CHAPTER 7

RESULTS - STUDY 2

This chapter presents the results of the second study and answers research questions raised earlier. There are four sections in the chapter. The first three, **Luria Study, Longitudinal Study** and **Interview Study** outline the results of the data analysis for the three parts of the second study. Conclusions concerning the research questions are presented also. The final section, **Overview - Study 2**, brings together the conclusions from the three parts of Study 2 and links these to the conclusions for Study 1.

Luria Study

Coding of responses in both parts of Questions 2, 3 and 4 in Study 1 showed a definite split into two different paths. These paths, labelled A and B, were observed in each of the Levels 5 through to 7. The Luria Study was designed to determine whether the different processing styles, successive and simultaneous, influenced the tendency to respond along the A or B path. A battery of six standard Luria Test items were used to measure the processing style, and these results were used to attempt to explain the choice of path made by the students in their responses to Study 1. This section presents the results of the Luria Study, and then draws conclusions from these results addressing the research question concerning a possible link between the path of a response and the processing style of the student.

Results

The analysis consisted of four components. First, the Luria scores were checked to determine whether the tests were a reliable measure. Second, the Luria scores were compared to the Rasch scores to assess the influence of processing style on statistical understanding. Next, the Luria scores were compared to Path scores to determine whether the processing style could help to explain the path selected. Finally, the possibility of a relationship between processing style and path selected were investigated separately for each question.

The combined results from the battery of six tests produced a successive score and a simultaneous score for each student. The successive scores ranged from -3.49 to 2.25 and the simultaneous scores ranged -2.37 to 2.85, the implication being that the higher the score the more tendency a student had to use that style of processing. A high simultaneous score did not automatically imply a low successive score. In fact, some students were low on both measures and some high on both.

To check on the reliability of the tests, as a measure of the two processing styles, a factor analysis was applied to the scores. The contributions of each of the six tests to the two factors, successive and simultaneous processing, are shown in Table 7.1.1. As any factor with a contribution of more than 0.75 is reliable, Tests 2 and 5 are the best indicators for simultaneous processing and Test 6 is the best for successive processing, with Tests 3 and 4 also being useful. Test 1, designed to measure attention span, loaded slightly towards successive processing. These results demonstrate that the tests have produced reliable measures of the two styles of processing as indicated in the research instrument description in Chapter 6.

Table 7.1.1 - Factor Analysis of Luria Test Scores

	Successive	Simultaneous
1 Number/Letter Search Test	0.54	0.48
2 Form Board	0.19	0.84
3 Auditory Word Span	0.78	0.26
4 Number/Letter Attention Span	0.73	0.13
5 Paper Folding	-0.05	0.80
6 Auditory Number Span	0.84	-0.18

To test whether statistical understanding could be explained by considering processing style, the Rasch score was regressed on the successive and simultaneous factors. The analysis of variance for the regression analysis is shown in Table 7.1.2. The result was highly significant ($p < 0.001$) indicating that the two factors accounted for a significant proportion of the variation in student understanding. Hence, it can be concluded that variation in understanding does depend on processing style. More specifically, both variables made a significant contribution to the regression. The coefficient for the successive variable being 0.08 which is significant at the $\alpha = 0.05$ level and for the simultaneous variable being 0.15 which is significant at $\alpha = 0.01$. In fact, 15% of the Rasch score could be accounted for by the successive and simultaneous scores for the students, indicating that these partially accounted for the level of understanding of students. This is consistent with research in other areas which has found that the preference for simultaneous or successive styles of processing can influence performance on mathematical tasks. These results were encouraging and reinforced the worth of checking whether the preference for Path A or B responses was influenced by processing style.

Table 7.1.2 - Regression of Rasch Score on Successive and Simultaneous Factors

Source	D.F.	S.S.	M.S.	F-Test
Regression	2	2.79	1.40	8.306
Residual	94	15.79	0.17	p=0.00
Total	**96**	**18.59**		

To test whether the preferred path was influenced by the processing style, the Path score, a measure for the preference of Path A over Path B, was regressed on the processing style. The analysis of variance produced a result which was close to being significant ($p = 0.06$) for the regression analysis. Details are reported in Table 7.1.3. The coefficient of the successive variable in the regression, -0.07, is not significant ($p > 0.50$) but the coefficient for the simultaneous variable, 0.27, is significant ($p < 0.05$). The different signs of the coefficients indicate that increased simultaneous processing tendencies increased the preference for Path A, but increased successive processing tendencies reduced the preference for Path A, making Path B preferable.

