A Model for the Promotion of Statistical Thinking?

Maxine Pfannkuch
The University of Auckland
<m.pfannkuch@auckland.ac.nz>

This paper presents a model for statistical thinking in empirical enquiry, based on practitioner behaviour. The conjectured types of thinking that are fundamental to the statistics discipline are then used to analyse an assessment item and teaching situation. It is concluded that such an analysis is possible and that more discussion and research is needed on defining, clarifying, and articulating the statistical thinking that is embedded and inherent in statistical practice.

Over the last thirty years introductory statistics teaching has been moving from mathematical statistics towards empirical statistics. At the same time significant shifts have been occurring in statistical practice with the increasing availability and power of computer technology and with many applied statisticians broadening their domain of practice to all parts of the problem solving empirical enquiry cycle. These shifts in practice are refocusing the emphasis in teaching from *how to do*, for example, a graph, mean or confidence interval, to *how to think about* them. However, the development of students' statistical thinking can only be promoted if we know the type of thinking that is characteristic of the statistics discipline. Many articles discussing and characterising the core elements of statistical thinking have been written (e.g., Moore, 1990; Biehler, 1994).

The Wild and Pfannkuch (1999) four–dimensional model (Fig. 1) was an attempt to characterise practitioner thinking in the statistics discipline. The model was developed as a result of: interviewing statisticians and tertiary students about statistical projects they had been involved in; interviewing tertiary students as they performed statistical tasks; and analysing the literature. The focus of the statistical thinking was at the broad level of the statistical enquiry cycle, ranging from problem formulation to the communication of conclusions. The multidimensional nature of statistical thinking is captured in the model. Thus a thinker operates in all four dimensions at once. For example the thinker could be categorised as being currently in the analysis stage of the Investigative Cycle (Dimension 1), dealing with some aspect of variation in Dimension 2 (Types of Thinking) by criticising a proposed analysis against contextual knowledge of the situation in Dimension 3 (Interrogative Cycle) driven by scepticism in Dimension 4 (Dispositions). This paper focuses on the thinking that we believe is fundamental to statistics and how we might use these defined characteristics to analyse assessment items and teaching situations for the promotion of statistical thinking.

Fundamental Statistical Thinking

From our research five types of thinking that are inherently statistical emerged (see Fig.1(b)). We believe that these are the fundamental elements of statistical thinking.

Recognition of the need for data. The foundations of statistical enquiry rest on the assumption that many real situations cannot be judged without the gathering and analysis of properly collected data. Anecdotal evidence or one's own experience may be unreliable and misleading for judgements and decision-making. Therefore data are considered a prime requirement for judgements about real situations.

Transnumeration. In order to make a judgement in a real situation, data that 'measure' or capture the qualities or characteristics of that situation must be found. For this type of thinking we coined the word transnumeration which means "numeracy transformation to facilitate understanding". Once the data have been collected transnumeration thinking operates again as raw data are transformed into multiple graphical representations, statistical summaries, and so forth, in a search to obtain meaning from the data. At the final stage transnumeration thinking occurs when the meaning from the data, the judgement, has to be communicated in a form that can be understood in terms of the original situation.

