Directions: Journal of Educational Studies Vol. 25 Nos 1 &2 June & Dec. 2003

An Exploration of High School Students' Understanding of Sample Size and Sampling Variability: Implications for Research

Sashi Sharma

Concerns about students' difficulties in statistical reasoning led to a study which explored Form Five (14 to 16-year-olds) students' ideas in this area. The study focussed on sampling, probability, descriptive statistics and graphical representations. This paper presents and discusses the ways in which these students made sense of sampling constructs obtained from the individual interviews. The findings revealed that many of the students used strategies based on prior experiences and intuitive strategies. While they showed competence with sample size, they were less competent on the sampling variability task. This could be due to instructional neglect of this concept or linguistic problems. The paper concludes by suggesting some implications for further research.

Introduction

In recent years statistics has gained increased attention in our society. Any newspaper or magazine is likely to contain statistical information. Decisions concerning business, industry, employment, sports, health, law and opinion polling are made, using understanding of statistical information (Wallman 1993). 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), New Zealand (Ministry of Education 1992), and the United Kingdom (Holmes 1994) these developments are reflected in official documents and in materials produced for teachers. In line with these moves, Fiji has also produced a new mathematics prescription at the primary level that gives more emphasis to statistics at this level (Fiji Ministry of Education 1994). The use of relevant contexts and students' experiences and understandings is recommended for enhancing the students' understanding of statistics (Watson 2000; Watson and Callingham 2003). Clearly the emphasis in these documents is on producing intelligent citizens who can reason with statistical ideas and make sense of statistical information. 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 and Garfield 1997). 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. Additionally, since most of this research has been done in just a very few western countries, it needs to be determined how culture influences conceptions of statistics; whether biases and misconceptions of statistics are artifacts of western culture or whether they vary across cultures.

Concerns about the importance of statistics in everyday life and in schools, the lack of research in this area and the difficulties students have in statistical reasoning, determined the focus of my study. Overall, the study was designed to investigate what ideas a group of Form Five students have about statistics, and how they construct these. Since the study is qualitative in nature, the sample was unavoidably small. By adopting a qualitative approach, I intended to gather rich, detailed and comprehensive data that enabled me achieve a better understanding of student thinking. The paper will not make any claims to universality or permanence. Rather, its interest lies in the specific, idiosyncratic context of *this* group of high school students. The study could be viewed as window or lens through which other people will be able to understand how some students construct meanings for statistical concepts. Prior to discussing the details of my own research, I will briefly discuss some existing literature on sampling and variation.

Research on sample size and sampling variation

Gal and Garfield (1997) express the need to study samples instead of populations and to infer from samples to populations. Additionally, they suggest that one goal for instruction in statistics is to help students understand the existence of variation. Metz (1997) and Moore (1990) highlight the key role that variation plays in students' understanding of chance. Moore writes that it is important for students to understand the idea that "chance variation, rather than deterministic causation, explains many aspects of the world" (Moore 1990:99). The New Zealand Ministry of Education (2004) states that since the idea of probability as long-run relative frequency needs to be addressed with students, variation can no longer be avoided.

Although it has been argued that sampling and variation play a fundamental role in students' understanding and application of statistics and chance, little research attention has been given to these concepts (Green 1993; Shaughnessy 1997; Shaughnessy, Watson, Moritz and Reading 1999; Torok and Watson 2000). Shaughnessy (1997) suggests some possible reasons for the lack of attention to research on variability concepts. The neglect of variation is noted in the National Assessment of Educational Progress (Zawojewski and Heckman 1997) studies which tested student achievement in grades 4, 8, and 12. The statistics assessment test items addressed concepts of descriptive statistics, mainly median and mean, as measures of central tendency, and range as a measure of variability. However, it was a low level computational task.

To illustrate the undue confidence that people put in the reliability of small samples, take Tversky and Kahneman's (1974) problem given to college 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. 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%. According to Tversky and Kahneman (1974) the representativeness heuristic underlies this misconception. People who rely on the representative heuristic tend to estimate the likelihood of events by neglecting the sample size or by placing undue confidence in the reliability of small samples.

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 (Shaughnessy and Zawojewski 1999). 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 growth 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.

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 variability: whether intuitive strategies and preconceptions such as representativeness are artifacts of western culture or whether they vary across cultures. Metz (1997) suggests that to adequately understand students' cognitive constructions and beliefs, we need to consider the culture in which students participate. Such information may help teachers to plan learning activities and students to overcome their difficulties. In the current interview-based study, both sampling and variability questions were used to determine specific student conceptions and the factors that contribute to these constructs.

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. Fourteen students were selected from this class and this group was representative of the larger group in terms of ability and gender.

To explore the full range of students' thinking about sampling concepts, open-ended questions were selected and adapted from those used by other researchers. The car question (Item 1) was used to assess whether students understand that it is more important to base decisions on large samples of data than on single cases or small samples, however compelling they might seem. The coin problem (Item 2) was used to explore students' ideas about sample variability.

