A Two-Year Study of Students' Appreciation of Variation in the Chance and Data Curriculum

Jane M. Watson
University of Tasmania
<Jane.Watson@utas.edu.au>

Ben A. Kelly
University of Tasmania
<Ben.Kelly@utas.edu.au>

This report considers the difference in performance of students in Grades 3, 5, 7, and 9 over a two-year interval, 2000 to 2002, on a survey of concepts in the chance and data curriculum. Two types of comparisons occur. First, longitudinal change within students is measured and this is compared for students in schools where an instructional intervention occurred and those in schools with no intervention. Second, in Grades 5, 7, and 9 in 2002 performance is compared with that in the same grades in the same schools in 2000. Reasons for differences are suggested based on the educational experiences that the 2002 students experienced in the intervening time. Students answered questions concerning basic chance and data understandings, and variation in chance, data and graphing, and sampling. Scores on these four subscales as well as the overall survey are used to make comparisons.

Research often proceeds in surprising ways and unexpected outcomes occur. This report is the story of such an occurrence in a study of the development of students' understanding of statistical variation and the trialling of intervention lessons to emphasise variation in relation to the chance and data curriculum.

Having had a long standing interest in chance and data (Watson, 1978) before its inclusion in the Australian mathematics curriculum (Australian Education Council [AEC], 1991, 1994) and an initial stake in research based on the curriculum (Watson, 1992), it was natural for the first author to extend this interest to the underlying concept of statistical variation in response to calls of Shaughnessy (1997) and others. With curriculum documents only implicitly recognising its importance, Shaughnessy, Watson, Moritz, and Reading (1999) took an item from a United States national survey and adapted it to allow students to show their appreciation of variation. This led to a proposal to study the change in understanding of variation within the chance and data curriculum across the middle school years and to intervene in the classroom to improve understanding. The study took place in Tasmania and various reports have been written on student understanding (e.g., Kelly & Watson, 2002; Watson & Kelly, in press) and on change in student performance after instruction (Watson & Kelly, 2002a; 2002b; 2002c). The survey outcomes were based on an instrument, the analysis of which has also contributed to the discussion on growth in student understanding (Watson, Kelly, Callingham, & Shaughnessy, 2003). Although positive outcomes have been achieved from instructional intervention, part of the design of the study included two-year longitudinal testing of students in the intervention schools and in schools matched for socio-economic status. Not only did this testing provide the opportunity to compare schools with specific project intervention to schools without it, but also there was the chance to compare students two years further along with those in the same grades in the initial year. This comparison gives some indication of whether the general school environment (intervention or not) had an impact on student performance. Figure 1 illustrates the two kinds of comparison that were possible.

Formal hypotheses could be stated but suffice it to say that the interest lies in comparing the change in two instances. In the first, using the paired comparisons: what changes occurred for grades over a two-year period in the left of Figure 1? Was the change observed on the left of Figure 1 the same or different for intervention and non-intervention schools? In the second instance, what change occurred within school groups for grades that

were first surveyed in 2000 compared to the same grades surveyed in 2002 that had been surveyed two years earlier in a grade two grades lower as shown on the right in Figure 1. Were these changes (if they occurred) different for the intervention and non-intervention schools? Are there any explanations for the observations of the study?

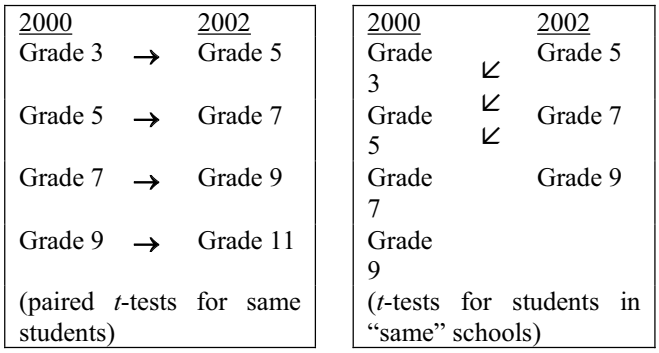

Figure 1. Possible Comparisons.