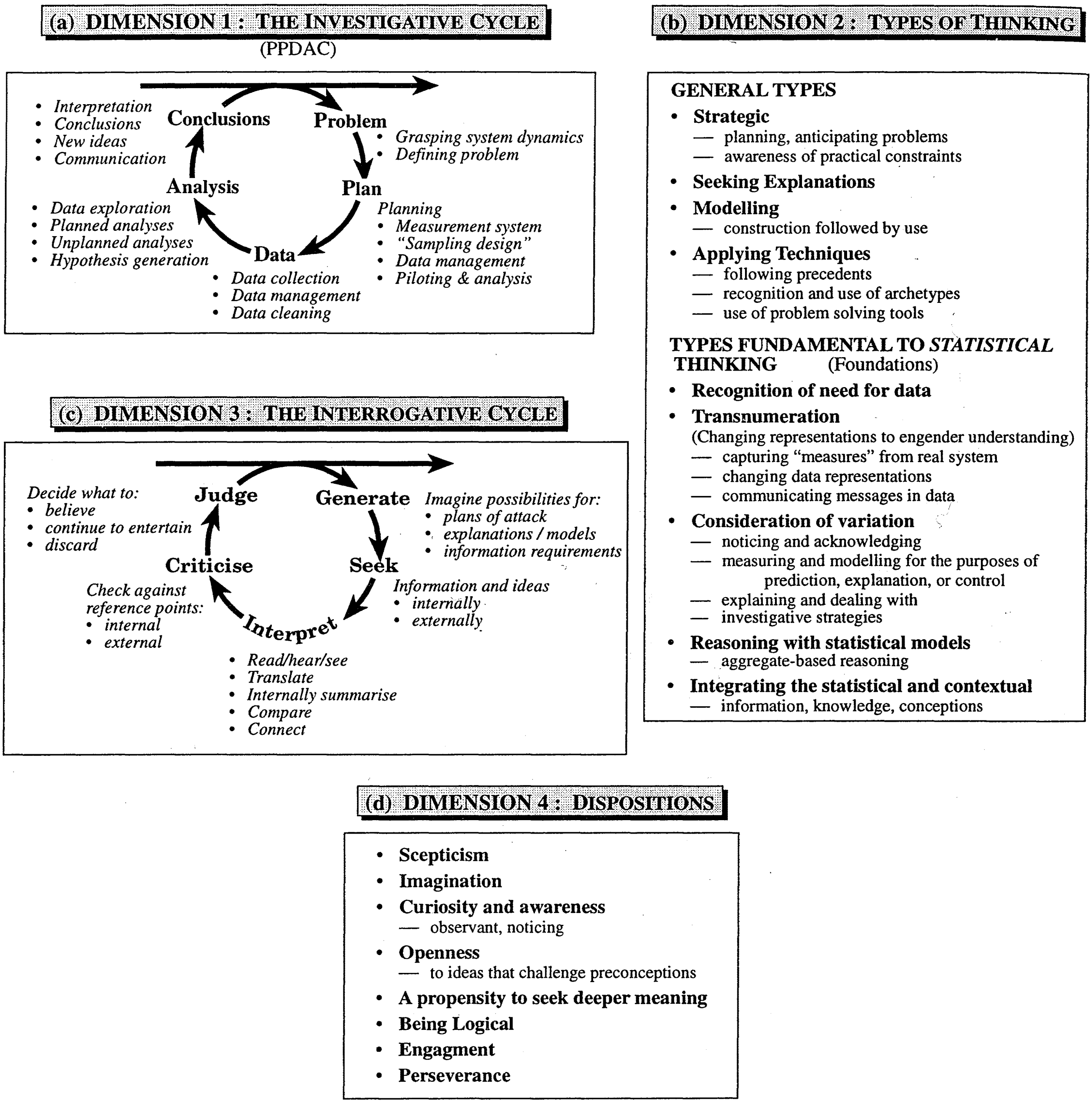


Figure 1. A four-dimensional model for statistical thinking in empirical enquiry

Consideration of variation. Making a judgement from data requires an understanding of variation during the process of statistical enquiry. It is a type of thinking that starts from noticing variation in a real situation, and then influences the strategies we adopt in the design and data management stages through determining whether we will ignore, plan for or control variation. It further continues in the analysis and conclusion stages through determining how we act in the presence of variation. If we accept that statistics is about making predictions, seeking explanations, and finding causes, then we will be looking for variation in data. It affects all thinking through every stage of the investigative cycle.

Reasoning with statistical models. The predominant statistical models are those developed for analysis of data. These models allow us to summarise data in multiple ways dependent upon the nature of the data. For example, centres, spreads, clusters, outliers, confidence intervals, and p-values are all read, interpreted and reasoned in an attempt to find evidence on which to base a judgement. According to Konold et al.'s (1997) research, students find it difficult, when dealing with data, to make the transition from thinking about and comparing individual cases to thinking about and comparing group propensities. Reasoning with statistical models requires the ability to do both aggregate-based and individual-based reasoning and to recognise the power and limitations of such reasoning across a variety of situations. This aggregate-based reasoning is fundamental to statistical thinking. For example, in mathematics one counter-example disproves a theory whereas in statistics one has a theory about group propensities and one counter-example (an individual case) does not disprove the theory. There is also a need to develop more statistical models for reasoning in the other stages of the investigative cycle. Quality management is one field of statistics that is attempting to address such model development.