Table 7.1.3 - Regression of Path Score on Successive and Simultaneous Factors

Source	D.F.	S.S.	M.S.	F-Test
Regression	2	7.34	3.67	2.86
Residual	94	120.57	1.28	p=0.06
Total	**96**	**127.90**		

This result prompted a consideration of separate regressions of the Path on each of the two variables. The analysis of variance for regression on the successive variable is shown in Table 7.1.4 and for the simultaneous variable in Table 7.1.5. The non-significant result for successive processing indicates that this variable is of little use in explaining the path chosen by the student. However, the simultaneous processing regression is significant ($p < 0.05$) indicating that the variable accounts for 5% of the variation in path selection.

Table 7.1.4 - Regression of Path Score on Successive Factor

Source	D.F.	S.S.	M.S.	F-Test
Regression	1	0.44	0.44	0.325
Residual	95	127.46	1.34	p=0.57
Total	**96**	**127.90**		

Table 7.1.5 - Regression of Path Score on Simultaneous Factor

Source	D.F.	S.S.	M.S.	F-Test
Regression	1	6.90	6.90	5.42
Residual	95	121.00	1.27	p=0.02
Total	**96**	**127.90**		

Considering this relationship in more detail, the paths for individual questions were analyzed to determine whether there was a link with successive and simultaneous processing. Students were designated as low, middle or high for each of the processing styles, as outlined in the design description in Chapter 6. The paths, arranged by processing level, for each question appear in Appendix O. Frequencies for successive processing across each of the questions are in Table O.1.1 and for simultaneous processing in Table O.1.2.

Results of the chi-squared tests for the contingency tables appear in Tables 7.1.6 and 7.1.7, for successive and simultaneous processing, respectively. When testing for an association between the path chosen and successive processing, all chi-squared tests gave a non-significant result indicating that there was no association for any of the questions. When the simultaneous processing tables were tested, all results were also not significant, except for Question 2 Part I, where a significant result ($p < 0.01$) indicates that the students with higher simultaneous processing style were more inclined to respond along Path A. Thus, the association suggested earlier between simultaneous processing style and a preference for Path A is most evident in the data tabulation and representation focus area.

Table 7.1.6 - Paths by Successive Processing

		χ^2	D.F.	Sig.
Question 2	**Part I**	1.00	2	N.S.
	Part II	0.81	2	N.S.
Question 3	**Part I**	expected numbers too small to test		
	Part II	0.16	2	N.S.
Question 4	**Part I**	1.36	2	N.S.
	Part II	2.46	2	N.S.

N.S. : $p > 0.05$; * : $p < 0.05$; ** : $p < 0.01$; *** : $p < 0.001$

Table 7.1.7 - Paths by Simultaneous Processing

		χ^2	D.F.	Sig.
Question 2	**Part I**	9.26	2	**
	Part II	2.84	2	N.S.
Question 3	**Part I**	expected numbers too small to test		
	Part II	0.17	2	N.S.
Question 4	**Part I**	0.51	2	N.S.
	Part II	0.38	2	N.S.

N.S. : $p > 0.05$; * : $p < 0.05$; ** : $p < 0.01$; *** : $p < 0.001$

Conclusion

Although the level of understanding is partially accounted for by the style of processing, there is less evidence that the preference for successive or simultaneous processing style provides a satisfactory explanation for the divergence of responses along the

A and B paths. Some of the tendency to prefer Path A to Path B can be explained by the preference for simultaneous style processing, this being most evident in the area of data tabulation and representation. Further investigation into student preference for particular processing paths is needed to seek other explanations for the divergence into separate paths.

Longitudinal Study

The Longitudinal Study was designed to verify the results observed and the conclusions drawn in Study 1 and to investigate any possible growth in student understanding over the twelve-month period. The results of these analyses are presented, preceded by the results of the validation of the 1992 test instrument.

