STUDENTS AND TEACHERS' KNOWLEDGE OF SAMPLING AND INFERENCE[1]

Anthony Harradine[1], Carmen Batanero[2], and Allan Rossman[3]

[1]Potts-Baker Institute, Prince Alfred College, Australia; [2]University of Granada, Spain;

[3]California Polytechnic State University, United States of America

aharradine@pac.edu.au; batanero@ugr.es; arossman@calpoly.edu

Abstract: *Ideas of statistical inference are being increasingly included at various levels of complexity in the high school curriculum in many countries and are typically taught by mathematics teachers. Most of these teachers have not received a specific preparation in statistics and therefore could share some of the common reasoning biases and misconceptions about statistical inference that are widespread among both students and researchers. In this chapter the basic components of statistical inference, appropriate to school level, are analysed, and research related to these concepts is summarised. Finally, recommendations are made for teaching and research in this area.*

1. INTRODUCTION

Statistical inference, in the simplest possible terms, is the process of assessing strength of evidence concerning whether or not a set of observations is consistent with a particular hypothesised mechanism that could have produced those observations. It is an essential tool in management, politics and research; however, people's understanding of statistical inference is generally flawed. The application and interpretation of standard inference procedures is often incorrect (see, for example Harlow, Mulaik, & Steiger, 1997; Batanero, 2000; Cumming, Williams, & Fidler, 2004).

Because of the relevance and importance of statistical inference, education authorities in some countries include a basic study of statistical inference in the curriculum of the last year of high school (17-18 year olds). For example, South Australian and Spanish students learn about statistical tests and confidence intervals for both means and proportions (Senior

[1] Published in C. Batanero, G. Burrill, & C. Reading (Eds.), Teaching Statistics in School-Mathematics-Challenges for Teaching and Teacher Education: A Joint ICMI/IASE Study (pp. 235- 246), DOI 10.1007/978-94-007-1131-0, Springer Science+Business Media B.V. 2011. The original publication is available at http://www.springerlink.com/

Secondary Board of South Australia, 2002; Ministry of Education and Sciences, 2007). New Zealand students learn about confidence intervals, resampling and randomisation (Ministry of Education, 2007).

Some of the fundamental elements of basic inference are implicitly or explicitly included in various middle school curricula, as well. For example, the National Council of Teachers of Mathematics (NCTM) Standards (2000) suggest that Grades 6–8 students should use observations about differences between two or more samples to make conjectures about the populations. NCTM further recommends that grades 9-12 should use simulations to explore the variability of sample statistics from a known population and to construct sampling distributions; they also should understand how a sample statistic reflects the value of a population parameter and use sampling distributions as the basis for informal inference. More recently, the American Statistical Association's Guidelines for Assessment and Instruction in Statistics Education (GAISE; Franklin et al, 2005) highlights the need for students to look beyond the data when making statistical interpretations in the presence of variability and urges that students in middle grades recognize the feasibility of conducting inference and that high school students learn to make inferences both with random sampling from a population and with random assignment to experimental groups.

This chapter analyses the basic elements of statistical inference and then summarises part of the wider research that is relevant to teaching this topic (see Vallecillos, 1999; Batanero, 2000 and Castro-Sotos, Vanhoof, Noortgate, & Onghena; 2007 for an expanded survey). The chapter finishes with some implications for teaching and research.

2. STATISTICAL INFERENCE – A RICH MELTING POT

Classical statistical inference consists primarily of two types of procedures, hypothesis testing and confidence intervals. These techniques build on a scheme of interrelated concepts including probability, random sampling, parameter, distribution of values of a sample statistic, confidence, null and alternative hypothesis, p-value, significance level, and the logic of inference (Lui & Thompson, 2009).

Consequently, statistical inference consists of three distinct, but interacting, fundamental elements: (a) the reasoning process, (b) the concepts and (c) the associated computations. Because the computations are often easily learned by students, and can be facilitated by user-friendly software, teachers of statistics must teach the three components and not just the mechanics of inference, because the main difficulties in understanding statistical inference lie within the other two elements.

