THINKING TOOLS AND VARIATION

MAXINE PFANNKUCH
The University of Auckland, New Zealand
m.pfannkuch@auckland.ac.nz

Editor's note: This article is a discussion of and reaction to five papers which appeared in the SERJ special issue 3(2) published in November 2004, on Research on Reasoning about Variation and Variability. These include four refereed papers reporting original research, by Hammerman and Rubin, Ben-Zvi, Bakker, and Reading, and Gould's invited paper providing a statistician's view of variation and its analysis. For brevity, these papers will be referenced here only by author names, without the year of publication.

SUMMARY

This article discusses five papers focused on "Research on Reasoning about Variation and Variability", by Hammerman and Rubin, Ben-Zvi, Bakker, Reading, and Gould, which appeared in a special issue of the Statistics Education Research Journal (No. 3(2) November 2004). Three issues emerged from these papers. First, there is a link between the types of tools that students use and the type of reasoning about variation that is observed. Second, students' reasoning about variation is interconnected to all parts of the statistical investigation cycle. Third, learning to reason about variation with tools and to understand phenomena are two elements that should be reflected in teaching. The discussion points to the need to expand instruction to include both exploratory data analysis and classical inference approaches and points to directions for future research.

Keywords: Statistics education; Variation; Reasoning; Thinking tools; Statistical investigation; Exploratory data analysis

1. INTRODUCTION

The five papers in the SERJ November 2004 special issue on reasoning about variation and variability not only open a window into an exciting new research area but also give an excellent insight into current research on developing cognitive and pedagogical theories. The opportunity to read how statisticians, teachers, and students reason about variation proved to be interesting and thought–provoking. Each of the papers gives its own unique perspective into and interpretation of reasoning about variation, whereas when considering the five papers as a whole the research on variation reasoning becomes more than the sum of all these findings. It then becomes a challenge to explicate and to extract some themes and issues about variation from all the papers.

Variation lies at the heart of statistical enquiry and is one of the "big ideas" that should underpin statistical teaching (Moore, 1990; Gal & Garfield, 1997). Consideration of the effects of variation influences all thinking and reasoning through every stage of the empirical enquiry cycle – problem, plan, data, analysis, and conclusion (Pfannkuch & Wild, 2004). Thinking about variation starts from noticing variation in a real situation, the problem. It influences the strategies that are adopted in the planning and data collection stages, for example by introducing planned variation into data production through the use of randomization or reducing known sources of variability. In the analysis and

Statistics Education Research Journal, 4(1), 83-91, http://www.stat.auckland.ac.nz/serj

conclusion stages the presence of variation determines how perceived patterns might be characterized and acted upon.

Research on how students might develop skills to reason about variation and variability throughout the empirical enquiry cycle has been limited (Shaughnessy, 1997). Further, although variation is a core concept in statistics, teaching has not focused on enculturating students into variation-type thinking and reasoning approaches. Much needs to be learnt about how to cultivate and develop this type of reasoning in students. The November 2004 of SERJ will be seen as a landmark special issue in establishing some beginning foundations for such a knowledge base, and is followed by additional papers on the same topic in the current issue.

With the juxtaposition of a statistician's (Gould's) view of how he deals with and reasons about variation, and the views of statistics education researchers of how students and teachers reason about variation (Bakker; Ben-Zvi; Hammerman & Rubin; Reading), many interesting avenues for enquiry are suggested. One noticeable feature of the five papers is the different thinking tools used by the statistician and students and teachers for the analysis of the data, such as graphs, tables of data, or diagrams. I view these as "thinking tools" because they are more than mere external representations; for example, in the analysis stage they allow the statistician and students to interrogate the data, to extract information contained in the data, and to think and reason with the data. Another noticeable feature is that the focus is on the variation in the *analysis* stage of the empirical enquiry cycle. Ideally teaching and learning data analysis should always be situated within the full empirical enquiry cycle. While specific grade levels may focus on developing particular statistical concepts, the overall cycle of enquiry should always be part of the teaching approach, otherwise students' thinking and reasoning can become disconnected (Pfannkuch & Wild, 2003). These two noticeable features raise two main issues. First, about the relationship between tools and the types of reasoning that they encourage or permit and second, about how reasoning about variation is interconnected to all parts of the enquiry cycle.

