BUILDING SAMPLING CONCEPTS FOR STATISTICAL INFERENCE: A CASE STUDY

PFANNKUCH, Maxine
The University of Auckland, New Zealand

This paper examines ways in which embryonic reasoning about sampling variability can be developed in instruction. Activities that could support such reasoning were developed and used by a teacher and her class of Year 10 (14 year-old) students. Results are reported from two web application activities that were designed to promote conceptual development in sampling reasoning and from an open-ended questionnaire completed by the students. From the teacher-student discourse the reasoning is analyzed to determine how conceptual growth was being promoted and what issues arise when the teacher is attempting to develop sampling variability concepts and the students are trying to understand them. The results suggest that the activities did start to promote conceptual growth and that teacher-student discourse and imagery were key factors. One issue that arose was the appropriate use of contextual knowledge.

OVERVIEW

Statistical inference lies at the heart of learning about real world situations from data. A myriad of concepts underpin statistical inference such as variation, sample, population, and distribution yet many school curricula (e.g., Ministry of Education, 1992) do not acknowledge or provide teaching pathways for such concepts to be developed over many years. For example, when Chance, delMas, and Garfield (2004) conducted five iterations of a specially designed course using computer software tools for developing tertiary students' reasoning about sampling distributions, they finally concluded that students' conceptual growth was being hindered by their limited understanding of related concepts such as distribution and variability. Similarly Saldanha and Thompson (2002) designed a teaching experiment to develop senior secondary students' concept of sampling distribution, a critical part of which was the relationship between sample proportion and population proportion, but again they were frustrated by students lack of a sense of variability and distribution. Such research clearly points to a lack of attention to conceptual development in earlier years. A particular concept that Pfannkuch (2007) identified in a previous study was sampling variability, the lack of which was limiting 15 year-old students' informal inferential reasoning. Also Pfannkuch (2006) noted in her analysis of a teacher's reasoning, that when the teacher compared two groups, she drew inferences about samples and also about populations from the samples. Hence research is needed on developing effective teaching approaches for introducing the conceptual building blocks for informal inferential reasoning, such as sampling reasoning, in order to provide a pathway towards students' understanding the logic of statistical inference.

The main research question addressed in this paper is: How can students be stimulated to start developing a conceptual structure with which they can conceive sampling variability? A subsidiary question is: What issues arise when the teacher is attempting to develop sampling variability concepts and the students are trying to understand them?

LITERATURE REVIEW

The development of technological tools has resulted in more research interest in finding ways to promote students' conceptual growth. Shaughnessy (2007, p. 995) states that the research he reviewed suggests "technological tools are very important for helping students to transition from those naïve conceptions to richer more powerful understanding of statistical concepts." Technology and the visual imagery offered, however, are not sufficient for conceptual growth. Important, also, is the teacher's and students' articulation of how they make sense of and explain in their own words what they see and understand and thereby create meaning from the images (Makar & Confrey, 2005). Bakker (2004) believes that such sense-making reflection between teacher and students is paramount in developing conceptual reasoning. He points out that if concepts are defined before students have intuitive ideas, then their development is hindered. In his experiment on developing students' concepts of variability, sampling, data, and distribution, where he wanted to dynamically relate samples and population without making the distinction between sample and population, he concluded that he was able to promote conceptual development. Since growth takes place over several years he felt that his students gained some intuitive notions, developed language, although imprecise, to describe and predict aggregate features of data sets, and gained an image that could be sharpened at a later stage. Shaughnessy (2007, p. 994) believes "that the type of classroom discourse that takes place during a statistics exploration also has a major impact on student conceptual growth." Therefore, it seems that for conceptual development in statistics, *image* and *language* are two key components.

The idea of image is prominent in the findings of Liu and Thompson (2007) on teachers' stochastic conceptions of probability. Since statistical inference and probabilistic notions are intimately linked, their theoretical constructs for three ways of thinking are pertinent to statistical conceptual development. Firstly, they believe that for teachers to understand and think about probability they need an image of a repeatable process and secondly, an image of a distribution of outcomes, both of which directly relate to sampling variability concepts. The third way of thinking involves understanding the conditions of the process. That is, whether the situation is conceived as probabilistic. Therefore, in relation to sampling variability concepts, for students to understand the situation presented they need to conceive that the underlying process of taking samples from a population is random.

