Tasmanian Research in Chance and Data

Jane Watson

University of Tasmania

Abstract

Over the past decade since the introduction of *A National Statement on Mathematics for Australian Schools*, research in Tasmania has focused on the development of students' understanding of the chance and data part of the curriculum. Branching off from this have been projects considering higher order statistical thinking while students work in collaborative groups, profiles of teachers' relationships to the curriculum, the influence of cognitive conflict on student understanding, and the impact of instruction specifically aimed at students' understanding of variation in relation to chance and data. A summary of some of the outcomes of this research will be presented, as well as potential directions for future research.

Introduction

In 1991, *A National Statement of Mathematics in Australian Schools* (Australian Education Council [AEC]) brought "chance and data" into the school curriculum as one of five content areas to be covered in the area of mathematics. This was followed in 1994 by *Mathematics – A Curriculum Profile* (AEC), which detailed expected outcomes for students at eight levels through the years of compulsory schooling. During this period New Zealand was grappling with similar curriculum revision and in 1992 *Mathematics in the New Zealand Curriculum* (Ministry of Education) presented "statistics" as one of the five corresponding content areas in its mathematics curriculum. This document also suggested eight levels but covering up to Year 13 of schooling. The objectives of the writers in the two countries were similar and reflected to some extent moves that had taken place in the United States in 1989 with the National Council of Teachers of Mathematics publication of its *Curriculum and Evaluation Standards for School Mathematics*. In that document "probability" and "statistics" were both covered in standards covering twelve years of schooling. Similar moves took place in other western countries.

Although a major step forward for statistics education, the difficulties associated with the production of these documents centred around the lack of research on children's understanding of statistical concepts, the lack of experience of school teachers in this area, and the consequent dependence on tertiary statisticians for advice on what should be in the school curriculum. Except for the work of Fischbein (1975) and Green (1983) in the area of probability, virtually no evidence had been gathered on what students could do at various ages. Still today research on students' statistical understanding is far behind that of other areas of the mathematics curriculum (e.g., Hart, 1981) but progress has been made. It is now becoming possible to provide teachers with an appreciation of what can be expected from students and with potentially useful activities. Research has also begun to provide a bottom-up (from the perspective of young learners) rather than a top-down (from the perspective of tertiary statisticians) view of how understanding can be developed. An example of this is the topic of variation, the foundational concept of all statistical investigation (no variation, no statistics). The documents noted earlier rarely use the word and focus instead on the centre, or arithmetic mean, of data sets in the middle years of schooling. The standard deviation, the classical measure of variation, being a complex algorithm, is not imposed until late in the school curriculum, and hence variation itself gets neglected. Building on young children's intuitions about change and difference, however, I believe it is possible to build a foundation that will make the standard deviation, when it is finally introduced, a natural measure rather than a magical mystery (as it still is today for many tertiary students). But this is getting ahead in the historical account of research into the chance and data curriculum. First I want to give an overview of Tasmanian research. This research relates mainly to school children and the development of their understanding of statistical concepts over the years of schooling.

Table 1 contains a summary of the "chance and data" projects funded by the Australian Research Council since 1993, and I will consider each briefly in turn, to illustrate some outcomes and how our ideas have evolved over time. A complete list of references is available and within these papers are references to many other researchers around the world who are now working in this area. Personal colleagues who have contributed significantly to the research reported here are Professor Kevin Collis, Jonathan Moritz, Dr Helen Chick, Rosemary Callingham, Ben Kelly, and Professor Mike Shaughnessy.