What strikes me most, however, from the juxtaposition of the statistician paper and education research papers, is the contrast in the purposes of investigating data. The nature of the analyses that the students are doing seems to me to be leading towards the future enculturation of students into classical inference. In comparison, the statistician's focus, in the statistical case studies described, is more about modeling variation to arrive at insights about underlying data-generated mechanisms. The thinking is less structured, more wide-ranging, and draws on many knowledge sources. Gould's focus is on modeling variation for the purposes of explanation (Lead-Level case study) and prediction (UCLA Rain, Drinking Trends and Chipmunk case studies). This Exploratory Data Analysis (EDA) approach (cf., Tukey, 1977; Shaughnessy, Garfield, & Greer, 1996), whose main aim is modeling and understanding phenomena, is in stark contrast to the simple estimation of parameters of populations from samples, which underscores most of current statistics education.

In light of the above, three main issues are addressed in this paper: the thinking tools used for displaying variation for analysis and the type of reasoning observed; the need to consider how reasoning about variation involves the whole empirical enquiry cycle; and the need to extend the statistics curriculum to include analytic or non-confirmatory situations. Questions will be raised in response to the papers and possible directions for future research will be explored.

2. THINKING TOOLS IN THE ANALYSIS STAGE

The five papers use a range of thinking tools in the analysis stage of the empirical enquiry cycle, such as tabular graphs, graphs, tables of data, and diagrammatic representations, to aid students' reasoning about variation. This section will examine what each paper presents in this regard, reflect on the findings about students' reasoning, and raise questions about the first main issue – possible interconnections between the tools used and the reasoning observed.

Hammerman and Rubin describe how teachers used a new dynamic data visualization tool (Tinkerplots™) to divide distributions into slices and consequently compared frequencies and percentages within these slices to make inferences. This tool, which I will call *the tabular graph* approach, is an integration of table and graph features. The type of thinking observed was slice-wise comparison across groups, which tended to ignore the distribution as a whole (see also Cobb, McClain, & Gravemeijer, 2003). According to Hammerman and Rubin, teachers were viewing the data from a classifier perspective, through their focus on comparing frequencies of particular attribute values within and between data sets, rather than seeing the data sets as a whole, a pre-cursor stage for aggregate-based reasoning. They argue that this new tool engendered and made visible thinking that had previously lain dormant or invisible, since the teachers' slice-wise comparison reasoning seemed to be an extension of pair-wise comparison reasoning that other researchers have documented (e.g., Moritz, 2000), whereby students compare two individual cases. The tool provided the means to divide the data up, but reasoning about variability also requires detecting patterns in the whole distribution. Is this new integrated tabular graph tool helping or hindering understanding? If the visible frequency count bars or unorganized dots in the bins were concealed by a bar graph in this tool, would the teachers' thinking change? Did this tool produce cognitive overload in the teachers with one of the consequences being subgroup or bin reasoning? Or has this tool opened up completely new ways of reasoning about variation?

Ben-Zvi gives a thorough insight into how two students grappled with variability within and between two distributions. The students built up their inferential reasoning by being introduced to a sequence of tools throughout the learning process, namely frequency table, percentage table, statistical measures, graph. This is a *tables plus graph* approach. Ben-Zvi's approach is to let the students focus and think with one tool at a time. First, with the table, the students were scaffolded to notice the variation and then formulate a hypothesis. To justify their hypothesis they argued with a table of percentages where slice-wise comparisons were made, as well as range comparisons involving variation within and between two sets of data in table format. The introduction of statistical measures seemed to produce few new steps in their thinking. The production of a series comparison graph by the students induced pairwise comparison of the bars, which led to seeing the pattern in the variation and to articulating a trend. Ben-Zvi gets his students to think with the table but other research (e.g., Pfannkuch & Rubick, 2002) shows students do not necessarily do this since the table is perceived as a means to organize data for graphing, not as a tool to think with. Did using the table aid the students' thinking? Was the percentage table a necessary precursor to their thinking with graphs? Did these students integrate the percentage table and graph into their thinking or did the thinking remain separated? How should the teacher integrate statistical measures into their graphical reasoning? Should thinking tools be introduced in a sequential way?