Validation of Test Instrument

Two aspects of the test needed to be validated, namely: the performance of questions as measures of statistical understanding; and, the adequacy of the fit of the Rasch model to the data. Second, the degree to which the 1992 test was capable of measuring the levels and steps between them, compared to the 1991 test. As with the written test for Study 1, the performance of the test was checked using item consistency and the infit mean square map for parameter fit. The item consistency for the test analysis was 0.7 which, as before, is considered sufficiently high (Vine, 1994) to assume that the items were consistent in testing the same construct, statistical understanding. The fit statistics for the question and student estimates appear in Figures P.1 and P.2, respectively (Appendix P). Both infit mean squares are not significant, indicating that the data fitted the Rasch model well. The infit mean square map for the four questions is shown in Figure P.3. Although the spread on mean squares is greater than for Study 1, all questions showed values within the acceptable limits except Question 3 Part I. This indicates that the parameter for Question 3 Part I is not as useful as the other questions, whose means lie within the boundaries, in the model used to calculate the estimate of understanding. These results were considered to be sufficient to allow the use of the 1992 test as a measurement instrument.

As mentioned in the design chapter, the written test used in the second study was similar to that used in the first, with some minor changes being made to the wording of the questions. To determine whether these changes altered the performance of the test in terms of the allocation of responses to the various levels of the hierarchy, the tau values and threshold values were compared. The tau values for 1991 and 1992 are presented in Table O.2. In testing whether there was a relationship between the two years, the correlation was 0.51 giving $t = 4.04$ (46 d.f.). This indicates strong evidence ($p < 0.001$) for a relationship between the tau values for each of the two tests. The positive correlation indicating that higher tau values (difficulties) are associated in each of the two years as expected. Hence, the difficulty experienced by any student in attaining each level of the hierarchy is relatively similar for the two tests.

Threshold values for 1991 and 1992 are presented in Table O.3. Testing for a relationship between the two years, the correlation was 0.96 giving $t = 23.3$ (46 d.f.) which represents strong evidence ($p < 0.001$) for a relationship between the threshold values for each of the two tests. The positive correlation indicates that higher threshold values are associated in each of the two parts, again, as expected. Consideration of the performance of each question separately gave significant correlations of 0.96, 0.98, 0.98 and 0.99 for Questions 1 to 4, respectively. In fact, all these correlations are highly significant ($p < 0.001$), indicating comparable performances of the two tests in 1991 and 1992, in terms of the level of understanding needed to attain successive levels of the hierarchy.

Overall, these results were considered strong enough to be able to conclude that the test was suitable as a measure of statistical understanding, and despite the minor alterations made to the test, the overall performance of the 1992 written test was similar to that of the 1991 written test. Having validated the test instrument it was possible to use the 1992 results to attempt to confirm Study 1 results and to further investigate students' growth in understanding.

Results

It was interesting to note that not only did some students not remember doing the test, twelve months previously, some actually denied that they had participated until shown their Study 1 responses. This is surprising given the different nature of the test compared to the routine mathematical tasks experienced at school. Nevertheless, the hierarchy developed in Study 1 proved to be sufficient to classify all responses which were supplied by the students in the Longitudinal Study. In this section, the results of the allocation of the responses for Study 2 are compared to those observed in Study 1. Then, the results of the analysis of the changes in individual student performances are presented.

Validation of Study 1 Results

Firstly, the results for the individual questions are compared for Study 1 and Study 2 and then, the assessment of overall understanding are compared for the two studies. The analysis of Study 1 included comparisons of performance on each question with varying academic year, mathematical ability and gender, the comparison of the two parts for each question and an analysis of the Rasch scores of overall understanding. The analysis of the Study 2 results mirrored this structure.