Another idea to develop in students' sampling reasoning is why and how a representative sample from the population is taken. Such a concept develops over many years and requires instruction that is more attune to developing social reasoning skills than calculation skills (Watson, 2005). Consequently many mathematics teachers do not attend to the discussion that is inherent in communicating sampling reasoning. Watson (2005) believes that improving instruction methods will improve students' sampling reasoning. Her research on Grades 3 to 13 students characterizes a six-tier hierarchy of developing concepts of sampling. The first tier describes students who take a small sample without regard to a selection method while the sixth tier describes students who take large samples of at least 20, are sensitive to bias of selection methods, and have notions of selection methods based on a random process. Considering her sixth tier descriptor for sampling reasoning, it seems that more tiers could be added to the hierarchy as an in depth understanding of sampling variability is not evident. Although she refers to sample size, understanding the effect of sample size with respect to qualitative and quantitative data is not considered but may be an important factor when designing instruction.

Building a sense and image of variability with repeated sampling from a population demands cognizance of how students might conceive variability. Shaughnessy (2007) reports on a spectrum of student thinking about variability, from attention to outliers only, to considering the range of possible outcomes, to focusing on the likely range, a precursor to forming conceptions of measuring variability. At a later stage students start to quantify variation by considering both the centre and spread. Therefore, when students are exposed to repeated sampling from a population attention needs to be paid to their conceptions of variability within and between distributions. With respect to sample distributions, Shaughnessy (2007, p. 982) found that in his lolly task experiments a "mental tug of war" was created in students between the expected value for one sample and the range of outcomes for repeated samples. He believes that probability instruction may be an obstacle to students' thinking about variability and that instruction should strike a balance by putting the two concepts together. "They should have a sense of the *reasonable expected variability around the expected value*, something like an intuitive confidence interval" (Shaughnessy, 2006, p. 87). Similarly, integration of the concepts of representativeness and variability need to be built up in statistical data situations, since students do not intuitively link chance-sampling experiences with data sampling experiences (Schwartz & Goldman, 1996).

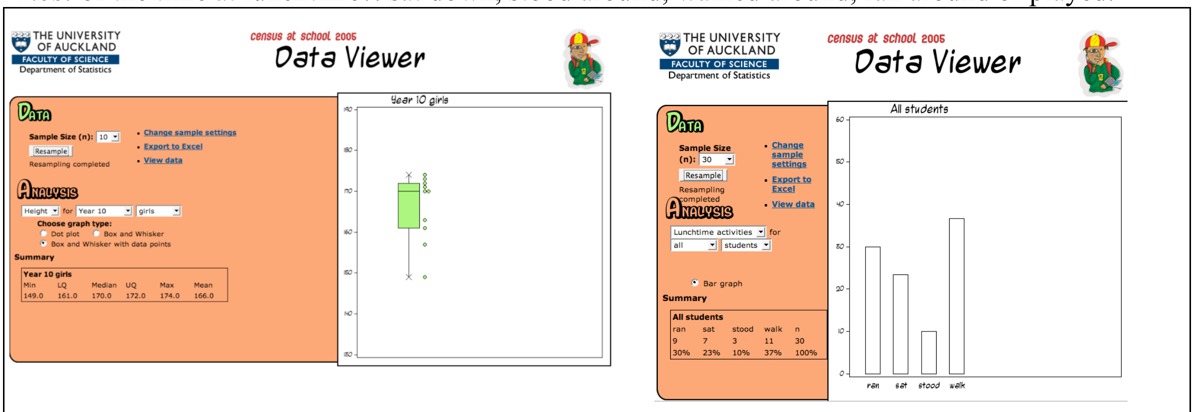
Into this milieu of issues that need to be considered about developing students' sampling reasoning is the effect of context on their reasoning. As Shaughnessy (2007) discovered, when he compared students' reaction to a lolly task and movie wait-time task context, the context of the problem can play a major role in how students perceive variability. Schwartz and Goldman (1996) also found that the context of the problem had a strong influence on students' understanding of sampling. They believe that instruction needs to expose students' prototheories and give them opportunities to reconcile discrepancies. Because the context of the problem has a powerful effect they suggest that different contexts will bring forth different prototheories and hence students can "discover organizing principles that will align the prototheories and improve progress towards normative theories" (p. S110)