Bakker, who deals with sample distributions and "grows" them to population distributions, has a purely *graphical or diagrammatic reasoning* approach. The students described the variability in predicted sample data distributions, comparing their predictions to real sample data sets. The focus is on transitioning students' informal language and notions to statistical language and notions and on transitioning the students from seeing variability in data plots to seeing variability in continuous distributions. The aim is to stay away from the data and foster the notion and concept of distribution by discussing averages and spread in relation to shape. An important part of using these graphical tools for reasoning is the role of reflection and the role of the teacher in facilitating students' conceptions. However, Bakker is not sure whether students recognized that there was more variability in small samples. What patterns were students conceiving within the variability of small samples? How did they integrate into their reasoning the sample distributions with the population distribution? If the students were given a population distribution and asked to take a sample of size 10 from it what would their predicted sample distribution look like?

Reading reports that her students did not use graphic tools for reasoning but rather used *tables of data*. The students noticed the variation, and looked for and detected patterns in the data. This is a natural human response, since all humans tend to see patterns even if none exist (Wild & Pfannkuch, 1999) and to find causal connections (e.g., Watson, Collis, Callingham, & Moritz, 1995). Reading distinguishes a hierarchy of reasoning from qualitative to quantitative. In the qualitative responses the students noticed and acknowledged the variation, whereas at the quantitative level there were emergent ideas of measuring and quantifying the variation such as reasoning with the minima, maxima, ranges, and notions of deviation. Yet, the teacher did not scaffold Reading's students' reasoning and few students chose to use graphs (with which they were familiar). What were the characteristics of the data table that induced students not to change the data to another representation?

In contrast, Gould, the statistician, uses multiple graphs and models as his thinking tools and keeps re-organizing the data to draw out the information contained in the variation. This is the transnumerative type of thinking promoted by Wild and Pfannkuch (1999). He chips away at the unexplained variation to gain more information about patterns in the data and about underlying factors that may be sources of variation in the realistic situations examined. He has a data detective approach – searching for patterns and trends, causes, and explanations for the variation.

These studies suggest that there is a link between the thinking tools used and the reasoning observed in the students or teachers studied. This raises the question as to what tools or sequence of tools will aid students or teachers to conceive both statistical patterns and deviations from those patterns. Prompting students to use multiple models or graphs, multiple ways of seeing and articulating patterns, would seem to be one solution suggested by this research. We do not know, however, how students or teachers link and interconnect their reasoning across tools or how the tools affect, limit, or empower their thinking.

With the above in mind, it is useful to look more broadly at the notion of a thinking tool. In mathematics education research, there is wide recognition that thinking tools or representations do affect students' thinking. Mesquita (1998) observes how the role and nature of the external representation in geometry affects students' thinking. She identifies external representations as having two roles, descriptive and heuristical. The descriptive representation "illustrates multiple relationships and properties involved in the problem without suggesting solution procedures" (p. 191) whereas the heuristical representation "acts as a support for intuition, suggesting transformations that lead to a solution" (p. 191). In another study, Hegedus and Kaput (2004) argue that their dynamic connective representations for algebra activities accounted for significant gains in student performance, which provides further evidence of a link between students' thinking and the thinking tool employed. Mesquita believes that external representations are relevant elements in cognitive theory. In fact, Thomas (2004) reports that there is no coherent cognitive theory on the role and nature of external representations.