Table 1. Time line of Tasmanian research projects in chance and data

1993 →	Students' understanding of concepts
1995 →	Students working in groups
1997 →	Profiling teachers (from 93)
1998 →	Concepts and cognitive conflict
2000 →	Teaching for appreciation of variation
2002 →	A model for statistical literacy

I. Students' Understanding of Statistical Concepts

At the beginning of implementation of the curriculum, the objective was to describe the development of students' understanding of the basic concepts within the chance and data curriculum. With this in mind surveys were developed using items from previous researchers (often used at higher levels of education) such as Green (1983), Konold and Garfield (1992), and Tversky and Kahneman (1971), as well as original items reflecting the new curriculum content and reflecting the application of these ideas in newspaper articles and graphs. Interview protocols were also developed and these were used with selected students who completed the surveys in Grades 3, 6, and 9. Longitudinal data were collected using the surveys two and four years later, and interviews three or four years later. The richness of the data on the specific topics has meant that in-depth reports have been written on chance measurement, sampling, average, beginning inference, and graphing topics, rather than combining data in a single scale to report student outcomes.

Several frameworks have been used in describing student outcomes but usually the underpinning model has been associated with the SOLO Taxonomy of Biggs and Collis (1982; 1991). This model, arising from a Piagetian perspective, is hierarchical in nature and has common features with other models employed in education today (Pegg, 2002). The SOLO model is based on modes of functioning and it is in the concrete symbolic mode, where most school-based learning experiences take place, in which most observations in our research occur. Within modes, however, the SOLO model describes a progression of levels as students integrate more elements of a particular problem context into a response. Briefly these levels reflect four types of reasoning: (a) ikonic, intuitive reasoning (IK); (b) unistructural reasoning: single ideas, contradictions not noticed (U); (c) multistructural reasoning (M): sequential use of ideas, recognition of contradictions but no resolution; and (d) relational reasoning (R): integration of ideas to form a whole, resolution of contradictions. The levels informed the analysis of many items and clusters of items, as did a derived hierarchy of statistical literacy that was particularly useful for the media-based items (Watson, 1997). The three tiers of the hierarchy are (1) understanding the basic terminology, (2) understanding terminology in social contexts, and (3) developing the ability to question statistical claims that are made without proper justification.

What are some of the things we learned about the development of students' understanding of statistical topics? The following tables highlight some of the increasingly complex structural features of responses for several topics. Similarities are also seen across topics. Table 2 considers chance measurement (Watson, Collis, & Moritz, 1997), Table 3 considers average (Watson & Moritz, 1999a; 2000), and Table 4 considers beginning inference (Watson & Moritz, 1999b). In each case there is an extension to the basic four levels of response, which represents a consolidation of an idea and the beginning of another potential cycle of response within the concrete symbolic mode. For the three problems used for chance measurement, the first two involving an isolated chance environment, whereas the third involves a comparison of two settings. A comparison of proportions is required for the third and this involves a consolidation of understanding in a second cycle.

Table 2. Levels of response for chance measurement

	Is a 1 or a 6 easier to throw?	13 boys & 16 girls names in a hat – boy or girl picked?	Box A (6 red, 4 blue marbles) Box B (60 red, 40 blue marbles) Which choose for blue?
IK	I get 1s	Girls because pretty	Blue favourite colour
U_1	(=) anything can happen, luck, not looking		
M_1	qualitative description		
	(=) one of each number	(=) boys/girls (g) more girls	(B) more marbles more blue (A) less marbles less red
R_1	quantitative measurement		
	(=) 1/6 chance	(g) 16/29 chance	Both more red, (A) Difference 20, 2
U_2^+			Proportional reasoning (=) same ratio, percent