Foundational knowledge for statistical inference includes understanding that an inference can be made about a population from a sample. However, "in statistics instruction it is uncommon to help students conceive of samples and sampling in ways that support their developing coherent understandings of *why* statisticians have such confidence in this practice" (Saldanha & Thompson, 2002). Learning about the connections between sample and population involves building a schema of many interrelated ideas such as representativeness, sampling variability, and distribution. It would seem that two key factors in building such a schema are image and language.

METHOD AND TASKS

This pilot study, which is focused on investigating how to build sampling concepts for statistical inference, is situated in one Year 10 (14 year-olds) classroom. The only technology available to the class was one computer with access to the internet and a data show. A sequence of lessons, based around a web application designed by the researcher, was used to introduce students informally to sampling reasoning. The instruction was designed in collaboration with the teacher and in cognisance of time, bookwork expectations, and resource constraints. The aim of the lessons was to build embryonic sampling concepts in recognition of the fact that concepts take several years to develop (Shaughnessy, 2007). Although the study covered 13 lessons, this paper focuses on, the seventh and tenth web application lessons. Before the seventh lesson the focus of instruction was on introducing students to dot plots and then making the transition to box plots. The teacher took the final responsibility for teaching and therefore could modify or add to the activities provided. The lessons involved about 30 above average ability students from a multicultural girls' school, which is below average in socio-economic terms. The results presented in this paper are based on video recordings that I made in the classroom and an open-ended questionnaire completed by the students. Many students did not wish to be videotaped and therefore the focus was on the teacher. A qualitative analysis based on the theoretical framework of Liu and Thompson (2007) and a discourse analysis method (Woodward & Irwin, 2005) were conducted on the transcripts. The focus of these qualitative analyses was on how the teacher promotes conceptual growth and how her class interacts with her discussion.

The two web applications (www. censusatschool.org.nz) will be referred to as the Height Task and the Lunchtime Task (Fig. 1). Both tasks generate random samples from a database of 32,000 Year 5 to Year 10 students. The Height Task data are Year 10 girls' heights and the Lunchtime Task data were from responses to a question, "At school last week, what did you do *most* of the time at lunchtime?: sat down, stood around, walked around, ran around or played."



Height Task	Lunchtime Task

Figure 1. Screenshots of web application tasks (www. censusatschool.org.nz)

At the beginning of both tasks the teacher wrote on the board all the new language the students would be hearing. Whenever she introduced new language she pointed to the word on the board. Before each task she used a story scenario, engaged the students in a substantial discussion, and built on students' initial ideas about population, sample, sample size, and representative sample. Hence this discussion and the previous work on comparing data sets with box plots and dot plots formed the backdrop to students' introduction to sampling variability. To engage the students in the web applications, they were asked, for example, during the Height Task to predict and plot samples of Year 10 girls' heights (see Bakker, 2004 for a similar lesson design), check with the person sitting next to them whether they had the same centre and spread, check with web application samples of the same size, and write down what they noticed.

RESULTS

Since the purpose of the two tasks was to build concepts of sampling variability, an adaptation of the framework of Liu and Thompson (2007) for ways of thinking that lead to stochastic conceptions emerged from the analysis. The results are reported in terms of the framework (Fig. 2).