Methodology

Sample. Ten schools took part in the study with the sample sizes used in the current study for Grades 3, 5, 7, and 9 given in Table 1. In each case the initial sample size in 2000 was larger but because of the need to compare individual students, only those who remained in the study in 2002 are included in this report. In each case there was one or more primary school, which was a feeder school for a local high school. These choices were made to make it possible to follow students from Grade 5 to Grade 7. Students initially in Grade 3 or 7 would be expected to be in the same school two years later, whereas Grade 9 students either changed schools or left school in the following two years.

Table 1
Number of Students per Grade and Sample

	Grade 3/5[1]	Grade 5/7[1]	Grade 7/9[1]	Grade 9/11[1]	Total
Intervention	56	61	72	30	219
Non-Intervention	67	44	68	31	210

1Grade in the longitudinal follow-up

Instrument. The survey instruments consisted of twenty-six items for all four grades, with additional items added for Grade 5, 7, and 9. All items are presented in Watson et al. (2003) and were clustered into four subscales. These four subscales measured understanding of Basic Chance and Data (BCD), Variation in Chance (CV), Variation in Data and Graphing (DV), and Variation in Sampling (SV). As well as this, an overall Total score for the survey was calculated.

Procedure. Teachers and the research team administered the survey to all grades in class time. After the administration of the survey, five of the ten schools took part in a series of lessons on chance and data with an emphasis on variation. For Grades 3 and 5 (3 schools), a 10-lesson unit was taught by the same teacher, who was supplied by the research team. The lessons are described in detail in Watson and Kelly (2002a, 2002b). The students in Grades 7 and 9 (2 schools) were taught by their usual mathematics teachers who were provided with a unit of six lessons devised by the researchers to cover a range of experiences relating to the chance and data curriculum with an emphasis on variation.

Suggestions were made to the teachers as to the order in which the lessons might be taught but the decisions of order of presentation and which lessons to choose, were left to the professional judgement of each teacher, as some teachers could not commit to the entire unit due to time constraints. A detailed description of the lessons provided by the researchers for Grades 7 and 9 and which lessons were presented can be found in Watson and Kelly (2002c). During the period of intervention, the remaining five non-intervention schools (3 primary and 2 high) did not receive any specialised instruction related to the research program. Students in intervention schools completed a post-test about six weeks after the unit finished and all grades showed significant improvement in the Total scores ($p<0.001$, Watson & Kelly, 2002a, 2002b, 2002c).

Two years after the survey was originally conducted, the researchers administered the same survey to students in Grades 5, 7, and 9 at the same ten schools. For students who were in Grade 9 when the survey was originally conducted, all had changed schools for Grade 11. The researchers located as many students as possible who had completed the original survey and administered the survey in four extra senior secondary colleges. The retention rate for students moving into Grade 11 from high school was smaller due to students leaving the education system. In 2002 all grades were administered the same survey as was given to the equivalent grade two years earlier in 2000.

Analysis. For the two-year longitudinal within-student change, 40 paired *t*-tests were performed, in each case for the original set of questions, and the conservative Bonferroni correction suggests a significance reduction from 0.05 to 0.0012. In the light of the information provided by the large number of *p*-values less than 0.05 (25) compared to the expected number (2), however, all *p*-values less than 0.05 are reported for consideration. To show the differences, if any, in the change in performance between the intervention and non-intervention schools, 20 *t*-tests were performed on difference scores (for each grade and subscale). One *p*-value less than 0.05 occurred, consistent in this case with chance. For the change within school types over the two-year period, 30 *t*-tests were performed, in each case for the questions for that grade level, with an expectation of 1.5 *p*-values less than 0.05. Since seven occurred, with six in the expected direction, all *p*-values less than 0.05 are reported for consideration.

Results

Longitudinal within Student Comparisons

Table 2 shows the mean scores and standard errors for each of the four subscales and the Total score for each grade in both the intervention and non-intervention schools. As can be seen there was a significant increase in the overall performance (Total) from the pre-test to the longitudinal follow-up for students originally in Grades 3 and 7 in the intervention schools, and students originally in Grades 3, 7, and 9 in the non-intervention schools. For Grade 3 students this represented an improvement of more than a standard deviation since the pre-test, whereas in Grade 7 it was between 2/3 and 3/4 of a standard deviation, and in Grade 9 it was only 3/8 of a standard deviation. There was, however, no significant difference between the pre-test and the longitudinal follow-up means for students originally in Grade 5 for either condition, with a minor decrease in performance for students originally in Grade 5 for the non-intervention schools. Significant initial improvement on the survey for the intervention schools occurred after the instruction supplied by the research team as measured by the post-test, which was administered six weeks after the instruction (Watson & Kelly, 2002a, 2002b, 2002c), and this was sustained

for Grades 3 and 7 after two years. It is not known when the improvement for the non-intervention schools occurred.