Item 1: The car problem

Mr Singh wants to buy a new car, either a Honda or a Toyota. He wants whichever car will break down the least. First he read in *Consumer Reports* that for 400 cars of each type, the Toyota had more break-downs than the Honda. Then he talked to three friends. Two were Toyota owners, who had no major break-downs. The other friend used to own a Honda but it had lots of break-downs, so he sold it. He said he would never buy another Honda. Which car should Mr Singh buy? Why?

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?

Each student was interviewed individually by myself in a room away from the rest of the class. During the interview, care was taken to avoid leading the students towards any particular viewpoint, so responses to questions were accepted as they were given and probing questions were asked simply to ascertain the reasons for what the student thought. The interviews were 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.

Analysis of data

Analysis of the interviews indicated that the students used a variety of intuitive strategies and prior knowledge to explain their thinking. I created a simple four stage-based model that could be helpful for describing research results relating to students' statistical conceptions, planning instruction in statistics and dissemination of findings to mathematics educators. The four categories in the model are: non response, non-statistical, partial-statistical and statistical. They are described in Table 1.

Table 1: Characteristics of the four categories of responses

Response Type	Sample size and sampling variation
Non-response	Complete silence, I don't know, I have forgotten the rule, I just guessed
Non-statistical responses	Refer to everyday and school experiences or make inappropriate connections with other learning areas Problems with language Hold the pervasive belief that they can control outcomes of events Causality perspective Belief in luck
Partial-statistical responses	Adapted the rules or applied them inappropriately. Refer to representativeness, unpredictability, equiprobability biases. Could not explain reasoning Inconsistent reasoning Focus on only a small subset of the information
Statistical responses	Able to justify reasoning by using frequentist interpretation Extend sampling rules to unfamiliar situations

The non-statistical responses were based on beliefs and experiences while the students using the partial-statistical responses applied rules and procedures inappropriately or referred to intuitive strategies. The term statistical is used in this paper for the appropriate responses. However, I am aware that such a term is not an absolute one. Students possess interpretations and representations which may be situation specific and hence these ideas have to be considered in their own right. This category has been used mainly to discuss and present results. Statistical simply means what is usually accepted in standard mathematics text-books. It would be reasonable to assume another level (advanced-statistical), equivalent to Shaughnessy's (1992) pragmatical statistical level, where students appear to have a very complete view that incorporates questioning of data but the need for such a category did not arise in my research and any responses that could have been categorised as advanced-statistical were simply grouped with the statistical responses.

Results and discussion

This section reports data on students' understanding of sample size and sampling variability. In this section the types of responses are summarised and the ways that the students made their errors are described. Extracts from typical individual interviews are used for illustrative purposes.

Interview Responses: Item 1

Results of student responses to Item 1 are summarised in Table 2.

Table 2: Response types for task involving sample size (*n* = 14) .

Response type	Number of students using it
Non-response	1
Non-statistical	9
Partial-statistical	-
Statistical	4

Four students were considered statistical on Item 1. They were able to apply the law of large numbers and thought that one should believe a consumer report because it represents a bigger sample.

***Non-statistical responses*:** The nine students who did not use sample size information on Item 1 based their responses on their beliefs and experiences. When asked to make predictions, students looked for factors that caused the behaviour or event under consideration. Rather than attending to sample size in a consumer report, three students in this study said that Mr Singh could buy either car because the life of a car depends on how one keeps it. They did not apply the idea of representativeness in this instance where it is really appropriate to do so. For example, student 29 explained:

He should buy any of the cars Honda or Toyota; it depends on him how he keeps and uses the car ... Ah ... Because it depends on the person, how he follows instructions then uses it. My father used to own a car and he kept it for ten years. He sold it but it is still going and it hasn't had any major breakdowns.

Two students thought that Mr Singh should buy a Toyota. They generalised inappropriately from experiences with small samples. For instance, student 17 responded:

Because his friends have had experience with their cars. They are saying that they had no major break downs. The other friend had Honda and he had many break downs.

The other four students drew upon their everyday experiences with consumer reports. Two of these students thought that Mr Singh should take advice from a consumer report because they were the right people to consult. The remaining students felt that Mr Singh should not take advice from the consumer reports because consumer people often give misleading information. For instance, student 9 explained:

Maybe they want to sell Toyota cars first. We read in the Fiji Times that people want to sell things; they just advertise that they are good. When people buy it, is not like that.

Interview Responses: Item 2

Results of student responses to Item 2 are summarised in Table 3.

Table 3: Response types for task involving sample variability (*n* = 14)

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

One student said that 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. As Table 2 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 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 depends 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 and 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.

The item produced very strong unpredictability reasoning. 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 seems to indicate a view that chance is naturally equiprobable. Even repeated probing by the interviewer 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:

Interviewer:	*Can you explain your answer?*
Student 2:	*Because a coin has two sides only heads or tails. One tosses it 10 times or 100 times, it makes no difference.*
Interviewer:	*So you think that both will get 80% or more heads?*
Student 2:	*80% or more heads, no [laughs]. Not 80%. They will be getting 50% or more heads.*
Interviewer:	*Can you say why?*
Student 2:	*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.*

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. Furthermore, students are not used to explaining their thinking or perhaps they had difficulty in explaining their reasoning in detail due to language difficulties. It takes considerable self-confidence to say something like *I can't describe my thinking in words or I don't understand the question.* Gal (1998) states that suggesting to students that a judgment 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.