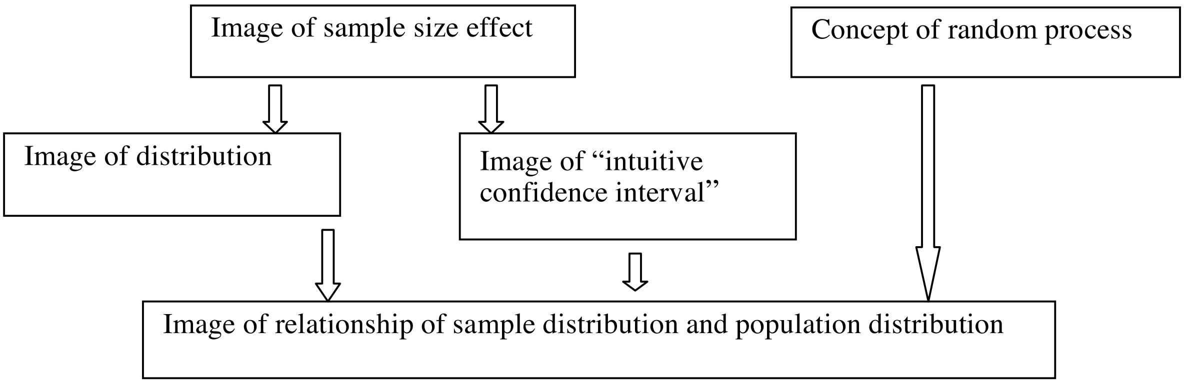


Figure 2. Framework for ways of thinking about sampling variability

Image of sample size effect

In order to obtain an image of the effect of sample size students need to notice variability among samples of the same size and variability among samples of different sizes and that as the sample size increases there is a point where the sample distributions are similar and therefore information can be gleaned about the population distribution. Since there are a myriad of concepts to attend to the teacher chose to focus on the images of the median and range for the Height Task, and the percentage of students who ran and the distribution shape for the Lunchtime Task. The teacher-student discourse to build up concepts of the sample size effect from images will now be discussed.

Image of "intuitive confidence interval"

As the samples were generated in the Height Task, attention was drawn to the median by marking on the board its place on the scale. In this way an interval of red marks emerged for samples of size 10, followed by an interval of black marks for samples of size 25, then green marks for size 63. A similar procedure was used for the percentage running in the Lunchtime Task.

When students were asked what they noticed as each sample was drawn, the teacher revoiced what they noticed and expanded their notice to consider new ideas such as variation.

Teacher: What is happening?
Student: Lower median
Teacher: Lower median, this time, variation in the median value
Teacher: For a sample size of 10 what's the median varied between?
Student: Between 160 and 166
Teacher: Between 160 and 166, all right, there's a lot of variability between the medians …

As she continued to revoice and expand on student responses, students started to tentatively use the new language they were hearing. For example, when the sample size was 25 the teacher drew attention to comparing the variability of the medians with the samples of size 10.

Teacher: Has the variability of the centre reduced as we increased the sample size?
Student: Yes
Teacher: Yes, slightly all right … the median variability decreased …
Student: The median variability seems to be about this distance (indicates distance)

By the teacher creating images of median intervals for different sample sizes and using revoicing as a way of associating new language to describe what the images meant the students were gradually induced to consider variability and the boundaries of variability.

Image of distribution

For the Height Task it was not clear that the students could build an image of the shape of the distribution as the teacher's focus was on the range and students intuitively noticed unusual values. From the start of the lesson the teacher used spread and range as synonyms. Even though students noticed and mentioned clustering of data she continued to draw their attention to the range. For example, when she compared samples of size 10 with size 25:

> Teacher: Did the range reduce or increase?
> Student: Increase
> Teacher: The range, the spread increased … and that is because we have more people.

In this excerpt another feature of her responses to students can be noted. She gives the grounds for the range increasing but does not give a warrant and say why sampling more people may increase the range. Such a method of argumentation was prevalent in the class discussion. Also students focussed very quickly on unusual values. The teacher responded by determining whether the unusual value was a mistake or a real person. For a height of 140cm, she compared the value with the shortest girl in the class, who was 145cm, and the class decided the value was possible. For another small value of 100cm, students posited reasons such as "midget", "might have no legs", "could be very smart and they were very young", all of which were deemed valid reasons.