The levels of ranking of responses, and path, where appropriate, for each question arranged by academic year (Tables O.4.1 to O.4.8), mathematical ability (Tables O.5.1. to O.5.8) and gender (Tables O 6.1. to O.6.8) appear in Appendix O. Comparing the results for the various academic years for each question produces similar trends over the two years. The most notable of these being the under-representation of senior students (Years 11 and 12) in the lower levels (Levels 0 to 2), few junior students (Years 7 and 8) in the higher levels (Levels 6, 7 and 8) and bulges of students generally in Levels 3 through to 6, except for Question 2 which has bulges of students at Level 2. The paths chosen by students also showed similar trends in 1992 to those taken in 1991. Question 2 has approximately the same number of students responding along Paths A and B, while Question 3 has approximately four times more Path As than Path Bs and Question 4 has about twice as many Path Bs as Path As.

The results of the chi-squared analyses of levels by mathematical ability for both studies are presented in Table 7.2.1. There were similar levels of significance in 1991 and 1992 for all questions, except Part II of Question 4. These results help to confirm that mathematical ability influences the level of understanding in data reduction and interpretation and inference. There is also an influence, when prompted, in the data collection area. The results of the chi-squared analyses of paths by mathematical ability are presented in Table 7.2.2. The only question for which there is an influence of mathematical ability on the path chosen is in working with graphs in data tabulation and representation, the more able students tending to respond along Path A while the less able respond along Path B.

Table 7.2.1 - Levels by Mathematical Ability

		1991			1992		
		χ^2	D.F.	Sig.	χ^2	D.F.	Sig.
Q.1	**Part I**	14.47	8	N.S.	7.13	6	N.S.
	Part II	42.36	10	***	20.14	6	**
Q.2	**Part I**	7.21	10	N.S.	14.63	8	N.S.
	Part II	12.41	10	N.S.	15.48	8	N.S.
Q.3	**Part I**	22.15	6	**	20.92	4	***
	Part II	21.90	6	**	13.18	2	**
Q.4	**Part I**	26.17	6	***	15.55	6	*
	Part II	20.49	6	**	7.98	6	N.S.

N.S. : $p > 0.05$; * : $p < 0.05$; ** : $p < 0.01$; *** : $p < 0.001$

Table 7.2.2 - Paths by Mathematical Ability

		1991			1992		
		χ^2	D.F.	Sig.	χ^2	D.F.	Sig.
Q.2	Part I	4.85	2	N.S.	2.35	2	N.S.
	Part II	9.60	2	*	7.33	2	*
Q.3	Part I	3.00	2	N.S.	0.50	2	N.S.
	Part II	1.45	2	N.S.	0.82	2	N.S.
Q.4	Part I	0.48	2	N.S.	3.22	2	N.S.
	Part II	4.31	2	N.S.	0.17	2	N.S.

N.S. : $p > 0.05$; * : $p < 0.05$; ** : $p < 0.01$; *** : $p < 0.001$

The results of the chi-squared analyses of levels by gender are similar for the two years and are presented in Table 7.2.3. Although the results are not as consistent as those for mathematical ability, it should be noted that while the 1992 sample still had similar numbers of students in all three levels of mathematical ability, the number of females who chose to participate in the second part of the study was greater than the number of males. This may cause the range of male responses not to be as well represented.

Table 7.2.3 - Levels by Gender

		1991			1992		
		χ^2	D.F.	Sig.	χ^2	D.F.	Sig.
Q.1	Part I	0.13	5	N.S.	2.26	4	N.S.
	Part II	9.92	5	*	8.80	3	*
Q.2	Part I	7.73	6	N.S.	7.01	4	N.S.
	Part II	7.24	5	N.S.	3.32	5	N.S.
Q.3	Part I	3.42	3	N.S.	0.38	3	N.S.
	Part II	10.37	3	*	2.34	3	N.S.
Q.4	Part I	13.12	5	*	1.23	4	N.S.
	Part II	7.12	5	*	4.02	4	N.S.

N.S. : $p > 0.05$; * : $p < 0.05$; ** : $p < 0.01$; *** : $p < 0.001$

The only result significant ($p < 0.05$) for both the years is Question 1 Part II, when students were prompted into giving more information in a sampling situation. It would appear that females have been able to benefit more from this information and produce higher level responses than the males. The 1991 results for Question 3 Part II and Question 4 Part I were significant ($p < 0.05$), with the females showing a higher level of ability in both, although the evidence was not strong. In 1992 the results for both Question 3 Part II and Question 4 Part I were not significant, and so in fact it may be that gender is not an influencing factor in these situation as first thought.