Table 2.
Mean Scores and Standard Errors for Each Subscale and Total Score for Each Grade in the Intervention and Non-Intervention Schools

		Intervention				Non-Intervention			
		G3/5[1] (*n*=56)	G5/7[1] (*n*=61)	G7/9[1] (*n*=72)	G9/11[1] (*n*=30)	G3/5[1] (*n*=67)	G5/7[1] (*n*=44)	G7/9[1] (*n*=68)	G9/11[1] (*n*=31)
Total									
Pre	Mean	24.09	37.11	42.31	54.40	21.64	36.50	49.94	62.16
	SE	1.246	1.189	2.096	3.198	1.073	1.533	2.034	2.763
Long	Mean	35.14	38.11	58.63	59.03	32.01	35.48	62.68	70.06
	SE	1.066	1.620	2.672	4.114	0.876	1.746	2.229	2.951
	t	-10.65	-0.78	-8.34	-1.67	-10.54	0.81	-6.55	-2.80
	p	*p*<.0001	NS	*p*<.0001	NS	*p*<.0001	NS	*p*<.0001	*p*<.005
Basic Chance and Data (BCD)									
Pre	Mean	9.27	11.72	12.17	14.27	8.66	11.50	12.69	15.71
	SE	0.351	0.246	0.578	0.856	0.342	0.354	0.594	0.640
Long	Mean	12.47	11.44	14.61	15.50	11.27	10.57	16.15	17.48
	SE	0.259	0.440	0.563	0.849	0.313	0.518	0.442	0.709
	t	-9.92	0.67	-4.40	-1.34	-8.12	2.47	-5.92	-2.22
	p	*p*<.0001	NS	*p*<.0001	NS	*p*<.0001	*p*<.009	*p*<.0001	*p*<.02
Variation in Chance (CV)									
Pre	Mean	5.95	6.67	7.94	9.47	4.31	7.00	9.88	9.94
	SE	0.429	0.365	0.460	0.617	0.387	0.464	0.328	0.554
Long	Mean	7.21	7.28	10.68	10.80	6.87	8.36	11.74	11.55
	SE	0.358	0.322	0.539	0.873	0.303	0.414	0.442	0.412
	t	-3.08	-1.73	-3.86	-1.25	-6.94	-3.09	-3.36	-2.34
	p	*p*<.002	*p*<.05	*p*<.0001	NS	*p*<.0001	*p*<.002	*p*<.0006	*p*<.02
Variation in Data and Graphing (DV)									
Pre	Mean	4.77	9.70	15.81	19.17	4.55	8.98	19.22	21.65
	SE	0.299	0.407	0.775	1.285	0.268	0.478	0.754	1.216
Long	Mean	5.75	9.54	21.19	19.87	6.13	8.82	22.54	23.74
	SE	0.220	0.513	1.038	1.452	0.190	0.513	0.818	1.215
	t	-3.03	0.34	-6.37	-0.63	-4.97	0.298	-4.44	-1.69
	p	*p*<.002	NS	*p*<.0001	NS	*p*<.0001	NS	*p*<.0001	NS
Variation in Sampling (SV)									
Pre	Mean	5.50	9.02	6.89	12.47	4.12	9.02	9.07	16.16
	SE	0.517	0.555	0.673	1.220	0.393	0.644	0.835	1.038
Long	Mean	9.73	9.85	13.28	14.10	7.75	8.52	13.76	18.87
	SE	0.582	0.712	0.891	1.520	0.442	0.906	0.839	1.286
	t	-7.46	-1.21	-7.56	-1.45	-7.08	0.65	-5.72	-2.12
	p	*p*<.0001	NS	*p*<.0001	NS	*p*<.0001	NS	*p*<.0001	*p*<.03

[1]Grade in the longitudinal follow-up

For Grades 3 and 7 in both groups, the pattern of performance for the Total score was repeated for each grade for the four subscales, indicating an even development across the topics included as part of the testing. For the Grade 9 students in the non-intervention schools the relatively smaller increase in performance was reflected in three of the four subscales, whereas for the intervention Grade 9 students, although all means increased, none reached a level of statistical significance. Grade 5 students showed the least consistent performance, with the intervention group improving from the pre-test on the CV subscale but none of the others, whereas the non-intervention group also did better on the longitudinal CV subscale but did worse on the longitudinal BCD subscale.