Integrating the statistical and contextual. Although the above types of thinking are linked to contextual knowledge, the integration of statistical knowledge and contextual knowledge is an identifiable fundamental element. The statistical model must capture elements of the real situation and thus the resultant data will carry their own literature base (Cobb & Moore, 1997). Information about the real situation is contained in the statistical summaries and therefore a synthesis of statistical and contextual knowledge must operate to draw out what can be learnt in the context sphere. This synthesis operates throughout the investigative cycle and enables some judgements to be eventually formed on the real situation.

Thus the question is raised as to whether this theoretical model will be useful in analysing assessment items and teaching situations for determining the type of statistical thinking that is being invoked. An exploratory analysis, based only on the fundamental types of statistical thinking, is now presented for each situation.

An Assessment Situation

Tertiary students in an introductory statistics course were given, as part of their assessment, a problem (Fig. 2) involving the construction and comparison of graphs. After the assessment they were given model solutions (Department of Statistics, 1999).

Analysis in terms of promoting statistical thinking

Recognition of the need for data. The data gathered and the questions asked do not promote this type of thinking, but illustrate its need, especially as the results are counter-intuitive.

Transnumeration. In this assessment problem this type of thinking is not done by the students, since the "measurement problem" has already been "solved" and the students are asked to construct particular plots. Comments are also not sought on which of the two plots give a better representation for communication about the real situation.

Consideration of variation. The first three parts of the problem deal with possible sources of variation in the design of the experiment such as gender, observer variability and researcher bias. This consideration of variation reinforces the thinking that, for the two groups of children, the effects of differences among the children and among the observers should be taken into account in the design of the experiment, so that the differences observed can only be attributed to the fact of whether the children were rewarded or not initially. The questions, however, do not get the students to think of other possible sources of variation in the design of this experiment. In the solutions to parts (d), (e) (not shown here), and (f), the interquartile range, the medians, and so forth are calculated (measured) and left. Comments on the plots and summaries, with regards to variation or patterns in the variability for the purposes of prediction and explanation about time spent on drawing, are not sought. Also, in the part (g) question, variation ideas are not triggered for the making of an inference, from the sample, about the population from which it is drawn.

Problem:

A study (Lepper, Greene & Nisbett, 1973, cited in Lockhart, 1998, p. 4) was conducted to investigate motivation in children's behaviour. Thirty-eight 5-year old children, all who had shown at least a moderate interest in drawing as a voluntary activity, were divided into two groups. All the children were asked to make some drawings and were given 6 minutes to do so. Beforehand, the children in one group were told that they would receive a "Good Player Certificate" as a reward for their participation. The second group were told nothing about the reward and received none. One week later, all the children took part in a one hour free play session. Drawing was one of the activities they could choose. Observers measured the amount of time (in minutes) each child spent drawing during this time. The resulting data are presented below:

With Reward: 5.2, 8.5, 1.9, 6.5, 8.8, 5.5, 2.3, 3.7, 16.0, 7.9, 1.9, 6.0, 5.2, 7.2, 0.0, 7.8, 2.4, 4.7, 7.8

No Reward: 8.4, 10.8, 8.2, 9.6, 2.5, 7.5, 12.6, 10.8, 13.8, 11.3, 8.9, 9.8, 10.1, 10.6, 11.3, 9.7, 9.4, 10.5, 11.5

(a) The 38 children used in the experiment were divided into the two groups randomly. What was the purpose of doing this at random?

(b) If the experimenters were worried that the gender of the child may influence the way they behave, how should the design of the experiment be modified?

(c) Can any form of blinding be used in this experiment? If so, briefly describe how.

(d) Construct a back-to-back stem-and-leaf plot of the two sets of data using an appropriate scale.

(e) Calculate the five-number summary for the With Reward data. Show your working.

(f) The five-number summary for the No Reward data is: (2.5, 8.9, 10.1, 11.3, 13.8). Using this and your five-number summary from (b) draw side-by-side boxplots for these two sets of data. Show your working.

(g) Using your plots, in plain English, compare the two sets of data.

