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

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

Concerns about students' difficulties in statistical reasoning and a lack of research led to a study which explored form five (14 to 16 year olds) students' ideas in this area. The students were interviewed using open ended questions which related to sampling and variation, graphical representations and probability. This paper presents and discusses the ways in which students made sense of sampling and variability questions. While some students demonstrated sound reasoning, the findings revealed that many of the students used strategies based on prior experiences. The use of non-statistical responses (beliefs, everyday and school experiences) were considerably more common than that discussed in literature. The paper concludes by suggesting some implications for teachers and researchers.

In recent years statistics has gained increased attention in our society. Decisions concerning business, industry, employment, sports, health, law and opinion polling are made using an understanding of statistical information. Paralleling these trends, there has been a movement in many countries to include statistics at every level in the mathematics curricula. In western countries such as Australia (Australian Education Council, 1991) and New Zealand (Ministry of Education, 1992) these developments are reflected in official documents and in materials produced for teachers. Clearly the emphasis in these documents is on producing intelligent citizens who can reason with statistical ideas and make sense of statistical information.

Lajoie and Romberg (1998) note that in spite of its decade long presence in mathematics education, statistics is an area still in its infancy. Research shows that many students find statistics difficult to learn and understand in both formal and everyday contexts and that learning and understanding may be influenced by ideas and intuitions developed in early years (Gal & Garfield, 1997). Additionally, most of the research in statistics has been done with primary school children or with tertiary level students, resulting in a gap in our knowledge about students' conceptions of statistics at the secondary level.

Concerns about the importance of statistics in everyday life and in schools, the lack of research in this area and students' difficulties in statistical reasoning, determined the focus of this study. Overall, the study was designed to investigate what ideas do form five students have about statistics, and how do they construct them. This paper presents and discusses data obtained from sampling and sampling variability tasks.

Theoretical Framework

Much recent research suggests that constructivist models of knowledge provide a useful model of how students learn mathematics. Constructivist theory, in its various forms, is based on the 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. The constructions, based on beliefs and past experiences, appear to be the foundation for most learning. Constructivism provides the theoretical basis for understanding how students learn in this study. The idea that knowledge is constructed in the mind of the learner on the basis of pre-existing ideas means that students develop a meaning for statistical concepts if they can relate new knowledge to relevant concepts and propositions they already knew. This research was therefore designed to identify students' ideas, and to examine how they construct them.

Lack of Research in Sampling and Sampling Variation

Although it has been argued that sampling and sampling variation play a fundamental role in students' understanding and application of statistics and chance (Metz, 1997; Watson & Moritz, 2000), little research attention has been given to these concepts (Shaughnessy, 1997; Shaughnessy, Moritz, & Reading, 1999). Shaughnessy (1997) suggests some possible reasons for the lack of attention to research on variability. One is that 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 centres 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. A lack of emphasis on variation is also found in the mathematics curriculum documents in New Zealand and Australia.

Jacobs (1999) and Watson and Moritz (2000) have researched students' understanding of samples and sampling. Jacobs investigated through individual interviews (17 students) and written activities (110 students) grades 4 and 5 students' understanding of sampling issues in interpreting, evaluating and drawing conclusions from multiple surveys. Specifically, the students were asked to reason about the sampling method: random, restricted and self-selected methods in a raffle scenario. Jacobs reported that while some children demonstrated sound reasoning: they were aware of potential bias issues in surveys and favoured random sampling, other children focused on issues other than the potential for bias. They tended to base their evaluations on fairness and practical issues. The fairness rationale led students to include categories of every possible group in the survey because they believed that everyone should have the chance to participate. Jacobs also found that some students did not use the survey results to draw conclusions, rather based their thinking on other information such as personal experiences or opinions. Watson and Moritz (2000) investigated grades 3, 6, and 9 students' construction of the concept of sample. Data were collected from sixty-two students through interviews and written responses to a questionnaire were also analysed. They reported that as with other terms in mathematics, the term sample can take on different meanings in different contexts and created difficulties for the students. Watson and Moritz report that the students did not understand the importance of the sample being a representation of the whole population and the importance of the sample having variation.

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.

To investigate the influence of the concept of sample variability on student's 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.

It must be reiterated that the research discussed above has been done in a very few western countries, it would be interesting to determine how culture influences conceptions of sampling and sampling variability. Metz (1997) suggests that to adequately understand students' cognitive constructions and beliefs, we need to consider the culture in which students participate. In order to help inform teachers and curriculum designers, it appears to be crucial to carry out investigations at the secondary school level. Such information may help teachers plan learning activities and students overcome their difficulties.

Overview of the Study

The secondary school selected for the research was a typical high school. The sample consisted of a class 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 first 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 time the whole class was 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 samples and sampling variability were selected and adapted from those used by other researchers. The student meeting (Item 1A) and Saturday class (Item 1B) tasks were used to explore students' understanding of sampling issues in selecting samples.