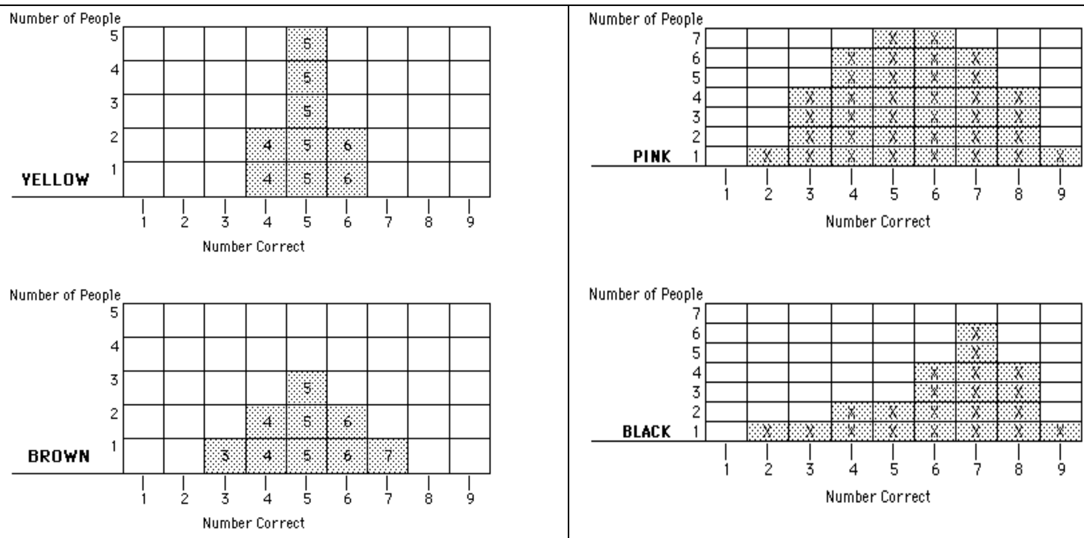
Of interest from our work on average (Table 3) is the observation that the mean is not the natural model that students use when asked to discuss the concept of average. In surveys of students in years 3 to 11, when students were asked, "What does it mean to be average?" they replied with ideas associated with the arithmetic mean 9% of the time, the median (middle) 60% of the time, and the mode (most) 36% of the time.

Table 3. Levels of response for average

	What does 'average' mean? What does 'students watch an average of 3 hours of TV per day' mean? How to obtain? On average family has 2.3 children. Work backward for 10 families with 2.3 children. Weighted mean problem.
IK	Imaginative stories: no explanation for average
U	"Normal", "okay", stories from context, "add up"
M	"Most", "middle", "add-and-divide", may recognise conflict of incorrect mean and the mode; very little progress on complex questions
R	"add-and-divide" in straight-forward context; understand 2.3; "most" compatible with "mean"; use visual features in preference to mean; partial progress on complex questions
Adv Avg	Basic representational understanding plus at least one problem in complex context based on arithmetic mean

For questions of comparing two groups presented in graphical form (Table 4), the distinction between sets of equal and unequal size was a task factor that split responses into two similar cycles. What we see in each of these cases is a building up in a three-step cycle of an initial concept: basic probability of an event, concept of average, and comparison of equal-sized data sets. Moving on to more complex problem solving involves consolidation of the first cycle understanding, plus for chance measurement and comparing graphs, proportional reasoning in unequal data sets; and for average, the ability to reverse and reconstruct the algorithm for the mean (also with underpinnings of proportional reasoning).

Table 4. Levels of response for inference from graphs



equal-sized sets	unequal-sized sets

U_1	Use single feature of graph for comparison of equal-sized sets – “Red more”
M_1	Use multi-step visual comparisons <u>or</u> numerical calculations for comparison of equal-sized sets – compare several columns <u>or</u> calculate totals
R_1	Integrate both visual and numerical information in comparing equal-sized sets
U_2	A single visual comparison to compare unequal-sized sets – “Black higher for the amount of people”
M_2	Multi-step visual comparison <u>or</u> comparison of means for unequal-sized sets
R_2	Both visual and numerical information integrated for comparison of unequal-sized sets

In terms of thinking about the school curriculum we might think of this process as building the basic conceptual understandings in the primary years and applying this understanding in more complex contexts in the secondary years. In all studies there were trends for students in higher grades to display higher levels of reasoning but the variation indicated that in any class a teacher could expect a wide range of observed outcomes.