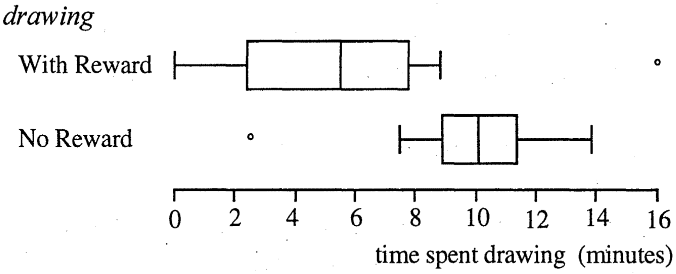
Some Solutions:

(a) The purpose of randomly allocating the children to the two groups was to try and ensure that the comparisons are fair in the sense that two groups are as similar as possible in every way except for whether or not they were rewarded initially.

(b) The experimenter should have blocked the children by gender.

(c) The only form of blinding possible would be for the people recording the observations to have been blinded.

(f) Box plot of time spent drawing



(g) The data for With Reward group is centred much lower and is more spread out than the data for No Reward group. The With Reward group data looks mildly left (negative) skewed. The No Reward data looks slightly right (positive) skewed. There is a possible low outlier in the No Reward group and a possible high outlier in the With Reward group.

Figure 2. Assessment item for introductory statistics course.

Reasoning with statistical models. Part (g) in the solutions is an example of how we would expect students to reason with these models. Skewness and outliers are noticed. The difference in the centres and spreads between the two groups is acknowledged. However, in order to draw out the meaning, in terms of the original situation, from these statistical models it is necessary to use context knowledge.

Integrating the statistical and contextual. If we believe that we undertake statistical investigations to learn more in the context sphere then the questions for this problem take no cognisance of the context of the problem. There is no request to interpret and communicate the information contained in the data about motivation in children's behaviour. There was sufficient context given at the beginning of the problem for students to communicate the information in the data contextually.

Discussion

This problem potentially provides a good base for assessing students' knowledge. It operates in the plan and analysis phases of the investigative cycle. Whilst I recognise that students need directed experiences such as the above item they also need undirected experiences in data-handling. With more consideration given to developing students' statistical thinking I believe that we could rethink the questions we ask and the model solutions we provide. For instance, the recognition of the need for data might be prompted with a question such as: *"Before looking at the data, read the details about the experiment. Use your own knowledge to predict what the results might demonstrate and then state the basis for your prediction. Think of an alternative prediction and justification."* Transnumeration thinking and the other types of thinking could be promoted by an open question such as: "*compare the two groups explaining what you learn from the comparison/s*". Another question that may be useful for triggering thinking on measurement issues is: "*write down some 'worry' questions you might have about defining the problem for this experiment and explain why"*. A question that could be asked to promote thinking about variation could be: "*write down some 'worry' questions you might have about the design of this experiment and explain why*". Questions such as: "*summarise what you learnt from this experiment*" followed up with: "*what would you investigate next if you were the researcher*?" might encourage the integration of statistical and contextual knowledge.

A Teaching Situation

This teaching situation (Shaughnessy, 1999) was observed by me in an undergraduate and a graduate class. Data sets for the time in minutes between blasts of the geyser Old Faithful are cut up into sixteen separate days. Each student is given a strip of data from *one* day of Old Faithful's activity. The students are asked: (1) to construct a graph of the data; (2) to make a prediction about what they think another day of this data would look like; and (3) to swap days with someone else, and repeat steps 1 and 2. They are also asked how long they would expect to wait between blasts of Old Faithful, and to explain their reasoning. As the activity progresses further questions are asked.

Analysis in Terms of Promoting Statistical Thinking

Recognition of the need for data. The giving out of one day's data and seeing that a neighbour's data look different promotes ideas about variability within days and from day to day, and about the need for more data to be able to understand the dynamics of this particular geyser system before making any informed judgement.