For the Lunchtime Task the teacher focussed on the shape. For samples of size 10 she recorded the shapes on the board and got the students to describe them. She revoiced their words, enhanced their words towards better descriptions, glossed over their contextual considerations, and got them to reflect on all the shapes generated.

> Student: It goes down and then up again.
> Teacher: [She generates another sample distribution] This time, what happens?
> Student: symmetrical
> Teacher: It's symmetrical, it's uniform
> Student: It could like depend on the weather
> Teacher: … Weather, yes. Ok what can you tell me about the shapes of all these distributions? They are all completely?
> Student: Different
> Student: Individual

As she increased the sample size from 10 to 30 to 150 to 500 to 1000 she drew students' attention to the distribution "settling down". For homework the students were required to record what they had noticed. When some of them read out their concluding sentences the language was imprecise in describing the images they had seen (students 1 & 2) and possibly muddled about which words were appropriate for describing variability in a percentage and distribution shape (student 3).

> Student 1: As the sample size increases the percentage of people running would be generally the same.
> Student 2: As the sample size increases the variation decreases. [Teacher: Variation in what decreases?] Variation in the movements.
> Student 3: As the sample size increases the distribution size decreases.

For each student conclusion, the teacher got the students to clarify their meanings, revoiced their language and discussed better ways of describing the concepts behind the images.

Concept of random process

Before the Height Task the students had not been exposed to the concept of random sampling. Hence *stories* about why a plot was different from previous plots emerged. This was particularly noticeable after students had been thinking of possible reasons for unusual height values. For example:

> Students: Oooh!
> Student: The range is really small
> Teacher: The range is really small with these 10 people, but the median is quite high, the highest it has ever been
> Student: Perhaps it is because the taller ones get more sleep.

The teacher acknowledged her story and said that could be a further investigation. In the Lunchtime Task the stories continued despite the students having had a lesson on random sampling and a reminder about that lesson:

Teacher: Take a random sample. Who knows how you do that?
Student 1: You can just take a whole lot of people and be like, I like you, you, you.
Student 2: That's biased. That's not random.
Teacher: That's not random
Student 1: No, no, no. I don't mean like that. I mean like you have got a lot of information and you are like, I'll take that number, that number.
Teacher: Good. You are numbering the people. You are numbering a whole list of people.
Student 1: And just choosing the numbers.
Student 3: That's kind of like bingo.

This excerpt demonstrates that some students have some appreciation of how to take a random sample but from the class discussion not an appreciation of how a random sample behaves as attributing causes or stories to the random samples continued. In the excerpt below the teacher reinforces the idea that stories are inappropriate for random samples but some students were determined to argue that the people in the sample could know each other.

Student 1: One person stands. They might be a loner.
Teacher: Remember. Did these people know each other?
Student 1: No
Teacher: No, they are randomly selected. They didn't know each other. So you can't say he's a loner. Might have lots of mates.
Student 2: The person standing could have been standing around the people sitting.
Teacher: No. Remember, these people didn't know each other. This is a random sample.
Student 2: How do you know they didn't know each other then?
Teacher: Oh, that's true. They may, but remember there are 32,000 people.

The teacher gives the grounds that people are randomly selected for her claim that people did not know each other, but her reason that there are 32,000 people does not fully explain to the students why she can make this inference from her argument. Such a brief discussion and the positing of stories suggest that some students did not conceptualise the underlying random process and how a random sample behaves in terms of the context.

Image of relationship of sample distribution and population distribution

For the Height Task the teacher did not convey to the students the overall purpose that a sample is taken to make an inference about the population. At one stage she asked the researcher to assist her when she realised that she had lost the main idea of the task, since her focus was on the variability of the median and the range. After the researcher invention she started to refer to the connection between the sample and population. At the end of the lesson the teacher showed the students the population distribution of all Year 10 girls' heights. From a student's response and her use of the word sample, there seems to some connection being made between sample and population.