2.1. The reasoning process

Garfield and Gal (1999) suggest that, across the primary, middle and high school years, teachers must develop students' statistical reasoning – the processes people use to reason with statistical ideas and make sense of statistical information. This process is supported by concepts such as distribution, centre, spread, association, uncertainty, randomness and sampling, some of which have been analysed in other chapters in this book. While most students may be able to perform the calculations associated with an inferential process, many students hold deep misconceptions that prevent them from making an appropriate interpretation of the result of an inferential process (Vallecillos, 1994; Batanero, 2000; Castro-Sotos, et al., 2007). In addition, Garfield (2002) remarks that some teachers do not specifically teach students how to use and apply types of reasoning but rather teach concepts and procedures and hope that the ability to reason will develop as a result. As a consequence, students reach their first inferential reasoning experience with a reasoning-free statistical background, giving rise to a mind-set that statistics is solely about the computation of numerical values. One possible reason for this unfortunate circumstance is that teachers responsible for teaching statistics at a high school level may have serious deficiencies in their knowledge that lead to inadequate understandings of inference (Liu & Thompson, 2009).

2.2. The concepts

Central to learning statistical inference is understanding that the variation of a given statistic (e.g. the mean) calculated from single random samples is described by a probability distribution – known as the sampling distribution of the statistic. When thinking about statistical inference it is necessary to be able to clearly differentiate between three distributions:

- The *probability distribution* that models the values of a *variable* from the population/process. This distribution usually depends on some (typically unknown) parameter values. For example, a normally distributed population is specified by two parameters - its mean and standard deviation, often denoted by μ and σ.

- The *data distribution* of the values of a *variable* for a *single* random sample taken from the population/process. From this sample sample statistics such as the mean and standard deviation, often denoted by $\bar{x}$ and s, can be used in the process of estimating the unknown values of the population parameters.
- The *probability distribution* that models the variability in values of a statistic from 'all' potential random samples taken from the population/process, called the sampling distribution. One example is the sampling distribution of a sample mean, which in many circumstances has an approximately normal distribution with mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$, where n represents the sample size. This result provides the basis for much of classical statistical inference.

Sampling distributions are more abstract than the distribution of a population or a sample and so are typically very challenging for students to understand (see section 3.2). One reason for this difficulty is that when thinking about both the population distribution and the single random sample's distribution, the unit of analysis (case) is an individual object. This is in stark contrast to the sampling distribution where the case is a single random sample (Batanero, Godino, Vallecillos, Green, & Holmes, 1994). The object of interest for each distribution might be the mean, for example, but in each case the distribution's mean has a different interpretation and a different behaviour. One strategy for helping students to understand these distinctions is to engage in activities that involve repeatedly taking random samples from a population. When working with such activities, high school students often struggle with moving between the various levels of imagery (Saldahna & Thompson, 2002). Proper application and interpretation of statistical inference requires mastery of the knowledge and techniques specific to each distribution and understanding of the rich links among these distributions.

3. DIFFICULTIES IN UNDERSTANDING STATISTICAL INFERENCE

Research reviewed in this section deals with understanding sampling and the sampling distribution, hypothesis tests and confidence intervals.

3.1. Understanding sampling

Research on inferential reasoning started with the *heuristics and biases* programme of research in psychology (Kahneman, Slovic, & Tversky, 1982), which established that most people do not follow the normative mathematical rules that guide formal scientific inference when they make a decision under uncertainty. Instead, people tend to use simple judgmental heuristics that sometimes cause serious and systematic errors, and such errors are resistant to change. For example in the *representativeness* heuristics, people tend to estimate the likelihood for an event based on how well it represents some aspects of the parent population. An associated fallacy that has been termed *belief in the Law of Small Numbers* is the belief that even small samples should exactly reflect all the characteristics in the population distribution.