II. Students Working in Groups

A classroom trial of a protocol that had proved too time-consuming for individual interviews, led to the exploration of what students working collaboratively in groups of three would produce in terms of hypothesizing and data representation from a set of 16 data cards containing information on children, including age, weight, favourite activity, eye colour, and number of fast food meals eaten per week. For part of this research students worked in isolated groups and for part, in groups in the classroom.

Part of the analysis in this project, associated with nine groups of grade 5/6 children working in groups of three with the 16 data cards, produced levels of observed outcomes again associated with the SOLO Taxonomy. These are summarised for interpreting and representing the data sets in Table 5 (Chick & Watson, 2001). Considering 26 of the students involved, 14 students interpreted and represented the data at the same SOLO level, whereas 12 could interpret at a higher level than they could represent.

Table 5. Levels of interpreting and representing for the data cards task for students working in groups of three

	Interpreting	Representing
U	Individual aspects	
	Reason based on single cards	Depict all data values in a table with no aggregation
M	Several aspects used in sequence	
	Consider all data but only one variable – more people like TV	Represent a single variable for all data – bar chart of eye colour
R	Integrated understanding of relationships	
	Propose cause-effect relationship of variables	Depict two variables – a scattergram

Another aspect of the observations of video tapes of these and other students working in groups, however, related to the factors that influence cognitive outcomes (Watson & Chick, 2001a) and to whether asking for or offering help was productive in terms of outcomes (Watson & Chick, 2001b). Seventeen factors were identified that contributed to improved ("lifting"), static ("hovering"), or reduced ("falling") outcomes for the group. These are clustered into three types in Table 6. They are not particularly surprising to people who work with children but what may be surprising is that all factors were observed at some point to be associated with each of lifting, hovering, and falling. Predictions are hence difficult to make and factors such as "collaborative type," overall environment, and grade level were found to be contributors to the imbalance of expected outcomes.

Table 6. Factors associated with outcomes in group work settings

Cognitive Factors	
Cognitive ability Previous experience Cognitive disagreement Doubt Misunderstanding	Tenacity of ideas The big picture Picking the easiest ideas Organisational collaboration
Social or Interpersonal Factors	
Leadership Social disagreement Egocentrism	Social collaboration Other social factors (e.g. gender balance)
External factors	
Task factors Outsider (e.g. teacher)	Environment (e.g. noise)

A detailed analysis of a video of three grade 6 boys working with the data cards confirmed the factors and how they combined to facilitate or otherwise seven aspects of emergent statistical understanding: association, average, graphing, the need to justify views with data, part/whole relationships, mathematical and statistical tools required, and the limitations of a data set (Chick & Watson, 2002). Although novices these boys dealt with some quite sophisticated concepts.

In terms of specific questioning, i.e., asking for help, which occurred with groups of three working on the data cards task and with two teacher/researchers in the classroom, the outcomes will not surprise many but will disappoint some. The answer to "Does help help?" was a qualified "Yes," depending on how and by whom it was requested and how and by whom it was provided. Teachers' questions were more productive in achieving higher level question-answer-outcome sequences than students' questions. Few student questions were observed at high levels. Boys asked more off-task questions and were associated with behaviour provoking more teacher interaction with them.

III. Profiling Teachers

One of the original objectives of the research was to gauge teacher understanding of the curriculum and pedagogical knowledge in relation to teaching chance and data (Shulman, 1987). Initial interviews with 72 teachers led to a detailed analysis of their understanding and views on sampling (Watson & Moritz, 1997) and further development of a teacher profiling instrument in association with a large professional development project for teachers involving technological innovation (Watson, 1998). The final profile reflects the need for criteria to assess teacher competence in relation to professionalism and the need for professional development as the curriculum changes (Watson, 2001). The main sections of the profile are given in Table 7.