Item 1A: Student meeting task

Five students are to be picked from your form to represent the form in a meeting. How would you pick the sample so that it is a fair representation of the class?

Item 1B: Saturday class task

It is thought that many people at the college do not like Saturday classes. To have the present practice changed, data needs to be collected. You have been asked to collect data to back up students' claims. Remember that not everyone at the school can be asked their opinion because there is not enough time. How would you go about doing this task?

The coin problem (Item 2) was used to explore students' ideas about sample variability.

Item 2: 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 the author 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 and Discussion

This section reports data on students' understanding of samples and sampling variability. The main focus 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 each of these sections the ways that the students have made their errors 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.

Sampling

All 14 students were able to respond in some reasoned kind of way to the two sampling tasks (Items 1A, 1B), with the reasoning of a minority on the tasks being clearly statistical. The reasoning of these statistical thinkers provides a marked contrast to that of the students with responses classed as non or partial-statistical. It is evident that their statistical thinking stemmed from both previous class activities and experiences beyond the class. When asked to select five students to represent the form in a meeting, four students used statistical approaches. For example, two students said they would take the form list, cut the names separately, roll them, put them in a box, shake the box to mix the names and then ask someone to pick five names. During teaching a number of sampling approaches were discussed. However, the approach used by these two students was not discussed in class. They appeared to have gained this understanding from outside class activities. Since fund-raising activities were going on in the school, students from this class were involved in organising raffles. The students used a procedure similar to the one described by the two students during the interview to select raffle winners. Student 25 used an approach that can be considered statistical. It seems that this student used an approach that was discussed in class. This is illustrated in the following response:

There are thirty students in our class, You want 5 students. So I will pick every sixth student from the form list.

Although four students were considered statistical on Item 1A, only two were considered statistical on the Saturday class task. The conversation with one of these students indicates that she believed in selecting a representative sample.

I: Can you tell me why you will select 10 students?

S12: Take 5 girls and 5 boys.

I: OK. How would you select these students?

S12: Like I have said before, if there are 30 students in each form, I will select the third student from the form list.

Non-statistical responses. Most student responses fell into this category. They dealt with the problems by ignoring elements of the questions or relying on irrelevant experiences. When asked to select five students to represent the form in a meeting so that it is a fair representation of the class, three students said they would select 2 boys and 3 girls or 3 boys and 2 girls. When pressed on the reasons for choosing 3 girls and 2 boys, the students said if they chose all girls, the boys would dispute the choice. For example, student 26 said, If I chose all girls then the boys will start arguing why I chose all girls. Seven students based their reasoning on classroom experiences. They did not take any care in picking a sample that was a fair representation of the whole class. The following discussion is indicative of this reasoning:

> It is a fair sample because some students in our form hardly talk and when they go and talk in the front the voice is so slow that we can't hear. We are sitting in front, still we can't hear and they hardly talk.

The students thought that they should choose students on the basis of intelligence or communication ability. They did not realise that their samples were biased towards bright and talkative students. The idea of representation had no statistical meaning for these students but seemed to be based on political astuteness. When asked to consider how student opinion could be obtained on whether to have Saturday classes, given that there was insufficient time to consult all students, four students thought that they should consult all students doing external examinations. In these responses, personal bias appeared to play a part in the students' reasoning too. The other four students tended to think that they should consult all students in the school about the Saturday class preference. These students preferred to ask everyone rather than take any sample. It must be noted that the all the sampling tasks used in this study focus on in-school situations because they are familiar to students. Perhaps including an out–of--school situation with a bigger population would have discouraged the students from asking everyone. Two students said that there was no need to have Saturday classes because fund-raising activities were going at the school and hence some students would like to organise activities such as soccer to raise funds for their class. Student 20 said that some students, especially the boys, could not come for Saturday classes because they had to work in the market.

Although this study provides evidence that reliance upon experience can result in biased, non-statistical reasoning, in some cases this strategy may provide useful information for other purposes. For example, the students' linguistic skills would have been reasonable. The responses raise further questions. First, is there a weakness in the wording of these tasks in that they are completely open-ended and do not focus the student to draw on other relevant knowledge? For instance, the term represent (Item1A) could have presented a linguistic problem as there are two perfectly reasonable interpretations. Perhaps, including strong cues in the items for the students to think of these in a statistical way would have been useful. Second, are the students aware of the differences in statistical reasoning compared with reasoning in other curriculum areas?

Partial-statistical response. The four students who gave partial-statistical responses on the Saturday class problem showed some understanding in selecting a sample. For example, they said that they should select a small sample from each form. However, they did not use statistical approaches for selecting students from each form. Student 5 used a similar approach. She said that her sample will be fair because she will select both boys and girls. Additionally, this student thought that she should choose students who are smart.