Most curricula at a high school level include some instruction on random sampling, which is mostly theoretical and includes descriptions of different methods of random sampling. The core message of such instruction is that if a sample is chosen in a *suitable random manner* and is *sufficiently big*, it will be representative of the population from which it has been drawn. Students therefore learn to think about a random sample as a *mini-me* of the population and that the purpose of drawing a random sample is to ensure *representativeness* in order to gain knowledge about the population from the sample. This conception constrains students' thinking to a single random sample only and provides no avenue to appreciate the range of possible samples that *might have been drawn* and the variability across that range.

Understanding the purpose of drawing a single random sample in the context of hypothesis tests and confidence intervals, requires the assimilation of "two apparently antagonistic ideas: sample representativeness and (sampling) variability" (Batanero et al, 1994). In these situations the purpose of drawing a single sample is to quantify that sample's level-of-unusualness relative to the many other samples that could have been drawn. Saldahna and Thompson (2002) observed that, without a suitable sense of the variation across many possible samples, which extends to the notion of the distribution of a statistic, 11th and 12th grade students tended to judge a sample's representativeness only in relation to the population parameter. Hence, when required to decide how rare a sample was, these students did so based on how different they thought it was to the underlying population parameter and not "on how it might compare to a clustering of the statistic's values" (Saldanha & Thompson, 2002).

3.2. Understanding sampling distributions

Reasoning about sampling distributions requires students to integrate several statistical concepts and to be able to reason about the hypothetical behaviour of many samples – an intangible thought process for many students (Chance, Delmas & Garfield, 2004). According to these authors, many students fail to develop a deep understanding of the sampling distribution concept and as a result can only manage a mechanical knowledge of statistical inference, leaving such tasks as interpreting a *p*-value well beyond those students.

Saldahna and Thompson (2002) studied the understandings of high school students when engaged in activities that used computer applets to simulate repeated random sampling from a population. The activity required students to randomly draw a sample from a population, compute a sample proportion and then repeat this process over and over. They found that most students had extreme difficulty in conceiving of repeated sampling in terms of three distinct levels: population, sample, collection of sample statistics. These difficulties led many students to misinterpret a simulation's result as a percentage of people rather than a percentage of sample proportions.

Chance et al. (2004) found that while students were able to observe behaviours and notice patterns in the behaviour (e.g. larger the sample size smaller the variation) shown by random sampling applets, they did not understand why the behaviour occurred. The authors noted that, after exposure to applets, students were unable to suggest plausible distributions of samples for a given sample size and agreed with Saldahna and Thompson that students did not have a clear distinction between the distribution of one sample of data and the distribution of means of samples. Simply being exposed to the applets was not sufficient to render a learning gain. The authors concluded that: (a) students need to become more familiar with the process of sampling, (b) activities associated with applets need to be both structured and unstructured, and (c) students need to discuss their observations after an activity so they could become focussed on what observations are most important, what important observations they did not make and how the important observations are connected.

3.3. Understanding the null and alternative hypotheses

Errors and misinterpretations in hypothesis tests can lead to a paradoxical situation, where, on one hand, a significant result is often required to get a paper published in many journals and, on the other hand, significant results are misinterpreted in these publications (Falk & Greenbaum, 1995). There is confusion between the roles of the null and alternative hypotheses as well as between the statistical alternative hypothesis and the research hypothesis (Chow, 1996). Vallecillos (1994) reported that many students in her research, including 6 out of 31 pre-service mathematics teachers, believed that correctly carrying out a test proved the truth of the null hypothesis, as in the case of a deductive procedure. Vallecillos (1999) described four different conceptions regarding the type of proof that hypotheses tests provide: (a) as a decision-making rule, (b) as a procedure for obtaining empirical support for the hypothesis being researched, (c) as a probabilistic proof of the hypotheses, and (d) as a mathematical proof of the truth of the hypothesis. While the two first conceptions are correct, many students in her research, including some pre-service teachers, held either conception (c) or (d).