Table 7. Profile of Teachers' Competence and Confidence to Teach Chance and Data

Significant factors for teaching chance and data (brainstorming)
Preparing to teach a unit in Chance and Data (1) overall planning
Preparing to teach a unit in Chance and Data (2) particular topic
Teaching practices: grades, time, resources
"Sample" in Chance and Data
"Average" in Chance and Data
Confidence – particular topics
Statistics in Everyday Life – attitudes
Student survey items – likely student responses (1) Survey in shopping centre (4) Chicago vs US sample (2) Pie chart adding to 128.5% (5) Cause-effect graph (3) Odds of 7-2 (6) Coin-tossing
Teacher background – details
Professional development – past exposure, future needs

Findings for the sample of Australian teachers included on one hand that although primary teachers taught many activity-based lessons, there was little evidence of coherent program planning. On the other hand at the senior secondary level there was good documentation of traditional programs, but there was little effort to provide activity-based sessions to assist students with difficulties and to reinforce theory. As might be expected primary teachers had less confidence in teaching topics in probability, odds, median, and sampling than secondary teachers. Some teachers were not aware of their students' difficulties and there appeared to be a need in many cases for professional development incorporating pedagogical content knowledge, topic content knowledge, and curriculum knowledge. Many were not aware of the important curriculum documents in the field.

IV. Concepts and Cognitive Conflict

The availability of video clips of students expressing their opinions on chance and data concepts and solutions to problems meant we could interview new students and provide them with conflicting views to those they had expressed and to some extent mimic a classroom environment where students express differing views (Watson, 2002a, 2002b; Watson & Moritz, 2001a, 2001b). The chance to use particular students' views repeatedly provided a control not present in a classroom discussion. The ability to argue back and forth was very limited, however, with the interviewer only able to reiterate or perhaps clarify the opinion expressed on video. In this project, 20 students from each of Grades 3, 6, and 9 were interviewed with some of the original protocols and during this time presented with alternative views at a higher or lower level. These were contained on video clips with typed transcripts, shown on a laptop computer. Of particular interest was the power of higher level arguments to convince interviewees in a short period of time. Also of concern was whether less viable responses at lower levels would dissuade students from their original views. Students were chosen for these interviews by their teachers as those who would express their views and enjoy being challenged. Hence the students were likely to be of above average ability in their grades.

Given the existence of data on longitudinal interviews after three or four years, and hence the documentation of improved levels of observed outcomes, it is possible to compare improvement in the two settings. Although this process is fraught with difficulties, particularly the retention and transfer of new levels of understanding obtained in a few minutes, it is still useful to make comparisons with a much longer period where specific intervention was not undertaken by the researchers. Table 8 contains some comparisons for different groups of students for several tasks.

Table 8. Percent of improved level of response after presentation of cognitive conflict and after a three- or four-year time span

Topic	Cognitive Conflict	3-4 Years
Beginning inference - equal sized sets	57%	NA
Representing in pictographs	60%	36%
Proportional chance measurement	33%	33% (surveys)
Beginning inference - unequal sized sets (proportional)	30%	31%
Sampling	22%	78%
Predicting from pictographs	30%	86%

It is interesting to note that for what might be considered easier tasks – comparing sets of equal size and creating a pictograph – improvement levels with cognitive conflict are similar. For more difficult tasks involving proportional reasoning, understanding of sampling methodology, and the ability to predict from a graph (at grade 3), the improvement rates are also similar to each other but about half of that for the easier tasks. Comparisons with 3-or-4-year improvement are not always as consistent. It is interesting to note however, that the two tasks relying on proportional reasoning for improvement were associated with outcomes that were similar for cognitive conflict interviews and longitudinal interviews. This is an area where further research could be very useful.