Student 17 applied the average rule inappropriately on the Saturday class task. She said that she would ask form four and form six students because they have to sit for examinations. However, when asked the number of students she would select from each form, she explained,

> If there are thirty students in each form, then the sum of the scores is sixty. The average is 30. Then I am going to ask 30 students altogether.

It seems that the students who inappropriately focused on the fairness issues when selecting their samples were not thinking of fair in the statistical sense of whether the sample would be fair, that is, whether everyone has an equal chance of being selected so that the sample is not biased. Rather, they were concerned about equitable issues in the school, everyone should participate.

Sample Variability

Of the 14 students, one student, student 9, said that both Anita and Shelley were equally likely to get 80% or more heads. However, when asked to explain, she said that she had just guessed the answer. None of the students 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. Four 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 other two students with responses in this category thought that the flipping of coins depends on luck or how one tosses the coin. For example, student 20 explained,

> 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. May be the student did not understand the question and readily performed arithmetic operations on the numbers given in the problem. 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.

Another common strategy used can be classed as equiprobability. 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 the interviewer did not induce any statistical thinking. 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. One student gave the correct answer with partially correct reasoning,

> Anita, because she does it fewer times.

Sampling and Variability: A Broader Context

It is clear from this study that students appear to have many misconceptions regarding sampling variability. According to Shaughnessy (1997) the representativeness strategy underlies the sample size misconception. The students did not rely on the representativeness strategy but based their thinking on the unpredictability and equiprobability bias. Moreover, results show that students did not explicitly use words dealing with sampling (random, bias) or variation (spread, deviate). They entertained several ideas about the concepts of sample, representative and fair, most of which were based on what they had absorbed informally from their everyday environments. These findings are similar to those reported by Jacobs (1999), Watson and Moritz (2000) and Shaughnessy et al. (1999). Although most grade 9 students in Watson and Moritz (2000) study appreciated the need for larger samples and appropriate methods of selection, they did not possess the language to provide convincing arguments based on representation and avoiding bias.

The finding that students base their thinking in statistics on their prior experiences is not new. The results certainly concur with the views expressed by Jacobs (1999) and Rubin et al. (1991). 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. In some respects, the findings of the present investigation go beyond those discussed above. The findings demonstrate that at times the in-school experiences appear to have had a positive effect on the students' thinking.

Implications of the Study for Teachers and Research

It must be reiterated that the open-ended nature of the tasks and the lack of guidance to students about the associated expectations certainly influenced how students explained their understanding. The students may not have been particularly interested in these types of questions as students are not used to having to describe their reasoning. Despite these drawbacks, the findings of the study have several implications for teachers and research.

First, when beginning instruction on sampling, it is important for the teachers to know the individual abilities of the students. Once the level of understanding has been explored, it is crucial for teachers to teach accordingly. They must probe for students reasoning behind the answers because sometimes students give correct answers for incorrect reasons.

Second, teachers 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 and toss was used to clarify the situation. This would have not been possible in a written survey.

Third, it seems that sampling and sampling variation are complex constructs and require more emphasis and explicit teaching. More space and time needs to be allocated to developing students' ability to make sense of and communicate ideas about these concepts.

Teachers should teach appropriate methods for selecting samples and help students see situations where bias can occur. Students should be encouraged to collect data from both in-school and out-of-school situations and verbalise the relationships and patterns among the different data sets. Such integration will enable students to construct meanings from data and build the foundation for an understanding of statistical variation.

Another suggestion relates to the variability concept. It would be helpful to include more items using different contexts in order to explore students conceptions of sampling variability and related concepts in much more depth and provide information on patterns of conceptual development.

References

Australian Education Council (1991). *A national statement on mathematics for Australian schools.* Canberra: Curriculum Corporation.

Gal, I., & Garfield, J. (1997). Assessing statistical knowledge as it relates to students' interpretation of data. In I. Gal & J. B. Garfield (Eds.), *The assessment challenge in statistics education* (pp. 1-13). Amsterdam, The Netherlands: IOS Press.

Garfield, J. B., & Gal, I. (1999). Teaching and assessing statistical reasoning. In L.V. Stiff. & F. R. Curcio (Eds.), *Developing mathematical reasoning in grades K*-12. (pp. 207-219). Reston, Virginia: National Council of Teachers of Mathematics.

Jacobs, V. (1999). How do students think about statistical sampling before instruction? *Mathematics Teaching in the Middle School, 5*(4), 240-246, 263.

Lajoie, S. P., & Romberg, T. A. (1998). Identifying an agenda for statistics instruction and assessment in K-12. In S. P. Lajoie (Ed.), *Reflections on statistics: Learning, teaching, and assessment in grades K-12. Mahwah* NJ: Lawrence Erlbaum Associates.

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

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 & K. Carr (Eds.), People in mathematics education. *Mathematics Education Beyond 2000. Proceedings of the 23[d] Annual Conference of the Mathematics Education Research Group of Australasia* (pp. 6-22). Hamilton: MERGA INC.

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

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.

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.

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 & 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.