Although it was expected for there to be some significant improvements over time from the pre-test to the longitudinal follow-up for both the intervention schools and the non-intervention schools, it was not necessarily expected that the improvement achieved by the intervention schools would be greater than the improvement for the non-intervention schools. The only difference in the *change* of performance between the intervention and non-intervention schools for each of the four subscales and overall (Total), as well as for each grade, was in relation to the CV Subscale where the non-intervention schools improved more than the intervention schools did for students originally in Grade 3 (5.44, $p<.03$). The Bonferroni correction suggests this is likely to be a chance outcome.

Change within School Types over Two-Years

Table 3 shows the means and standard errors for each of the subscales for the same grade two years apart in both the intervention and non-intervention schools. To clarify, in the intervention schools the students in Grade 5 for the longitudinal follow-up in 2002 were originally in Grade 3 for the pre-test in 2000, and hence had taken part in the intervention instruction, whereas the students with whom they are being compared, who were originally in Grade 5 in 2000, had not. For the non-intervention schools, no students took part in the instruction, so the comparison is between two groups of students in the same grade, two years apart, both without intervention. Comparisons are made on the basis of students having answered the same questions; for example, Grade 5 students in Grade 7 in 2002 answered the same items as Grade 7 in 2000. The sample sizes for each grade in 2000 and in 2002 are reported with the Total score.

As can be seen from Table 3, for the intervention schools, the Grade 7 students in the longitudinal follow-up performed at a significantly higher level than the original Grade 7 students in 2000 on the overall survey (Total). There was no significant overall increase in performance (Total) within grades for the non-intervention schools. In relation to the BCD subscale only the Grade 5 students in the intervention schools in 2002 performed better than their earlier counterparts, whereas, in the CV subscale there was an increase in performance for the Grade 7 students in the intervention schools and improvement for the Grade 9 students in the non-intervention schools. When compared to the non-intervention schools, the Grade 7 students in the intervention schools in the CV subscale had a lower mean in 2000, perhaps leading to the expectation of greater improvement than the non-intervention school Grade 7 students.

For the DV subscale, there was no significant increase in performance between the grades two years apart in either condition, however, there was a significant decrease in performance in the Grade 7 students in the non-intervention schools with the mean score dropping almost three marks between the 2000 Grade 7 students and those who were in Grade 7 in the 2002. Finally, the greatest improvement in performance within intervention schools occurred in the SV subscale where both Grade 5 and Grade 7 students in 2002 performed better than students in the same grade in 2000. There was no increase for grades after two years for the non-intervention schools for this subscale. Out of the thirty comparisons, five were positively significant for the intervention schools, compared to only one for the non-intervention schools.

Table 3
Means and Standard Errors for the Total and Subscales for Each Grade Two Years Apart[1]