Belief that rejecting a null hypothesis means that one has proven it to be wrong was also found in the research by Lui and Thompson (2009) when interviewing 8 high school statistics teachers, who seemed not to understand the purpose of statistical tests as mechanisms to carry out statistical inferences.

3.4. Understanding statistical significance and p-values

Two particularly misunderstood concepts are the significance level and the *p*-value. The significance level is defined as the probability of falsely rejecting a null hypothesis. The *p*-value is defined as the probability of observing the empirical value of the statistics or a more extreme value, given that the null hypothesis is true. The most common misinterpretation of these concepts consists of switching the two terms in the conditional probability: interpreting the level of significance as the probability that the null hypothesis is true once the decision has been made to reject it or interpreting the *p*-value as the probability that the null hypothesis is true, given the observed data. For example, Birnbaum (1982) reported that his students found the following definition reasonable: "A level of significance of 5% means that, on average, 5 out of every 100 times we reject the null hypothesis, we will be wrong". Falk (1986) found that most of her students believed that α was the probability of being wrong when rejecting the null hypothesis at a significance level α. Similar results were found by Krauss and Wassner (2002) in university lecturers involved in the teaching of research methods. More specifically they found that 4 out of every 5 *methodology instructors* have misconceptions about the concept of significance, just like their students. Vallecillos (1994) carried out extensive research on students misconceptions related to statistical tests (n=436 students from different backgrounds) that included 31 pre-service mathematics teachers (students graduating in mathematics), 13 of whom interpreted the level of significance as the probability that the null hypothesis is true, once the decision to reject it has been made.

Lui and Thompson (2009) remark that the ideas of probability and unusualness are central to the logic of hypothesis testing, where one rejects a null hypothesis when a sample from this population is judged to be sufficiently unusual in light of the null hypothesis. However, they found that teachers “conceptions of probability (or unusualness) were not grounded in a conception of distribution and thus did not support thinking about distributions of sample statistics and the fraction of the time that a statistic’s value is in a particular range (p. 16).” While a single random sample is a critical part of statistical inference, probably more important is an appreciation of the "could-have-been" – all the other random samples that could have been drawn but were not. “Sampling has not been characterized in the literature as a scheme of interrelated ideas entailing repeated random selection, variability, and distribution.” (Saldahna & Thompson, 2002, p. 258).

3.5. Understanding confidence intervals

Fiddler and Cumming (2005) asked a sample of 55 undergraduates and postgraduate science students to interpret statistically non-significant results and gave the results in two different ways (first as *p* values and then as confidence intervals or vice versa). Students were asked to indicate whether the results provided support for the null hypothesis (considered as a misconception), provided support against the null hypothesis, or neither. The authors found that students misinterpreted *p*-values twice as often as they mis-interpreted confidence intervals. There was also evidence that students who were given the confidence interval results first gave the correct answer on the *p* value presentation more often than students who were given the *p* value results first. The author concluded there are benefits of teaching inference via confidence intervals rather than hypothesis tests.

Cumminget al. (2004) reported an internet study in which researchers were given results from an experiment (simulated in an applet) and were asked to show where they thought the 10 means from 10 'new' samples could plausibly fall. The results suggested that a majority of the researchers held a misconception that a r% confidence interval will, on average, capture r% of the means of the 'new' samples.

4. IMPLICATIONS FOR TEACHING AND RESEARCH

Castro-Sotos (2009) reported slightly lower percentages of students with certain misconceptions related to hypothesis testing when compared to similar studies from years before. The author suggests that innovation in statistics education in the last decade may be resulting in some level of improved understanding of statistical inference. While this is merely conjecture, it highlights the idea that students must develop an understanding of many challenging probabilistic and statistical concepts and the relationships between them before meeting statistical inference. Given the difficulty learners have integrating the concepts involved in statistical inference, it makes sense that the underpinning ideas need to be developed over years, not weeks.