Fischbein (1987, p. 125) states that thinking tools or models are useful for the learner to gain access to an understanding of a concept. For productive reasoning the tool must "constitute an intervening device between the intellectually inaccessible and the intellectually acceptable and manipulable." He warns, however, that the properties inherent in the tool may lead to an imperfect mediator and hence can cause incomplete or incorrect interpretations. In the light of what Fischbein is saying the results from the empirical investigations reviewed here suggest that some statistical thinking tools developed by statisticians may be considered as imperfect thinking mediums and may lead to misinterpretation. One avenue to overcome such misinterpretations is for the teacher and learner to become aware of these intuitive obstacles. Another avenue to pursue is that advocated by Fischbein (1987, p. 191): "to create fundamentally new representations." Therefore statistics educators and statisticians may need to reassess the existing thinking tools in the discipline and perhaps create new thinking tools that are more closely aligned to human reasoning, an approach taken by the Tinkerplots™ software creators Konold and Miller (2004).

3. VARIATION AND THE EMPIRICAL ENQUIRY CYCLE

Statistical investigations are conducted to expand contextual knowledge and to improve understanding of situations, as well as to solve data-based problems. In reporting on statistical investigations, usually the problem and the measures taken are defined, the method of data collection is explained, and the main results are communicated. It is important, however, to ask how much information is provided to learners about the *context* from which data emerged, and what role does contextual information have. Gould, the statistician, briefly gives details of where his four sets of data have come from, how data were collected and how measurements were taken. For the reader this allows extra thinking tools for understanding his data and findings. This raises the issue as to whether lack of such information might inhibit or prevent students' thinking from reaching an optimal level. This type of information feeds into the analysis and conclusions parts of the cycle. In order to demonstrate how seeking explanations for the variation is grounded in the empirical enquiry cycle, questions will be raised below about the sources of the data for each of the statistics education papers. These remarks do not assume that these aspects were neglected in the four classroom studies; rather, they aim to raise the second main issue about considering how students' reasoning about variation is interconnected with the whole enquiry cycle.

Ben-Zvi analysed reactions to a task which asked students to compare the lengths of Israeli and American last names. Information that might help an analysis would be: Where did these names come from? What parts of the respective countries? What grades were the two classes selected? What year were they collected? Bakker states that the teacher used real data sets about the weights of groups of children. Were they eighth graders from another Dutch school? Did the school have the same socio-economic level, same ethnicity as the students' school? Were the weights of one gender or both genders? How accurate were the data collected—self-report or actual weighting with a scale?

Hammerman and Rubin ask their teachers to compare homework hours studied by students at two schools. Before comparing the data it would be helpful to know the following: Were the students the same age? How were the data collected – recall of number of hours for one particular week, diary over one week, or was it students' perception of how much homework they usually did each week? Were the two schools similar in ethnicity, socio-economic level? Were both genders included? Was a random sample taken from each school or were particular classes selected? Reading has students analyse real weather data from the students' town. She has a three-year census of data but each student is given a sample of one month's rainfall and later on – in another teaching segment – one month's temperature data to comment on. Questions that need to be considered are: How were these data measured? Were they averages from a number of weather stations or from one weather station? In what parts of the town were these measurements taken? How is a wet day defined in that area? What is the difference between a "wet" day and a "showery" day and how is this information captured in the rainfall measurement? What range of temperatures is considered best for outdoor activities?

The variation that is considered and dealt with in the first two stages of the empirical cycle, problem and plan, should be part of the information or story that is presented to students. One could argue that there would be too much information for students to grasp, hence there should be a balance between providing enough background information to inform the analysis and enable students to engage with their data, versus providing no information, which may leave teachers wondering why students did not consider certain questions. Hammerman and Rubin notice that the analyses appeared to be context dependent. Their teachers seemed to reach an aggregate perspective with the AIDS data and one wonders whether the teachers were given or had more background information about the data.

All the above issues about the problem and plan stages may well have been discussed with the classes but the statistician included his story in his paper while the researchers left their stories out and the question is why? Is it just that their focus was different when writing their papers? Or is teaching different from the practitioner way of doing data analysis? Or should teaching incorporate more of the statisticians' practice into their teaching? Whatever the answer, research that explores students' reasoning about variation with and without the investigative cycle information may provide insights into pedagogical practice and students' cognition.