Sampling and variability: a broader context

Intuitive Strategies: 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 uncertainity 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.

Outcomes can be controlled: The results show that quite a number of students think that outcomes on random generators such as coins or dice 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). 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' (i.e. not loaded) in item 2 would have aided in the interpretation of this question. Are the students aware of the differences in probabilistic reasoning compared with 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. It does not make sense to say that the coin has a probability of one-half to be heads because the outcome can be completely determined by the manner in which it is thrown.

Relevant contexts: 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 and Querelle 2001; New Zealand Ministry of Education 1992) and research showing the effects of contexts on students' ability to solve open-ended tasks (Cooper and Dunne 1997; Sullivan, Zevenbergen and Mousley 2002). For instance, the study by Cooper and Dunne showed that realistic problems disadvantaged working class children since middle class children had greater linguistic facility.

Limitations

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 students in this sample 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 important for these students, for whom the language of instruction is a second language for them, one that is not spoken at home. Another reason could be that such questions do not appear in external examinations. Although the study provides some valuable insights into the kind of thinking that high school students use, the conclusions cannot claim generality because of the small sample. Additionally, the study was qualitative in emphasis and the results rely heavily on my skills to collect information from students. Some implications 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 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 student's understanding of the context contributes to his/her interpretation of sampling data. Research on what makes this translation difficult for students is needed.

The picture of students' thinking in regard to sampling is somehow limited because students responded to only one item related to sample size and one item related to variability. There is a need to include more items using similar contexts in order to explore students' conceptions of sampling in much more depth. It would be interesting to explore how changing the context of the task influences the types of responses exhibited by the students.

Researchers can accurately assess their subjects' understanding through individual interviews. The interview results provide evidence that students often experience difficulty when speaking about tables. However, in the present investigation I overcame these difficulties by restating a task or changing the wording. This would have not been possible in a written survey.

Finally, the place of statistics has changed in the revised mathematics prescription. Statistics appears for the first time at all grade levels (Fiji Ministry of Education, Women, Culture, Science and Technology 1994). 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

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

Cooper, B. and M. Dunne. 1999. *Assessing children's mathematical ability.* Open University press, Buckingham.

Fiji Ministry of Education, Women, Culture, Science and Technology. 1994. *Primary mathematics prescriptions.* Curriculum Development Unit, Suva.

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.* Lawrence Erlbaum Associates, Mahwah NJ: 275-295.

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

Green. D. R. 1993 Data analysis: What research do we need? In L. Pereira-Mendoza (ed) *Introducing data analysis in the schools: Who should each it and how?* Proceedings of the International Statistical Institute Round Table Conference. Lennoxville, Canada, August 19-14, 1992. International Statistical Institute, Voorburg: 219-239.

Holmes, P. 1994. Teaching statistics at school level in some European countries. In L. Brunelli and G. Cicchutelli (eds) *Proceedings of the first scientific meeting.* University of Perugia: 3-11.

Metz, K. E. 1997. Dimensions in the assessment of students' understanding and application of chance. In I. Gal and J. B. Garfield (eds) *The Assessment challenge in statistics education.* IOS Press, Amsterdam, The Netherlands:1-13.

Meyer, M., T. Decker and Querelle. 2001. Context in mathematics curricula. *Mathematics Teaching in the Midde School* 6 (9): 522-527.

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

Ministry of Education. 1992. *Mathematics in the New Zealand curriculum.* Ministry of Education, Wellington.

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

Rubin, A., B. Bruce and Y. Tenney. 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.* International Statistical Institute, Voorburg, The Netherlands:314-319.

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

Shaughnessy, J. M., J. Watson, J. Moritiz and C. Reading. 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. 1992. Research in probability and statistics: Reflections and directions. In D. Grouws (ed) *Handbook of research on mathematics teaching and learning*. Macmillan, New York: 465-494.

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

Sullivan, P., R. Zevenbergen and J. Mousley. 2002. Context in mathematics teaching: Snakes or ladders? In B. Barton, K.C. Irwin, M. Pfanncuch 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.

Torok, R. and J. Watson. 2000. Development of the concept of statistical variation: An exploratory study. *Mathematics Education Research Journal*, 12(2): 147-169.

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

Wallman, K. K. 1993. Enhancing statistical literacy: Enriching our society. *Journal of the American Statistical Association* 88 (421):1-8.

Watson, J. M. 2000. Statistics in context. *The Mathematics Teacher* 93(1): 54-58.

Watson, J. M. 2004. Quantitative literacy in the media: An arena for problem solving. *Australian Mathematics Teacher* 60(1): 34-40.

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

Zawojewski, J. S. and D. S. Heckman. 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*. National Council of Teachers of Mathematics, Reston, VA: 195-223.