4.1. Inference-friendly views of a sample

Statistical inference is applied to a wide variety of situations. However, understanding why it can be validly applied to one situation does not mean learners will understand why it can (or cannot) be validly applied to another, e.g. a situation involving the mean of a finite population compared to a situation involving measurement error (where a population does not exist, but a true value of the measurement does). Students need to hold multiple views of a sample, appreciating the source(s) of the variability that give rise to the samples characteristics, to deeply understand statistical inference and its many applications. Context is clearly critical in supporting a student to develop different views of a sample. Konold and Lehrer (2008) discuss three contexts from which samples are produced: measurement error, manufacturing processes and natural variation.

A critical view of a sample is as the result of a *target-error* process, which aims to consistently produce a single value but fails due to the unavoidable variation in the process (e.g. the machine process that aims to cut fruit bars to be exactly 7 cm long). This can be referred to as the *target-error-view* of sample. Opportunities to develop this view are rarely, if ever, provided at a school level. Natural variation contexts (e.g. the weight of all female quokkas on Rottnest Island) are the most common contexts students meet at school but do not help in developing this critical view of a sample.

Students also need opportunities, over a period of years, to develop a *view* of a sample as a single instantiation of the random sampling process from a population and to develop the appreciation that each possible random sample carries with it an associated level of unusualness (the probability of being drawn). This is referred to as the *population-view* of a sample. While this is the most common view, and current school curricula attempt to develop this using contexts associated with natural variation, it is possible that the target-error-view of a sample should be developed prior the population-sample view. Konold, Harradine, and Kazak (2007) describe activities in which middle school students build *data factories* with the aim of assisting in the development of the target-error-view. Their approach also develops the notion that data result from chance based processes and as such make explicit the relationship between data and chance; a relationship critical to understanding statistical inference and that has been lost (or was never present) in many current school curricula (Konold & Kazak, 2007). Without such views of sample, it is difficult to develop a deep understanding of, and validly apply, statistical inference.

4.2. Developing an understanding of the population-view of a sample

Many interactive applets are now available that provide dynamic, visual environments within which students can engage in the construction of sampling distributions. Chance et al. (2004) reported on a series of studies that investigated the impact that interacting with such applets had on students' understanding when learning about sampling distributions. In the first studies, students tended to look for rules when answering test items and did not understand the underlying relationships that caused the visible patterns they noticed as a result of using the applets. In later studies, the authors asked the students to make predictions about sampling distributions of means before using the applets to validate their predictions. This strategy proved to be useful in improving the students' reasoning about sampling distributions.

4.3. Alternative ways to introduce statistical inference

Most students' first introduction to statistical inference is via a first course in classical statistical inference. In recent years the literature has included thinking about what is termed *informal inference*. While informal inference, as a concept, is not yet universally agreed upon, a consistent feature of informal inference is that suggested activities engage students in the reasoning process of statistical inference without relying on probability distributions and formulas.

Some see informal inference as the collection of the fundamental ideas that underpin the understanding of classical statistical inference. These fundamentals include discriminating between signal and noise in aggregates, understanding sources of variability, recognizing the effect of sample size, and being able to identify tendencies and sources of bias (Rubin, Hammerman, & Konold, 2006). Other views of informal inference include (Zieffler, Garfield, Delmas, & Reading, 2008): (a) reasoning about possible characteristics of a population from a sample of data, (b) reasoning about possible differences between two populations from observed differences between two samples of data and, (c) reasoning about whether or not a particular sample statistic is likely or unlikely given a particular expectation about the population.