Teacher: Tell me about this whole population.
Student: The median is roughly the same as the sample
Teacher: The median is roughly the same as the sample, good and as we increased the sample size so it got closer and closer to the population median.

In the Lunchtime Task the teacher kept the focus on connecting the sample distribution to the population distribution as these two excerpts demonstrate. Sometimes she used the word average or mean instead of percentage.

Teacher: There is some mythical true percentage for what actually happens, how many people run at lunchtime. … From a sample of size 10, look at the variation, do you think the true population average is somewhere between these values for all Year 5 to 10 students? [points to interval 20 to 70%]. (Students were asked to put their hands up and most did agree.)
Teacher: As the sample size increases the variation in the sample proportion or percentage decreases. So if we go and select 500 people at random … do you think we are getting more confident about what our true population mean is going to be?

From the student responses there appeared to be some appreciation that as the sample size increased the sample could say something about the population. Two weeks after this lesson the students were given an open-ended questionnaire, in which one of the questions asked them to write down what they found interesting about the four-week statistics unit. Half the class mentioned what they had learnt from the Height and Lunchtime Tasks. Three responses were:

- I found it interesting when you only need a sample to work out problems from a large population. It's just really weird and cool at the same time that you just take any sample that is representative and it gives you near enough accurate answers.
- Comparing distributions with different sized sample graphs because the variation of the graph changing each time when the sample was small but having less variation when the sample size got larger.
- The part where we compares and saw all the different medians from different sample sizes. Coz it was really interesting to see it change. It's buzzy…

Their responses to the questionnaire coupled with their responses in class suggest students gained some intuitive notions of sampling variability, some linking of sample to population, some language associated with sampling, although tentative and imprecise, and dynamic images.

DISCUSSION

The main research question addressed was whether students could be stimulated to start developing a conceptual structure with which to conceive sampling variability. These findings suggest that the web application images, coupled with sense-making classroom discourse, stimulated in the students, in terms of sampling behavior, embryonic understandings of: the sample size effect; "intuitive confidence intervals" for medians and proportions; distributions for categorical data; and the relationship between sample distributions and population distributions. It seemed that ideas about the distributions for measurement data were not stimulated nor was there a connection made between sampling behavior and the underlying random generating process. However, as Shaughnessy (2007) stated, it is impossible with many interrelated concepts to attend to them all at the same time and that teachers should introduce concepts gradually without telling the students everything. The teacher, in this study, did not attend to all concepts, choosing to focus on the median and range in the Height Task and distribution, proportion of one category, and random sampling ideas in the Lunchtime Task.

The students' and teacher's attention on outliers and the range in the Height Task suggest that conceptions of variability started, according to Shaughnessy's (2007) hierarchy, at the very beginning. However, through the imagery of the "intuitive confidence intervals" conceptions were being formed about the possibility of measuring variability, which is in the later stages of his hierarchy. Therefore instruction with this web application can start with students' intuition and start to scaffold students towards higher-level concepts. This study also suggests that more tiers could be added to Watson's (2005) sampling reasoning hierarchy as knowing how to select a sample based on a random process does not mean that students know how such samples behave.

The findings seem to support Shaughnessy's (2007) claim that technology can enrich and deepen students' conceptions. The visual imagery presented drew students' attention to notice how particular features of the plots behaved when many random samples were drawn from a population. Even though access was limited to one computer and a data show, the images and the teacher-student discussion seemed to support concept formation, which confirm Bakker's (2004) findings. Since half the students mentioned the web-based applications in an open-ended questionnaire, their responses suggest that dynamic images linger in their minds with some associated statistical language and ideas.

When the teacher built meaning from the images she engaged in them noticing what had happened to a particular feature such as the median and built on their responses by using a revoicing technique to gradually introduce new ideas and language. In this way students tentatively started to articulate what they were seeing with some new language. The students struggled to articulate what they saw and experienced. Similarly, the teacher struggled at times to use the correct language, to verbalize the purpose of the tasks, and to synthesize the main ideas that were meant to emerge from the tasks. Developing sampling variability concepts in statistics is not the usual practice in classrooms (Watson, 2005), and therefore the teacher was a learner also when using the web application imagery with discussion.