The results of the chi-squared analyses of paths by gender are presented in Table 7.2.4. The paths chosen by the students showed similar results for 1991 and 1992 when analyzed by gender. Gender was consistently found not to influence the path chosen in all questions except Question 4 Part II in 1992 where females showed preference for Path B and males for Path A.

The Study 2 results have, in most instances, confirmed trends observed in Study 1 concerning the influence of academic year, mathematical ability and gender on the level of statistical understanding and the path chosen. Next, the comparisons of Part I and II of each question are considered for the two years. The level of ranking of responses (and path where appropriate) for 1992, arranged by question part, for each question are presented in Tables O.7.1 to O.7.4.

Table 7.2.4 - Paths by Gender

		1991			1992		
		χ^2	D.F.	Sig.	χ^2	D.F.	Sig.
Q.2	**Part I**	1.23	1	N.S.	0.08	1	N.S.
	Part II	0.00	1	N.S.	0.03	1	N.S.
Q.3	**Part I**	1.23	1	N.S.	0.04	1	N.S.
	Part II	0.01	1	N.S.	1.50	1	N.S.
Q.4	**Part I**	1.15	1	N.S.	0.00	1	N.S.
	Part II	3.44	1	N.S.	13.66	1	***

N.S. : $p > 0.05$; * : $p < 0.05$; ** : $p < 0.01$; *** : $p < 0.001$

The results of the chi-squared analyses of levels by question part are presented in Table 7.2.5. Results for Question 1 were the same for both years, highly significant ($p < 0.001$), confirming that when prompted with some information concerning data collection students were able to produce higher level responses. The other three questions, although showing some significance in 1991, were all not significant in 1992. This suggests that perhaps the level of understanding of a student's response does not depend on whether the information is presented in raw form or as a graph. The results of the chi-squared analyses of paths by question part, presented in Table 7.2.6., were not significant for both years confirming that the form of presentation of data does not affect the path chosen when responding to questions dealing with data.

Table 7.2.5 - Levels by Question Part

	1991			1992		
	χ^2	D.F.	Sig.	χ^2	D.F.	Sig.
Question 1	109.50	6	***	93.09	5	***
Question 2	13.53	6	*	6.22	5	N.S.
Question 3	16.62	6	*	6.46	4	N.S.
Question 4	7.94	5	N.S.	1.95	5	N.S.

N.S. : $p > 0.05$; * : $p < 0.05$; ** : $p < 0.01$; *** : $p < 0.001$

Table 7.2.6 - Paths by Question Part

	1991			1992		
	χ^2	D.F.	Sig.	χ^2	D.F.	Sig.
Question 2	2.39	1	N.S.	0.01	1	N.S.
Question 3	3.28	1	N.S.	1.32	1	N.S.
Question 4	0.02	1	N.S.	0.85	1	N.S.

N.S. : $p > 0.05$; * : $p < 0.05$; ** : $p < 0.01$; *** : $p < 0.001$

Finally, the Rasch scores for 1991 and 1992 need to be compared to check whether conclusions made concerning students' overall level of understanding of statistics are confirmed by the 1992 results. The analysis of variance for the Rasch scores for 1992 is shown in Table 7.2.7 and these results are now compared to those for 1991, which appeared in Table 5.1 in Chapter 5. As in 1991, the 1992 tests for the effect of academic year or mathematical ability on the level of understanding are very significant ($p < 0.001$) confirming that both factors have a strong influence on statistical understanding. However, in 1992 the gender factor gave a non significant result compared to a significant result in 1991.