Cobb (2007) proposes teaching the logic of inference with randomisation tests rather than using normal distributions as approximate models for sampling distributions, noting that such an approach is what Ronald Aylmer Fisher advocated, but which was not realistic in his day due to the absence of computers. Rossman (2008) claims that teachers could use randomisation tests to connect the randomness that students perceive in the process of collecting data to the inference to be drawn. He provides examples of how such a randomization-based approach might be implemented, while Scheaffer and Tabor (2007) propose such an approach for the secondary curriculum and provide relevant examples.

4.4. Teacher knowledge

Research results summarised in this chapter primarily concern students' misconceptions and difficulties in learning about statistical inference. The little research available about teachers' understanding of statistical inference (Vallecillos, 1994; 1999; Krauss & Wassner, 2002; Lui & Thompson, 2009) indicates it is possible that some teachers share the same misconceptions as the students. In addition, teachers who have not studied statistical inference prior to having to teach it are likely to have the same difficulties in learning the concepts as students do. If this is the case and the situation is not addressed, then it is unlikely that widespread improvement in student understanding will be seen any time soon.

4.5. Some research priorities

The valid application of statistical inference is of critical importance in a broad range of human endeavours. Areas in which research attention is needed include:

- The creation and critical evaluation of a curriculum that systematically develops the key ideas that underpin statistical inference across a number of years in the middle and high school years, so a proper foundation is laid for the formal instruction of statistical inference.
- The study of the current level of understanding and professional knowledge, both at a school and university level, of those teachers charged with teaching statistical inference.
- The critical evaluation of the use of alternative methods (e.g. randomisation tests) when first introducing statistical inference. Great care should be taken in this area given the widespread and long-term use of classical statistical inference.

REFERENCES

Batanero, C. (2000). Controversies around significance tests. *Mathematical Thinking and Learning, 2*(1-2), 75-98.

Batanero, C., Godino, J. D., Vallecillos, A., Green, D. R., & Holmes, P. (1994). Errors and difficulties in understanding elementary statistical concepts. *International Journal of Mathematics Education in Science and Technology*, *25* (4), 527–547.

Birnbaum, I. (1982). Interpreting statistical significance. *Teaching Statistics, 4*, 24–27.

Castro-Sotos, A. E. (2009). How confident are students in their misconceptions about hypothesis tests? *Journal of Statistics Education 17* (2). Online: www.amstat.org/publications/jse/.

Castro-Sotos, A. E., Vanhoof, S., Noortgate, W. & Onghena, P. (2007). Students' misconceptions of statistical inference: A review of the empirical evidence from research on statistics education. *Educational Research Review*, *2*, 98–113

Chance, B., delMas, R. C., & Garfield, J. (2004). Reasoning about sampling distributions. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 295-323). Amsterdam: Kluwer.

Chow, L. S. (1996*). Statistical significance: Rationale, validity and utility*. London: Sage.

Cobb, G. (2007). The introductory statistics course: A Ptolemaic curriculum? *Technology Innovations in Statistics Education*, *1*(1). Online: repositories.cdlib.org/uclastat/cts/tise/.

Cumming, G., Williams, J., & Fidler, F. (2004). Replication, and researchers' understanding of confidence intervals and standard error bars. *Understanding Statistics, 3*, 299-311.

Falk, R. (1986) Misconceptions of statistical significance, *Journal of Structural Learning, 9*, 83-96.

Falk, R., & Greenbaum, C. W. (1995) Significance tests die hard: The amazing persistence of a probabilistic misconception*, Theory and Psychology, 5* (1), 75-98.

Fidler, F., & Cumming, G. (2005). Teaching confidence intervals: Problems and potential solutions. *Proceedings of the International Statistical Institute 55th Session.* Sydney, Australia: International Statistical Institute. Online: www.stat.auckland.ac.nz/~iase/publications.

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., & Scheaffer, R. (2005). *Guidelines for assessment and instruction in statistics education (GAISE)* *report: a preK-12 curriculum framework.* Alexandria, VA: American Statistical Association. Online: www.amstat.org/Education/gaise/.