One main issue that arose when the teacher was attempting to develop sampling variability concepts was contextual interference. My conjecture is that learners do not know when it is appropriate to *tell a story* and when it is not. It is appropriate to tell stories for extreme individual quantitative values, and it is appropriate to tell a story about the population from a random sample, but it is not appropriate to tell stories about individuals in the sample distributions and to compare sample distributions contextually when a random process generates them. Contextual knowledge seems to interfere with students being able to connect the hidden random generating process with the images they were seeing. In a similar vein, Rubin, Bruce, and Tenney (1991, p. 318) report that students believed that randomness was "not sufficient to explain sampling variability – some mechanism or bias must be postulated to explain it." Shaughnessy (20007) and Schwartz and Goldman (1996) also report the effect of context on students' reasoning. The ability to conceive that perceived patterns in data may be due to chance and not causes is part of learning about statistical inference. However, the dual ways of thinking about data, deterministic and non-deterministic, may take some years to develop.

The limitations inherent in this study are the focus on one teacher and her class, and the lack of other supporting evidence such as student interviews and responses to tasks. This pilot study, however, suggests that dynamic images produced by technology coupled with sense-making classroom discourse can start to promote conceptual growth in sampling reasoning. Furthermore, the creation of a sampling-reasoning pathway through the curriculum is necessary if students are to understand the logic of inference. More research should be conducted in this area.

REFERENCES

Bakker, A. (2004). Reasoning about shape as a pattern in variability. *Statistics Education Research Journal*, 3(2), 64-83, http://www.stat.auckland.ac.nz/serj

Chance, B., delMas, R., & Garfield, J. (2004). Reasoning about sampling distributions. In D. Ben-Zvi and J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 295-324). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Liu, Y., & Thompson, P. (2007). Teachers' understandings of probability. *Cognition and Instruction*, 25(2), 113-160.

Makar, K., & Confrey, J. (2005). "Variation-Talk": Articulating meaning in statistics. *Statistics Education Research Journal*, 4(1), 27-54, http://www.stat.auckland.ac.nz/serj

Ministry of Education (1992). Mathematics in New Zealand curriculum. Wellington, New Zealand: Learning Media.

Pfannkuch, M. (2006). Comparing box plot distributions: A teacher's reasoning. *Statistics Education Research Journal*, 5(2), 27-45, http://www.stat.auckland.ac.nz/serj

Pfannkuch, M. (2007). Year 11 students' informal inferential reasoning: A case study about the interpretation of box plots. *International Electronic Journal of Mathematics Education*, 2(3), 149-167, http://www.iejme.com/

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, pp. 314-319). Voorburg, The Netherlands: International Statistical Institute.

Saldanha, L., & Thompson, P. (2002). Conceptions of sample and their relationship to statistical inference. *Educational Studies in Mathematics*, 51, 257-270.

Schwartz, D., & Goldman, S. (1996). Why people are not like marbles in an urn: An effect of context on statistical reasoning. *Applied Cognitive Psychology*, 10, S99-S112.

Shaughnessy, M. (2007). Research on statistics learning and reasoning. In F. Lester (Ed.), *Second handbook of research on the teaching and learning of mathematics*, (Vol. 2, pp. 957-1009). Charlotte, NC: Information Age Publishers.

Watson, J. M. (2004). Developing reasoning about samples. In D. Ben-Zvi and J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 277-294). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Woodward, J., & Irwin, K. (2005). Language appropriate for the New Zealand Numeracy Project. In P. Clarkson, A. Downton, D. Gronn, M. Horne, A. McDonough, R. Pierce & A. Roche (Eds.), *Proceedings of the 28th Annual Meeting of the International Group for the Psychology of Mathematics Education*, (Vol. 2, pp. 799-806). Melbourne, Australia: PME.