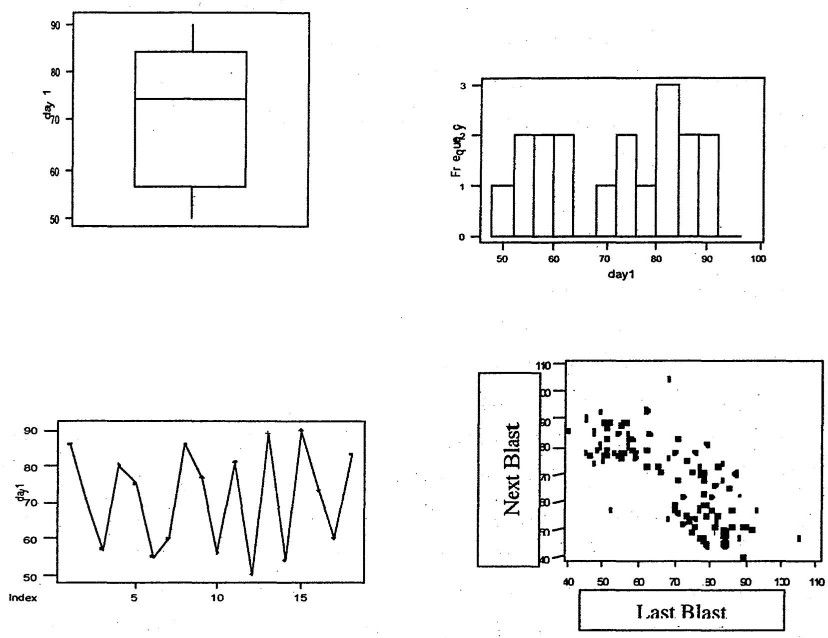


Figure 3. Minutes between blasts – Some possible graphs for teaching situation

Transnumeration. Students start with a strip of one day's data and they need to think how they must change the data representation to facilitate a prediction. The transnumeration thinking is partially done by the students as they are asked to draw a graph but they are not told what to draw. Most students draw a single graph. Students are then asked to share their graphs with the class (see Fig. 3). It soon becomes apparent with data from sixteen days that the dot plots, stem-and-leaf plots, time series line graphs, box-and-whisker plots, the histograms with small class intervals and with large class intervals, all reveal something different about the activity of Old Faithful. The histogram with large class intervals obscures information and hence the students must recognise this and try several different class interval widths. This sharing of graphs promotes the transnumeration type of thinking and the need to look at multiple representations.

Consideration of variation. The box-and-whisker plots reveal the considerable variation within a day and between days even though sometimes the median is approximately the same. The dot plots, histograms with small class intervals, the stem-and-leaf plots and the line plots reveal the bimodal nature within a day's data for most of the days (though not all) and thus the linking of these plots with the boxplots accounts for the observed variability. The students report that the mean time of 70 minutes is how long they expect to wait before the next blast. However after much discussion and looking at the line plot with its oscillating variability and the bimodal histograms gives them the idea that if the last blast waiting time was around 60 minutes then they would expect to wait about 80 minutes for the next blast while if the last blast waiting time was about 80 minutes they would expect to wait about 60 minutes for the next one. Thus the students are beginning to model and measure variation for the purpose of prediction. The question then arises as to whether this prediction is valid. Students must think of a way of representing the data (transnumeration type thinking) in order to validate the prediction. A scatterplot of last blast versus next blast in minutes (Fig. 3) reveals that there is more variability in the next blast prediction when the last blast is greater than 70 minutes than when the last blast is less than 70 minutes.

Reasoning with statistical models. The multiple representations of these data (Fig.3) require the students to be able to read and interpret the graphs and also interrelate the graphs. The oscillating nature of the line plots requires the students to reason that a short blast is often followed by a long blast, that there is an underlying bimodal distribution, that the activity is not random, and the data have an underlying model from which predictions could be made about another day's data. The histograms require the students to reason that there may be an underlying bimodal distribution and there are possibly two distinct clusters of data that are reasonably symmetric with their own underlying means and standard deviations. The boxplots require students to reason comparatively among the days, to reason with group propensities, and to notice that these are summary plots and hence may be obscuring features of the distribution of the data. For predicting the next blast of Old Faithful, the students reasoned that the line plots were the most appropriate graphs for giving measurable predictions which were conditional upon knowing the wait time of the previous blast. Reasoning with the centre, which many students did, is inappropriate for predicting the next blast. Above all, students need to read, interpret, reason, and make sense of the *connections* and *interrelationships* among all the graph representations.