V. Variation

Growing interest in students' understanding of variation (Green, 1993; Shaughnessy, 1997) led to a more recent project intended both to document students' understanding and to carry out a teaching intervention designed to emphasize variation in relation to the teaching of the chance and data curriculum. This project involved interviews with 73 students in Pre-Grade-1 and Grades 3, 5, 7, and 9 to profile understanding. It also involved pre test, post test, and two-year longitudinal tests (the first and last with control groups) with over 700 students in order to monitor change associated with the teaching units. For Grades 3 and 5, a teacher was provided for a 10-lesson unit on chance and data. For Grades 7 and 9, units were provided for teachers to choose from in their usual mathematics planning. Although not all analyses are completed, the pretest-posttest results for the experimental classes are summarised by grade in Table 9 (Watson & Kelly, 2002a, 2002b, 2002c), where the levels of improvement for four subscales – basic chance and data, variation in chance, variation in data, and variation in sampling – plus the total score are given. These are based on paired *t*-tests. For high school grades there was variation among classes, with one class in each grade not improving. These are encouraging outcomes but the preliminary comparison for some items and subscales with the control groups after two years, indicates little difference.

Table 9. Improvement following teaching intervention emphasizing variation

Scale	Grade 3 (n = 72)	Grade 5 (n = 82)	Grade 7 (n = 92)	Grade 9 (n = 90)
Basic chance and data	$p < 0.001$	NS	$p < 0.001$	$p < 0.01$
Variation in chance	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$
Variation in data	$p < 0.01$	$p < 0.01$	$p < 0.001$	$p < 0.001$
Variation in sampling	$p < 0.01$	$p < 0.001$	$p < 0.001$	$p < 0.01$
Total score	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$

VI. A Model for Statistical Literacy

Having collected survey data from over 4000 students on a wide range of items related to chance and data, including variation and applications in media contexts, it is time to put it all together and propose a model of development of understanding (Watson & Callingham, 2003). The wide range of contexts within which items were set permits an interpretation of a variable associated with the goal of statistical literacy and achieving a level of critical questioning by the time students leave school. Contributing to this are the mathematical and statistical skills required of tasks, the understanding of concepts alone and in context, and then the ability to question in various social contexts. Initial work in this area has employed the Rasch model (Masters, 1982; Rasch, 1980) to produce a variable map that simultaneously plots student ability and item difficulty on the same graph. A student on the same value as an item has a 50% chance of achieving success on that item. With multiple coding levels applied to items reflecting hierarchical structure, the objective is to be able to explain the overall distribution of items in terms of a global structure. Initial work suggests a developmental sequence similar to that in Table 10, where it is seen that engagement with context is a salient feature. Using items with contexts ranging from very "classroom-mathematical" like tossing a die, to "classroom-social" like planning a school survey, to "unfamiliar-social" like critiquing a media article, brought out how there is a hierarchy of contexts present in items as well as other hierarchical aspects.

Table 10. Six hypothesized levels of development of Statistical Literacy

Level 1	Idiosyncratic-personal engagement with context using basic graph/table reading skills.
Level 2	Colloquial-informal engagement with context using basic chance, graph, and numeracy skills.
Level 3	Selective engagement with context involving qualitative interpretation of statistical ideas.
Level 4	Appropriate non-critical engagement with context using basic statistical skills.
Level 5	Critical-questioning engagement with context using appropriate statistical terminology.
Level 6	Critical-questioning engagement with context using sophisticated mathematical-statistical understanding.

Recent work in this area has focused on developing two shortened survey forms for use in classrooms (Watson & Callingham, 2004), and using 10-year longitudinal data to confirm the developmental aspects of the model (work still in progress).

Conclusion

Returning to where we started, it is hoped that the 10-year longitudinal data will also inform us on the effect that the chance and data curriculum has had in Tasmania in 1993. Several other areas of research also deserve attention with respect to statistics education in schools. These include

- the development and testing of activities to assist students to improve their levels of performance.
- work with teachers to increase their content and pedagogical content knowledge in a research context that will measure evidence of change both for them and their students.
- work with technology to enhance learning of students (and maybe teachers); e.g. software like "Tinkerplots" (Konold, 2003) and "Fathom" (Key Curriculum Press, 2003), and learning contexts such as websites and Web CT.