A secondary but related issue to the above discussion is that the thinking tools used by statisticians and hence statistics educators tend to be tools for the analysis stage of the empirical enquiry cycle. Wild and Pfannkuch (1999), however, argue that new thinking tools need to be created and used for the other stages of the cycle. The quality management field has created thinking tools for the enquiry cycle such as Joiner's (1994) seven-step process for projects, which gives lists of critical questions and diagrams to prompt, stimulate, and trigger thinking. Quality management defines statistics in terms of variation, and hence variation underpins all its thinking tools. Such tools are successfully used in the quality management field and there seems to be no reason why the statistics discipline and statistics education researchers should not adapt and expand these tools for the general field (see Section 4 of Wild & Pfannkuch, 1999).

4. PURPOSE OF THE STATISTICS CURRICULUM

The third main issue raised upon reading the five papers is that the nature of the analyses seems to be leading students towards a classical inference perspective. This raises the question about the purposes of the thinking tools that were used by the students. EDA and classical inference should be two major strands of statistical investigation in a modern curriculum. It was W. E. Deming in 1953 who first raised the distinction between what he termed enumerative and analytic studies. Enumerative studies are concerned with describing the current situation and inference is limited to the population or process sampled. Current statistics education curricula seem to me to be geared towards formal statistical inference and the carrying out of tests to determine information about the population from the sample data. Causal thinking or seeking explanations or making predictions for phenomena are inadvertently discouraged in such an approach, as probabilistic thinking is a key outcome. Analytic studies involve using data from current processes or populations for understanding and seeking explanations for observed phenomena in order to control or predict the future behavior of processes or populations. It is the measuring and modeling of variation for these purposes as well as explaining and dealing with it that Wild and Pfannkuch (1999) considered important. Gould, the statistician, demonstrates such an approach in his case studies as he attempts to learn more about real world situations through interacting and interrogating the data and to provide some possible solutions (e.g., the identification of chipmunk families).

The differentiation between analytic and enumerative studies suggests a dual approach to reasoning about variation. Consideration of variation should allow for the seeking of explanations, looking for causes, making predictions, or improving processes. Yet at the same time awareness that not all variation can be explained should be raised. That is, there is a need to determine whether the observed differences in sample data reflect the underlying differences in populations, the explained systematic variation, or are due to sampling variability—the unexplained variation statisticians choose to model as random variation. An analytic approach involves learning more about observed phenomena. Each statistics education paper is now considered to raise questions about what the students might have learnt when they interacted with their data if they were afforded an EDA perspective. Again, these remarks do not assume that the EDA perspective was not adopted. Rather, the issue raised is the need to query how teaching could adopt both EDA and classical inference approaches.

In Ben-Zvi's students' conceptions of variability in data, they used their contextual knowledge to realize that a source of variation was in the structure of the two languages. This interaction between data and context enabled them to continue to believe that there was a difference between the two sets of data. What did his students learn about the real world situation? Perhaps they became aware that English surnames are usually longer than Israeli surnames a fact that they had not considered before. Ben-Zvi mentioned that the sources of variation were cultural, historical, and hereditary but he did not seem to explore these ideas with his students.

Bakker's students, when reasoning about the shape of a weight distribution, used their contextual knowledge of weight to explain the type of variability they would expect. Bakker's students were not using actual data when discussing a graph shape for weight; rather, they used "everyday data" that are collected as one operates in the daily environment. They were interrogating data that had not been collected in a planned way. Tversky and Kahneman (1982), Snee (1999), and Pfannkuch and Wild (2004) believed it was possible to have a statistical perspective and understand variation when not being able to collect data. What these students were learning was how to reason about variation using their contextual knowledge. Contextually they learned that weight has a positively skewed distribution, that is, for the Dutch eighth graders population, weights which are much greater than average are more common than weights which are much smaller than average.