Table 7.2.7 - Estimates of Understanding - Analysis of Variance

Source	S.S.	D.F.	M.S.	F-Ratio	Prob.
Year(Y)	4.451	4	1.113	7.005	0.000
M.Ability(M)	4.820	2	2.410	15.172	0.000
Gender(G)	0.341	1	0.341	2.147	0.147
Y x M	1.266	8	0.158	0.996	0.445
Y x G	2.442	4	0.610	3.842	0.006
M x G	1.480	2	0.740	4.658	0.012
Y x M x G	0.583	8	0.073	0.458	0.882
Error	13.821	87	0.159		

Further investigation showed that there were in fact significant interactions between gender and academic year, and gender and mathematical ability in 1992 and the combined effect of these was to show no significant gender differences between levels of understanding. The 1991 results had shown females to be responding consistently at a higher level than males but the 1992 results showed Year 8 and Year 11 males, and also the males with lower mathematical ability, having a superior performance to that of the females in corresponding years or of comparable mathematical ability. However, these are comparisons where the sample numbers for male and female differ the most, that is by three. In each case, the male number being the lower. This lack of representation of males may partially explain higher performances compared to females than those observed in 1991. Taking this into consideration, a gender effect on statistical understanding may still exist.

The mean estimates for each level of the main factors are presented in Table 7.2.8 and the differences are now compared to those for 1991 which are presented in Table 5.2. The matrix of pairwise mean differences and the matrix of comparison probabilities for the mean differences for each of the three factors academic year, mathematical ability and gender are shown in Tables O.8.1 to O.8.6. Some of the comparisons differ from the results observed in Study 1. For the academic year means, the only significant differences in 1992 were Year 8 being significantly ($p < 0.01$) lower than all other years. In the 1991 comparisons, it was Year 12 who were significantly higher than all other years. Although these results differ in focus, what is most important is that generally, between successive academic years, there is not a significant difference in level of understanding. A number of years are needed before there is an obviously significant increase in understanding.

Table 7.2.8 - Estimates of Understanding - Main Effect Means

Year 8	-0.32
Year 9	0.25
Year 10	0.19
Year 11	0.15
Year 12	0.31
Low Ability	-0.13
Middle Ability	0.12
High Ability	0.36
Female	0.17
Male	0.08

The Study 2 results confirmed a significant difference between students with low and middle mathematical ability as observed in 1991 and found a significant result between the middle and high ability students. This confirms that a relationship exists between the mathematical ability of a student and the level of statistical understanding. Although the mean estimate for females in 1992 was still higher than for males, it was not significantly higher. However, the interaction discussed earlier needs to be taken into consideration. This casts doubt on the significantly higher result observed in 1991, although it must be remembered that the Scheffe comparisons are very conservative and may not detect important differences which occur. The possibility that females achieve at a higher level than males should not be totally rejected.

Again, graphs have been prepared showing the spread of the individual understanding scores and the means. The overall understanding is shown in Figure P.4 while the spread of the estimates for academic year, mathematical ability and gender are shown in Figures P.5, P.6 and P.7, respectively. Comparing these to the spreads of estimates for Study 1, in Figures J.4, J.5 and J.6, highlights some differences. Results for gender are similar in both years in terms of means and spread, as are results for mathematical ability, except for low ability students who show less spread of estimates in 1992. However, the pattern of scores for academic years show different pictures for the two years. Results for

1992 show, in general, a steady increase in understanding as in 1991. However, the gaps between year means are smaller than in 1991 and the Year 9 mean result for 1992 is unexpectedly higher, above the means for Years 10 and 11. Also noticeable is that the spread of scores for Years 10 and 11 are smaller than for other academic years and also smaller than for the 1991 results. The upper tail of the spread is missing which probably contributed to the means for these two years being slightly lower than the previous year's pattern would suggest. These unexpected features may be due to the smaller number of students in each year, as low as 22 (Year 8) and 23 (Years 10, 11 and 12) compared to 30 in 1991. In fact, in Year 12 there were no students of high mathematical ability, male or female, who took part in Study 2.

Generally, the 1992 results and conclusions confirmed those observed in Study 1. Some of the differences observed may be due to the stratification of the sample available for Study 2. The non-significant gender effect and significant gender by mathematical ability interactions for estimates of understanding may be attributed to the slight female-male imbalance in the sample. The apparently superior Year 9 performance may be due to the fact that Year 9 were better represented than any other year. However, results were consistent enough to accept the hierarchy developed in Study 1 and the conclusions drawn.