There is much more to be done and there are exciting times ahead.

Non Tasmanian References

Australian Education Council. (1991). *A national statement on mathematics for Australian schools*. Carlton, Vic. : Author.

Australian Education Council. (1994). *Mathematics - A curriculum profile for Australian schools*. Carlton, Vic.: Curriculum Corporation.

Biggs, J.B., & Collis, K.F. (1982). *Evaluating the quality of learning: The SOLO taxonomy*. New York: Academic Press.

Biggs, J.B., & Collis, K.F. (1991). Multimodal learning and the quality of intelligent behaviour. In H. A. H. Rowe (Ed.), *Intelligence: Reconceptualization and measurement* (pp. 57-76). Hillsdale, N. J.: Lawrence Erlbaum.

Fischbein, E. (1975). *The intuitive sources of probabilistic thinking in children*. Dordrecht: D. Reidel Publishing Company.

Green, D.R. (1983). A survey of probability concepts in 3000 pupils aged 11-16 years. In D. R. Grey, P. Holmes, V. Barnett, & G. M. Constable (Eds.), *Proceedings of the First International Conference on Teaching Statistics: Vol. 2* (pp. 766-783). Sheffield, UK: Teaching Statistics Trust.

Green, D. (1993). Data analysis: What research do we need? In L. Pereira-Mendoza (Ed.), *Introducing data analysis in the schools: Who should teach it?* (pp. 219-239). Voorburg, The Netherlands: International Statistical Institute.

Hart, K.M. (Ed.) (1981). *Children's understanding of mathematics: 11-16*. London: John Murray.

Key Curriculum Press. (2003). *Fathom: Dynamic statistics*. Emeryville, CA: Author.

Konold, C. (2003). *Tinkerplots: Tools and curricula for enhancing data analysis in the middle school*. NSF Grant ESI-9818946. Amherst, MA: Author.

Konold, C., & Garfield, J. (1992). *Statistical reasoning assessment: Part 1. Intuitive Thinking*. Scientific Reasoning Research Institute, University of Massachusetts, Amherst, MA.

Masters, G.N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, *47*, 149-174.

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

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, Va.: Author.

Pegg, J.E. (2002). Fundamental cycles of cognitive growth. In A. D. Cockburn & Elena Nardi (Eds.), *Proceedings of the 26th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 41-48). Norwich, UK: PME.

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests*. Chicago: University of Chicago Press (original work published 1960).

Shaughnessy, J.M. (1997). Missed opportunities in research on the teaching and learning of data and chance. In F. Biddilph & K. Carr (Eds.), *Proceeding of the Twentieth Annual Conference for the Mathematics Education Research Group of Australasia: People in mathematics education* (pp.6-12). Waikato, NZ: MERGA.

Shaughnessy, J.M., Watson, J., Moritz, J., & Reading C. (1999, April). *School mathematics students' acknowledgment of statistical variation.* In C. Maher (Chair), There's more to life than centers. Presession Research Symposium conducted at the 77th Annual National Council of Teachers of Mathematics Conference, San Francisco, CA.

Shulman, L.S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57*, 1-22.

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin, 76(*2), 105-110.

Selected Tasmanian References

Chick, H.L., & Watson, J.M. (2001). Data representation and interpretation by primary school students working in groups. *Mathematics Education Research Journal, 13*, 91-111.

Chick, H.L., & Watson, J.M. (2002). Collaborative influences on emergent statistical thinking—A case study. *Journal of Mathematical Behavior, 21*, 371-400.

Watson, J.M. (1997). Assessing statistical literacy using the media. In I. Gal & J.B. Garfield (Eds.), *The Assessment Challenge in Statistics Education* (pp. 107-121). Amsterdam: IOS Press and The International Statistical Institute.