Reading facilitates her students to make predictions about future behavior, by requiring students to look at one month of weather data and to argue whether that particular month would or would not be a good time to hold a youth festival. Her students were at the stage of noticing variability within the data and arguing verbally with scant summarization and modeling of data. It seemed that the students were relying on their experience of the context or their "everyday" data that they had collected on the weather and not the information in the data. This raises another question as to how do students learn when to argue by referring to or using the given data, their “everyday” data, or both? What did her students learn about the real world situation? Presumably they learnt about variability present in the weather during one month. Sources of variation such as the season, year, and data collection method could be considered for further exploration of the data as well as wondering if they took the same month of data in different years whether they would detect similar patterns in the data. From the classical inference perspective the students could consider that they had a sample from a population. Hence, with these weather data both an EDA and classical inference approach could be adopted.

Hammerman and Rubin describe teachers comparing slices when making decisions about whether there was a difference between groups. The teachers reduced the variability in the data by continually breaking up the data into categories. In particular, when comparing hours of homework studied between two groups, the teachers argued with a subgroup of the data to state that Amherst students studied more homework than the Holyoke students. If these hours of homework studied were considered, then hypothetically the next step would be to seek an explanation for the variation between the two schools. If there was a belief or research evidence that hours of homework were linked to improved performance, a specific improvement in the Holyoke system, based on the explanation of the variation, could be implemented in order to raise the level of homework hours per week. Another question about homework hours was to consider whether the difference was real or whether it was due to sampling variation. Herein lays the duality of classical inference and EDA. Assuming that the number of homework hours was real data and that the teachers knew the schools and students since they explained the upper end variation by commenting on “lifestyle”, what did these teachers learn about the real situation? When they looked at the “typical student” and found Amherst students did more homework there seemed to be no explanation for this source of variation. It was not clear whether this finding was a revelation to the teachers or a confirmation of what they already knew.

5. CONCLUSIONS AND FURTHER RESEARCH

Variation is at the core of statistical thinking (Moore, 1990) and the research papers reviewed are beginning to uncover how variation reasoning might be integrated into and revealed within the statistics discipline. The task for the students and teachers who participated in the studies reviewed was noticing and distinguishing variability within and between distributions in many forms, such as:

- tabular graphs which can be an integration of information represented as frequencies, percentages, unorganized dots, or count bars;
- graphs which can be dot plots, continuous distributions, or series comparison graphs;
- tables which can contain raw data, frequencies, or percentages.

The question remains about how to enculturate students into reasoning about variation with a variety of analysis thinking tools, as well as into focusing on learning more about observed phenomena. Perhaps Gould provides an avenue for consideration when he states that confirmatory analysis should be de-emphasized and more attention paid to the noise.

Overall, the five papers examined here will contribute greatly towards variation reasoning becoming an integral part of statistics students’ experience and thought. The juxtaposition of these papers raised three key issues which have been the focus of this discussion.

The first key issue is the link between the types of tools students were exposed to and the type of reasoning about variation that was observed. Future research should explore how representations affect, limit, or empower students’ thinking; particularly as statistics learning may require representations that suggest multiple ways of transforming and reorganizing data in order to seek out the information contained in the data. For example, the integrated tabular graph representation does not exist as a conventional statistics tool. Therefore research could explore whether the tools actually assist in enculturating students towards reasoning with conventional statistical thinking tools. A comparison of students' reasoning about variability between the Hammerman and Rubin integrated approach and the Ben-Zvi sequential approach is another possible research avenue. Furthermore, the integrated tabular graph representation that was explored by Hammerman and Rubin has not featured in mathematics education research (Thomas, 2004) and therefore statistics education research should be able to make a contribution to cognitive theories on external representations.

The second key issue is the grounding of students' reasoning about variation within the whole enquiry cycle. Research should not only explore students' reasoning about variation and variability within and between each stage but also consider how the reasoning is connected to the overall cycle. Researchers could also develop thinking tools for all parts of the empirical enquiry cycle and then explore how students reason about variation using these tools.

