne, 1997; Sullivan, Zevenbergen & Mousley, 2002). For instance, the study by Cooper and Dunne found that some pupils have a greater facility in recognising whether they are being asked to play a 'school maths' game or an 'everyday life' game.

It must be acknowledged that the open-ended nature of the tasks and the lack of guidance given to students regarding what was required of them certainly influenced how students explained their understanding. The students may not have been particularly interested in these types of questions as they are not used to having to describe their reasoning in the classroom. Some of the misconceptions addressed in this study may actually be due to misinterpretation of the questions. Given the subtleties of interpretations that have been reported in literature, it is unlikely that the item used in the research described in this paper would have discriminated finely enough. Some students clearly had difficulty explaining explicitly about their thinking. Students who realised that Anita was more likely to get 80% or more heads had a difficult time explaining their responses. The issues of language use are particularly more important for these students, because the students face schooling in a second language that is not spoken at home. Another reason could be that such questions do not appear in external examinations. Instruction in Fiji centers mostly around presentation of short answers of the correct/incorrect type to procedural questions. Although the study provides some valuable insights into the kind of thinking that high school students use, the conclusions cannot claim generality because of a small sample. Additionally, the study was qualitative in emphasis and the results rely heavily on my skills to collect information from students. Some directions for future research are implied by the limitations of this study.

Implications for Further Research

One direction for further research could be to replicate the present study and include a larger sample of students from different backgrounds so that conclusions can be generalised.

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 one item. There is a need to include more items using different contexts in order to explore students' conceptions of variation and related concepts in much more depth. Perhaps extending the question to include range and choice versions (Shaughnessy *et al.*, 1999) would be useful. The interview results show that personalisation of the context can bring in multiple interpretations of tasks and possibly different kinds of abstractions. At this point, it is not clear how a learner's understanding of the context contributes to their interpretation of data. Research on what makes this translation difficult for students is needed.

Fourth, researchers can accurately assess their students' understanding through individual interviews. The interview results provide evidence that students often experience difficulty when speaking about chance events. However, in the present investigation I overcame these difficulties by restating a task or changing the wording. For instance, some students were not familiar with the word flip, so toss was used to clarify the situation. This would have not been possible in a written survey. Moreover, researchers must probe for students' reasoning behind the answers. Some students give correct answers for incorrect reasons. For example, when comparing sample sizes one student said that the second sample was more likely to produce eighty percent or more heads. It would be easy to conclude from the response that the student had a well developed concept of sampling variation.

However, the justifications provided by the student indicate that the student did not have a coherent statistical explanation for the response.

Another implication relates to culture. Unlike the Shaughnessy (1997) study, none of the students in my study used the representativeness construct on the coin toss Item. One explanation for this could be the cultural context. Additionally, Watson and Callingham (2003) note that students in 'other cultural settings' may respond differently to their Australian counterparts, particularly to context-based items used in their studies. It would be interesting to determine how cultural practices influence conceptions of probabilistic reasoning.

An additional important implication for further research is the development and evaluation of teaching strategies that help students move beyond their prior conceptions towards those of mathematics.

Finally, like the secondary school students, primary school students are likely to resort to non-statistical or deterministic explanations. Research efforts at this level are crucial in order to inform teachers, teacher educators and curriculum writers.

References

Amir, G. & Williams, J. (1994). The influence of children's culture on their probabilistic thinking, In J. da Ponte & J. Matos (Eds.), *Proceedings of the Eighteenth International Conference for the Psychology of Mathematics Education, 29 July-3 August, 1994*, Vol. II (pp. 24-31). Lisbon, Portugal: The Program Committee of the 18th PME Conference

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

Cooper, B. & Dunne, M. (1999). *Assessing Children's Mathematical Ability*. Buckingham: Open University Press.

Fischbein, E. & Schnarch, D. (1997). The evolution with age of probabilistic, intuitively based misconceptions. *Journal for Research in Mathematics Education*, 28, 96-105.

Gal, I. (1998). Assessing statistical knowledge as it relates to students' interpretation of data. In S. P. Lajoie (Ed.), *Reflections on Statistics: Learning, Teaching, and Assessment in Grades K-12* (pp. 275-295). Mahwah, NJ: Lawrence Erlbaum Associates.

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

Lecoutre, M. (1992). Cognitive models and problem spaces in purely random situations. *Educational Studies in Mathematics*, 23, 557-568.

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.

Meyer, M. R., Dekker, T. & Querelle, N. (2001). Context in mathematics curricula. *Mathematics Teaching in the Middle School*, 6(9), 522-527.

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

Ministry of Education. (1992). *Mathematics in the New Zealand curriculum*. Wellington, NZ: Ministry of Education.

Moore, D. S. (1990). Uncertainty. In L. Steen (Ed.) *On the shoulders of giants: New approaches to numeracy* (pp. 95-137). Washington, DC: National Academy Press.

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

Rubin, A., Bruce, B. & Tenney, Y. (1991). Learning about sampling: Trouble at the core of statistics. In D. Vere-Jones (Ed.), *Proceedings of the third international conference on teaching statistics. Vol. 1. School and general issues* (pp. 314-319). Voorburg, The Netherlands: International Statistical Institute.

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 23d Annual Conference of the Mathematics Education Research Group of Australasia* (pp. 6-22). Hamilton, New Zealand: 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.

Shaughnessy, J. M. & Zawojewski, J. S. (1999). Secondary students' performance on data and chance in the 1996 NAEP. *The Mathematics Teacher*, 92(8), 713-718.

Sullivan, P., Zevenbergen, R., & Mousley, J. (2002). Contexts in Mathematics teaching: snakes or ladders? In B. Barton, K. C. Irwin, M. Pfannncuch and M. Thomas (Eds.), *Mathematics education in the South Pacific*. Proceedings of the 25th annual conference of the Mathematics Education Research Group of Australasia. Auckland, NZ. pp 649-656.

Truran, K. M. (1994). Children's understandings of random generators in C. Beesey & D. Rasmussen (Eds.), *Mathematics Without Limits, Proceedings of the 31st Annual Conference of the Mathematical Association of Victoria* (pp. 356-362). Melbourne, Australia: Mathematics Association of Victoria.

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

Watson, J. M. & Callingham R. (2003). Statistical literacy: A complex hierarchical construct. *Statistics Education Research Journal*, 2(2), 3-46.

Watson, J. M. & Moritz, J. B. (2000). Developing concepts of sampling. *Journal for Research in Mathematics Education*, 31(1), 44-70.

Zawojewski, J. S. & Heckman, D. S. (1997). What do students know about data analysis, statistics, and probability? In P. A. Kenney and E. A. Silver (Eds.), *Results from the sixth mathematics assessment of the national assessment of educational progress* (pp. 195-223). Reston, VA: National Council of Teachers of Mathematics.