Watson, J.M. (1998). Professional development for teachers of probability and statistics: Into an era of technology. *International Statistical Review, 66*, 271-289.

Watson, J.M. (2001). Profiling teachers' competence and confidence to teach particular mathematics topics: The case of chance and data. *Journal of Mathematics Teacher Education*, *4*, 305-337.

Watson, J.M. (2002a). Creating cognitive conflict in a controlled research setting: Sampling. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics: Developing a statistically literate society, Cape Town, South Africa*. Voorburg, The Netherlands: International Statistical Institute.

Watson, J.M. (2002b). Inferential reasoning and the influence of cognitive conflict. *Educational Studies in Mathematics, 51*, 225-256.

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

Watson, J.M. & Callingham, R.A. (2004). *Statistical literacy: From idiosyncratic to critical thinking*. Paper submitted to the IASE Roundtable Conference, Lund, Sweden.

Watson, J.M., & Chick, H.L. (2001a). Factors influencing the outcomes of collaborative mathematics problem solving—An introduction. *Mathematical Thinking and Learning, 3*(2&3), 125-173.

Watson, J.M., & Chick, H.L. (2001b). Does help help?: Collaboration during mathematical problem solving. *Hiroshima Journal of Mathematics Education, 9*, 33-73.

Watson, J.M., Collis, K.F., & Moritz, J.B. (1997). The development of chance measurement. *Mathematics Education Research Journal, 9*, 60-82.

Watson, J.M., & Kelly, B.A. (2002a). Can grade 3 students learn about variation? In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics: Developing a statistically literate society, Cape Town, South Africa*. Voorburg, The Netherlands: International Statistical Institute. Reprinted in B. Phillips (Ed.), *ICOTS6 Papers for School Teachers* (pp. 83-88). Voorburg, The Netherlands: International Statistical Institute.

Watson, J.M., & Kelly, B.A. (2002b). Grade 5 students' appreciation of variation. In A. Cockburn & E. Nardi (Eds), *Proceedings of the 26th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 385-392). Norwich, UK: University of East Anglia.

Watson, J.M., & Kelly, B.A. (2002c). Variation as part of chance and data in grades 7 and 9. In B. Barton, K.C. Irwin, M. Pfannkuch, & M.O.J. Thomas (Eds.), *Mathematics education in the South Pacific* (Proceedings of the 26th Annual Conference of the Mathematics Education Research Group of Australasia, Auckland, Vol. 2, pp. 682-289). Sydney, NSW: MERGA.

Watson, J.M., & Moritz, J.B. (1997). Teachers' views of sampling. In N. Scott & H. Hollingsworth (Eds.), *Mathematics creating the future* (Proceedings of the 16th Biennial Conference of the Australian Association of Mathematics Teachers Inc., pp. 345-353). Adelaide, SA: AAMT.

Watson, J.M., & Moritz, J.B. (1999a). The development of concepts of average. *Focus on Learning Problems in Mathematics, 21*(4), 15-39.

Watson, J.M., & Moritz, J.B. (1999b). The beginning of statistical inference: Comparing two data sets. *Educational Studies in Mathematics, 37*, 145-168.

Watson, J.M., & Moritz, J.B. (2000). The longitudinal development of understanding of average. *Mathematical Thinking and Learning, 2*(1&2), 11-50.

Watson, J.M., & Moritz, J.B. (2001a). Development of reasoning associated with pictographs: representing, interpreting, and predicting. *Educational Studies in Mathematics, 48*, 47-81.

Watson, J.M., & Moritz, J.B. (2001b). The role of cognitive conflict in developing students' understanding of chance measurement. In J. Bobis, B. Perry, & M. Mitchelmore (Eds.) *Numeracy and beyond* (Proceedings of the 24th Annual Conference of the Mathematics Education Research Group of Australasia, Vol. 2, pp. 523-530). Sydney, NSW: MERGA.