	Intervention			Non-Intervention		
	Pre-test (2000) (Mean, SE)	Longitudinal (2002) (Mean, SE)	*t, p*	Pre-test (2000) (Mean, SE)	Longitudinal (2002) (Mean, SE)	*t, p*
			Total			
Grade 5	37.11, 1.189 (*n*=61)	40.07, 1.340 (*n*=56)	-1.66, NS	36.50, 1.533 (*n*=44)	35.97, 1.034 (*n*=67)	0.30, NS
Grade 7	42.31, 2.096 (*n*=72)	50.33, 2.348 (*n*=61)	-2.56, $p<.006$	49.94, 2.034 (*n*=68)	47.61, 2.764 (*n*=36)	0.69, NS
Grade 9	54.40, 3.198 (*n*=30)	59.61, 2.714 (*n*=72)	-1.11, NS	62.16, 2.762 (*n*=31)	64.00, 2.280 (*n*=68)	-0.48, NS
			Basic Chance and Data (BCD)			
Grade 5	11.72, 0.246	12.45, 0.254	-2.05, $p<.03$	11.50, 0.354	11.27, 0.313	0.48, NS
Grade 7	12.17, 0.578	13.33, 0.534	-1.46, NS	12.69, 0.594	12.45, 0.662	0.26, NS
Grade 9	14.27, 0.856	14.61, 0.563	0.86, NS	15.71, 0.640	16.15, 0.522	-0.49, NS
			Variation in Chance (CV)			
Grade 5	6.67, 0.365	7.21, 0.358	-1.06, NS	7.00, 0.464	6.87, 0.303	0.25, NS
Grade 7	7.94, 0.460	9.18, 0.446	-1.91, $p<.03$	9.88, 0.328	9.50, 0.565	-0.30, NS
Grade 9	9.47, 0.617	10.68, 0.539	-1.31, NS	9.94, 0.554	11.74, 0.442	-2.38, $p<.01$
			Variation in Data and Graphing (DV)			
Grade 5	9.70, 0.407	9.52, 0.459	0.31, NS	8.98, 0.478	9.21, 0.304	-0.43, NS
Grade 7	15.81, 0.775	16.97, 0.881	-0.99, NS	19.22, 0.754	16.32, 1.011	2.34, $p<.02$
Grade 9	19.17, 1.285	21.19, 1.038	-1.12, NS	21.65, 1.216	22.54, 0.818	-0.61, NS
			Variation in Sampling (SV)			
Grade 5	9.02, 0.555	10.89, 0.655	-2.20, $p<.02$	9.02, 0.644	8.63, 0.511	0.48, NS
Grade 7	6.89, 0.673	10.85, 0.815	-3.78, $p<.0002$	9.07, 0.835	9.16, 1.003	-0.07, NS
Grade 9	12.47, 1.220	14.26, 0.930	-1.09, NS	16.16, 1.038	15.09, 0.901	0.71, NS

[1]Some mean values change in this table because more questions were answered in some higher grades

Discussion

The discussion focuses on the various educational issues surrounding the results reported over two years for the intervention and non-intervention schools.

For the intervention schools, Grade 3 and 7 students retained improvement or improved further on results from the post-test that were reported in Watson and Kelly (2002a, 2002c) six weeks after instruction. Over the same two-year period, students in the non-intervention schools showed similarly significant improvement. For Grade 5 students in the intervention schools, improvement seen in the post-test for the Total score, DV, and SV subscales was not retained in the two-year follow-up (Watson & Kelly, 2002b). In the non-intervention schools a similar lack of improvement was observed for most scales for Grade 5. The largest disparity for intervention and non-intervention schools occurred for Grade 9 students where for the subgroup continuing to Grade 11, the non-intervention schools showed significant improvement whereas the intervention schools lost ground compared to the post-test outcomes for their larger cohort of students in the original year (Watson & Kelly, 2002c).

The data indicate that there can be no reason to claim that the intervention units on chance and data had any long-term effect that was significantly better in terms of survey results than the ordinary mathematics program at the non-intervention schools. Possible

explanations for this outcome include the following: (1) It may be that having had an intervention specifically aimed at chance and data in the first year of the study, the intervention schools neglected these two topics over the next two years. (2) It is known that several teachers in the non-intervention high schools attended Quality Teacher Programs, including sessions on chance and data led by the first author. Anecdotal evidence suggests that these high schools included lessons on units of work similar to those used in the intervention schools at some time between the pre-test and longitudinal follow-up. This might explain the improvement for Grades 7 and 9. (3) It may be that the strength of the intervention was not great enough to be sustained without further reinforcement.

The overall lack of sustained improvement for Grade 5 students in the intervention schools and the regression for this same grade in non-intervention schools suggests that other factors may be playing a role. Evidence from other research is accumulating about the lack of educational progress made by students at the middle school level (e.g., Hill, Rowe, Holmes-Smith, & Russell, 1996; Callingham & McIntosh, 2002). All Grade 5 students in this study changed from a local primary to a local high school at the end of Grade 6 and were in the new school for about six months before taking the longitudinal survey. One suspects that the data from this study add to the results of other studies suggesting concern and the need for further research on middle school students' outcomes. For Grade 9 students, except for the relatively smaller sample sizes, the authors have no explanation for the lack of sustained change for the intervention students, especially since the non-intervention school students improved over the two years. When difference scores were considered, however, there was not a significant difference between the two groups.