Garfield, J. B. (2002) The challenge of developing statistical reasoning. *Journal of Statistics Education, 10* (3). Online: http://www.amstat.org/publications/jse/.

Garfield, J., & Gal, I. (1999), Teaching and assessing statistical reasoning. In L. Stiff (Ed.), Developing mathematical reasoning in grades K-12 (pp. 207-219). Reston, VA: National Council Teachers of Mathematics.

Harlow, L. L., Mulaik, S. A., & Steiger, J. H. (1997). *What if there were no significance tests?* Mahwah, NJ: Lawrence Erlbaum Associates.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases.* New York: Cambridge University Press.

Konold, C., Harradine, A., & Kazak, S. (2007). Understanding distributions by modeling them. *International Journal of Computers for Mathematical Learning, 12* (3), 217-230.

Konold, C., & Lehrer, R. (2008). Technology and mathematics education: An essay in honor of Jim Kaput. In L. D. English (Ed.), *Handbook of international research in mathematics education* (2nd ed.) (pp. 49–71). New York: Routledge.

Konold, C., & Kazak, S. (2008). Reconnecting data and chance. *Technology Innovations in Statistics Education, 2*(1). Online: repositories.cdlib.org/uclastat/cts/tise/.

Krauss, S., & Wassner, C. (2002). How significance tests should be presented to avoid the typical misinterpretations. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics.* Cape Town: International Statistical Institute and International Association for Statistical Education. Online: www.stat.auckland.ac.nz/~iase/publications.

Liu, Y., & Thompson, P. W. (2009). Mathematics teachers' understandings of proto-hypothesis testing. *Pedagogies, 4* (2), 126-138.

Ministry of Education and Sciences (2007). *Real Decreto 1467/2007, de 2 de noviembre, por el que se establece la estructura del bachillerato y se fijan sus enseñanzas mínimas* (Royal Decree establishing the structure of high school curriculum).

Ministry of Education, (2007). *The New Zealand Curriculum*. Wellington, New Zealand: Learning Media Limited.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics.* Reston, VA: Author.

Rossman, A. (2008). Reasoning about informal statistical inference: One statistician's view. *Statistics Education Research Journal*, *7* (2), 5-19. Online: www.stat.auckland.ac.nz/serj/.

Rubin, A., Hammerman, J. K. L., & Konold, C. (2006). Exploring informal inference with interactive visualization software. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics.* Cape Town, South Africa: International Association for Statistics Education. Online: www.stat.auckland.ac.nz/~iase/publications.

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

Saldanha. L., & Thompson, P. (2007) Exploring connections between sampling distributions and statistical inference: an analysis of students' engagement and thinking in the context of instruction involving repeated sampling. *International Electronic Journal of Mathematics Education, 3,* 270-297.

Scheaffer, R., & Tabor, J. (2008). Statistics in the high school mathematics curriculum: Building sound reasoning under uncertainty. *Mathematics Teacher*, *102* (1), 56-61.

Senior Secondary Board of South Australia (SSABSA), (2002). *Mathematical studies curriculum statement*. Adelaide, Australia: SSABSA.

Vallecillos, A. (1994). *Estudio teórico-experimental de errores y concepciones sobre el contraste estadístico de hipótesis en estudiantes universitarios* (Theoretical and experimental study on errors and conceptions about hypothesis testing in university students). Unpublished Ph. D. University of Granada, Spain.

Vallecillos, A. (1999). Some empirical evidence on learning difficulties about testing hypotheses. *Proceedings of the International Statistical Institute 52nd Session.* Helsinki: International Statistical Institute. Online: www.stat.auckland.ac.nz/~iase/publications.

Zieffler, A., Garfield, J. B., delMas, R., & Reading, C. (2008). A framework to support research on informal inferential reasoning. *Statistics Education Research Journal*, *7*(2), 5-19. Online: www.stat.auckland.ac.nz/serj/.