Integrating the statistical and contextual. The patterns seen in the variability stimulates questions about how this geyser works. The data is telling these students about properties of the eruption of this geyser. At this stage more scientific information is needed on geysers so that the statistical knowledge gained can be integrated with the contextual knowledge about geysers. It may be that this type of pattern is repeated in other geysers and that there are good reasons, concerning the dynamics of geysers, for this pattern to occur. Other questions naturally arise as to what other data are needed, such as the duration time of the blasts, to further understand and to be able to predict the waiting time between blasts.

Discussion

This rich teaching situation actively develops and promotes statistical thinking according to the Wild and Pfannkuch paradigm. It could be changed in three ways, although from a teaching perspective I would not recommend these changes for this particular activity. The following changes are only mentioned to highlight that students need some other experiences with data to further promote statistical thinking. Initially, before the data-sets are given out, the students could be asked to predict the minutes between a series of blasts and the bases for their predictions. This may encourage the notion of the need for data. Secondly, each student could be given the same data-set and then it would be up to the students to suggest that more data is needed. Thirdly, there should be no suggestion that they construct a graph. Through not asking them to plot the data the transnumeration type of thinking could be promoted from the beginning of the activity. Thus only the two prediction questions need to be asked. Or taking another step, the students could be given the data for one or several days and asked what sort of questions (i.e., "their notices and wonders" (Shaughnessy, 1997)) they could investigate with these data. This would promote thinking of the variation type and the use of statistical and contextual knowledge.

Conclusion

The assessment situation is an illustration of a structured task whereas the teaching situation illustrates an unstructured task. Rich statistical experiences can be given to students by choosing tasks that portray how statistical thinking is applied to particular statistical situations. Students need to practise that new knowledge and the way of thinking in structured questions. However, if students are to independently model statistical practice and thinking they must be presented with unstructured tasks during teaching and assessment. This will enable students to create for themselves the linkages and connections between what they have learnt and the statistical situation presented. This would suggest that in every assignment and its equivalent teaching unit there should be at least one unstructured task that promotes statistical thinking.

The Wild and Pfannkuch model was an attempt to make implicit practitioner thinking in statistics more explicit. The above analysis was an exploration of whether we could understand our teaching and assessment from a statistical thinking perspective. The analysis demonstrates how this model may be used for assessing the presence of the broad fundamental statistical thinking skills. The model may also identify improvements for teaching and assessing, and clarify why tasks are particularly good for developing thinking. Most introductory tertiary statistical courses have commonalities for the content that should be covered. However, teaching such courses requires an understanding of how to develop the content and thinking that exists in statistical practice. More discussion is needed in the statistical community to develop commonalities for understanding, communicating, articulating, and teaching statistical thought. The new curricula in the future should interconnect, align, and focus on both content and thinking as the statistics discipline evolves with new technology and new statistical practice.

References

Biehler, R. (1994). Probabilistic thinking, statistical reasoning and the search for causes: Do we need a probabilistic revolution after we have taught data analysis? In J. Garfield (Ed.), *Research papers from the fourth International Conference On Teaching Statistics, Marrakech 1994.* Minneapolis, MN: University of Minnesota.

Cobb, G., & Moore, D. (1997). Mathematics, statistics and teaching. *The American Mathematical Monthly, 104*(9), 801-823.

Department of Statistics (1999). Introduction to statistics: Assignment 1, Semester 2, 1999. (Available from Department of Statistics, The University of Auckland, Private Bag 92019, Auckland, New Zealand)

Konold, C., Pollatsek, A., Well, A., & Gagnon, A. (1997). Students analysing data: Research of critical barriers. In J. Garfield & G. Burrill (Eds.), *Research on the role of technology in teaching and learning statistics* (Proceedings of the 1996 International Association of Statistics Education round table conference, pp. 151-167) Voorburg, The Netherlands: International Statistical Institute.

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

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* (Proceedings of the 20th annual conference of the Mathematics Education Research Group of Australasia, pp. 6-22). Rotorua, NZ: MERGA.

Shaughnessy, J. M. (1999). Using data to make predictions about Old Faithful. (Available from Professor J. M. Shaughnessy, Department of Mathematical Sciences, Portland State University, Portland, OR 97207)

Wild, C., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review, 67*(3), 223-265.