In considering difference scores for intervention and non-intervention schools, changes were generally comparable. In considering differences of students in Grades 5, 7, and 9 two years apart there was some indication that for Grade 5 and 7, there may have been some benefit in being involved in the intervention program in Grades 3 and 5 respectively. Of 10 comparisons of mean scores for these grades, all except two were in the expected direction if the intervention had had an influence and five of these differences were significant at the 0.03-level or less. In the non-intervention schools there were two differences at less than the 0.02-level, one in each direction, with no indication of an overall difference for the two cohorts. As there is no other way of comparing the cohorts, except that they are from the same schools, this is a very tentative outcome but at least it is encouraging in terms of the intervention implemented in those schools.

Although the Grade 5 students in the intervention schools did not improve overall longitudinally, the increased level of performance of these same students in Grade 7 in 2002 when compared to the original Grade 7 students in 2000 on the CV subscale is perhaps not surprising as there was in fact a significant improvement longitudinally on this subscale. When surveyed two years later in 2002, this group of students (originally in Grade 5) not only performed better than the 2000 Grade 7 students, but performed almost as well as the 2000 Grade 9 students (see Tables 2 and 3). A similar phenomenon occurred in the SV subscale, where there was a significant increase in performance when comparing Grade 5 students two years later in Grade 7 with the students originally in Grade 7 but this time there was no improvement longitudinally for the Grade 5 students. This result is more than likely an artefact of the poor performance of the original Grade 7 students on this subscale before the intervention (see Table 2).

The outcomes of this study suggest that although it is possible to intervene in classrooms and expect short-term positive outcomes for students, ensuring long-term sustained performance is likely to require continued reinforcement. As chance and data are

still relatively new parts of the mathematics curriculum, reinforcement and appreciation across the curriculum may not yet be occurring. One might hope that as quantitative literacy achieves a higher profile across the school curriculum, sustained and continued improvement will be observed in relation to the objectives of this study.

Acknowledgement. This research was funded by an Australian Research Council grant (No. A00000716).

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: Author.

Callingham, R., & McIntosh, A. (2002). Mental computation competence across years 3 to 10. In B. Barton, K. C. Irwin, M. Pfannkuch & M. O. J. Thomas, (Eds.), *Mathematics education in the South Pacific: Proceedings of the 25th annual conference of the Mathematics Education Research Group of Australasia, Auckland* (pp. 155-163). Sydney: MERGA.

Hill, P. W., Rowe, K. J., Holmes-Smith, P., & Russell, V. J. (1996). *The Victorian Quality Schools Project: A study of school and teacher effectiveness. Report (Volume 1).* Melbourne: Centre for Applied Educational Research, University of Melbourne.

Kelly, B. A., & Watson, J. M. (2002). Variation in a chance sampling setting: The lollies task. 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* (pp. 366-373). Sydney: MERGA.

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*: *Proceeding of the 20th Annual Conference for the Mathematics Education Research Group of Australasia* (pp. 6-12). Rotorua, 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.

Watson, J. M. (1978). A current event for the mathematics classroom. *Mathematics Teacher, 71*, 658-663.

Watson, J. M. (1992). What research is needed in probability and statistics education in Australia in the 1990s? In B. Southwell, B. Perry, & K. Owens (Eds.), *Proceedings of the Fifteenth Annual Conference of the Mathematics Education Research Group of Australasia* (pp. 556-567). Kingswood, NSW: MERGA.

Watson, J. M., & Kelly, B. A. (2002a). Can grade 3 students learn about variation? In B. Phillips (Ed.), *Developing a statistically literate society*: *Proceedings of the Sixth International Conference on Teaching Statistics, Cape Town.* 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* (pp. 682-689). Sydney: MERGA.

Watson, J. M., & Kelly, B. A. (in press). Expectation versus variation: Students' decision making in a chance environment. *Canadian Journal of Science, Mathematics and Technology Education.*

Watson, J. M., Kelly, B. A., Callingham, R. A., & Shaughnessy, J. M. (2003). The measurement of school students' understanding of statistical variation. *International Journal of Mathematical Education in Science and Technology, 34*, 1-29.