The third key issue is the need to extend teaching approaches and the curriculum towards an EDA perspective. Herein a tension arises in two goals for learning: learning to use a tool to reason about variation; and reasoning about variation to understand a phenomenon. Thinking of sources of variability and then re-organizing the data to produce more distributions to reason with is part of the process. Reasoning about variation involves the interrogation of the data and requires the interplay between context and the variation in data. The context can be general knowledge about the subject area, knowledge of how the data were collected, how the measures were defined, or the "everyday" data that one collects. Reasoning about variation also involves detecting patterns through the interplay between the centre or trend and the variation. All the data interrogated by these students could have led to them understanding and learning more about an observed phenomenon. Unlocking the story in the data requires both an analytic EDA and an enumerative classical inference approach. Future research needs to consider how students can be enculturated into both ways of reasoning and in particular examine how students reason about variation for the purposes of explanation and prediction.

REFERENCES

Cobb, P., McClain, K., & Gravemeijer, K. (2003). Learning about statistical covariation. *Cognition and Instruction*, *21*(1), 1–78.

Fischbein, E. (1987). *Intuition in Science and Mathematics*. Dordrecht, The Netherlands: Reidel.

Gal, I., & Garfield, J. (1997). Curricular goals and assessment challenges in statistics education. In I. Gal & J. Garfield (Eds.), *The assessment challenge in statistics education* (pp. 1–13). Amsterdam: IOS Press.

Hegedus, S., & Kaput, J. (2004). An introduction to the profound potential of connected algebra activities: Issues of representation, engagement and pedagogy. In M. Hoines & A. Fuglestad (Eds.), *Proceedings of the 28th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 3, pp. 129-136). Bergen, Norway: Bergen University College.

Konold, C., & Miller, C. (2004). *Tinkerplots™ Dynamic Data Explorations*. Emeryville, CA: Key Curriculum Press.

Joiner, B. (1994). *Fourth generation management*. New York: McGraw-Hill Inc.

Mesquita, A. (1998). On conceptual obstacles linked with external representation in geometry. *Journal of Mathematical Behavior*, *17*(2), 183-195.

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

Moritz, J. B. (2000). Graphical representations of statistical associations by upper primary students. In J. Bana & A. Chapman (Eds.), *Mathematics education beyond 2000: Proceedings of the Twenty-third Annual Conference of the Mathematics Education Research Group of Australasia* (Vol. 2, pp. 440-447). Sydney: MERGA.

Pfannkuch, M., & Rubick, A. (2002). An exploration of students' statistical thinking with given data. *Statistics Education Research Journal*, *1*(2), 4–21. [Online: www.stat.auckland.ac.nz/serj]

Pfannkuch, M., & Wild, C. J. (2003). Statistical thinking: How can we develop it? In *Bulletin of the International Statistical Institute 54th Session Proceedings, Berlin, 2003* [CD-ROM]. Voorburg, The Netherlands: International Statistical Institute.

Pfannkuch, M., & Wild, C. (2004). Towards an understanding of statistical thinking. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 17-46). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Shaughnessy, J. M. (1997). Missed opportunities in research on the teaching and learning of chance and data. In F. Biddulph & K. Carr (Eds.), *People in mathematics education: Proceedings of the Twentieth Annual Conference of the Mathematics Education Research Group of Australasia* (Vol. 1, pp. 6-22). Sydney: MERGA.

Shaughnessy, J. M., Garfield, J., & Greer, B. (1996). Data handling. In A. Bishop, K. Clements, C. Keitel, J. Kilpatrick, & C. Laborde (Eds.), *International handbook of mathematics education* (Part1, pp. 205–237). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Snee, R. (1999.) Discussion: development and use of statistical thinking; a new era. *International Statistical Review*, *67*(3), 255–258.

Thomas, M. (2004). Personal communication, 14 October 2004, Auckland, New Zealand.

Tukey, J. (1977). *Exploratory data analysis.* Reading, MA: Addison-Wesley.

Tversky, A., & Kahneman, D. (1982). Judgment under uncertainty: Heuristics and biases. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 3–20). New York: Press Syndicate of the University of Cambridge. (Originally published in *Science*, *185*(1974), 1124–1131)

Watson, J., Collis, K., Callingham, R., & Moritz, J. (1995). A model for assessing higher order thinking in statistics. *Educational Research and Evaluation*, *1*, 247-275.

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

MAXINE PFANNKUCH
Department of Statistics
The University of Auckland
Private Bag 92019
Auckland
New Zealand