Longitudinal Comparison

Once the differences between the performance of successive academic years had been observed in Study 1 to be only small, it was considered unlikely that there would be a significant growth of student understanding over one year. However, it was still considered worthwhile to study the change, if any, that had taken place in the understanding of individual students. The correlation of 1991 and 1992 estimates of understanding was 0.61 (115 d.f.) which is very significant ($p < 0.001$). This indicates a link between the estimates for the students for the two years and the positive correlation shows that higher estimates are linked for each year, so that there is a consistency of performance.

Firstly, considering the change in understanding of each student, the mean change was -0.03. There is evidence ($t = -0.76$, 116 d.f., $p > 0.05$) to indicate that over the twelve-month period there does not appear to be an increase or decrease in statistical understanding.

Investigating the differences in more detail, the means and t values for the varying factors academic year, mathematical ability and gender are shown in Tables 7.2.9, 7.2.10 and 7.2.11, respectively. All levels of mathematical ability and both genders showed no significant change in the level of understanding exhibited. However, when the academic years were tested, the students who had progressed from Year 8 to Year 9 showed a significant ($p < 0.05$) increase in the level of understanding. This is not surprising as it is the stage in a student's secondary schooling when more emphasis is placed on data reduction which may have contributed to the increase in understanding.

Table 7.2.9 - Estimate Differences - Academic Years

Year	Mean	t	D.F.	Sig
8	-0.18	-1.53	20	N.S.
9	0.14	2.26	26	*
10	0.07	0.90	22	N.S.
11	-0.12	-1.34	22	N.S.
12	-0.10	-1.01	22	N.S.

N.S. : p>5% ; * : p<5% ; ** : p<1% ; *** : p<0.1%

Table 7.2.10 - Estimate Differences - Mathematical Ability

Math. Ability	Mean	t	D.F.	Sig
Low	-0.04	-0.54	36	N.S.
Middle	-0.07	-1.04	38	N.S.
High	0.02	0.23	40	N.S.

N.S. : p>5% ; * : p<5% ; ** : p<1% ; *** : p<0.1%

Table 7.2.11 - Estimate Differences - Gender

Gender	Mean	t	D.F.	Sig
Female	-0.09	-1.57	61	N.S.
Male	0.04	0.69	54	N.S.

N.S. : p > 0.05 ; * : p < 0.05 ; ** : p < 0.01 ; *** : p < 0.001

The results of the analysis of variance of the differences for the factors academic year, mathematical ability and gender appear in Table 7.2.12. All main effects are not significant, although the academic year factor was almost significant (p = 0.056). This indicates that any change in understanding over the twelve-month period does not depend on a student's academic year, mathematical ability or gender. The only significant interaction is mathematical ability by gender. Consideration of individual mean differences showed that the significant comparison was due to a difference between female and male mean differences for low ability students compared to middle and high ability students. Females showed a greater increase in level for both middle and high ability students, however for low ability students, the females and males showed a decrease in level of understanding, with the female decrease being greater.

Table 7.2.12 - Estimate Differences - Analysis Of Variance

Source	S.S.	D.F.	M.S.	F-Ratio	Prob.
Year(Y)	0.774	4	0.193	2.402	0.056
M.Ability(M)	0.067	2	0.033	0.414	0.662
Gender(G)	0.171	1	0.171	2.118	0.149
Y x M	0.803	8	0.100	1.246	0.283
Y x G	0.682	4	0.170	2.116	0.086
M x G	0.572	2	0.286	3.552	0.033
Y x M x G	0.875	8	0.109	1.358	0.226
Error	7.006	87	0.081		

The Longitudinal Study showed very little growth overall in the level of students' statistical understanding within a twelve-month period and these results were generally consistent irrespective of academic year, mathematical ability and gender. The exception was evidence of some improvement in the movement from Year 8 to Year 9.

Conclusion

The written test proved to be a valid instrument. The fit of the data to the Rasch model suggested that the written test used in Study 2 was a suitable measure of statistical understanding. The performance of this test was comparable to the test used in Study 1, despite minor alterations made. So it was possible to use the results of the test to validate Study 1 conclusions and to consider students' growth in understanding over time.

Concerning the query as to whether similar conclusions would be reached were the Study 1 test to be repeated at another time, as presented in research question Q2.9 in Chapter 2, most conclusions reached in Study 1 were reinforced by the Study 2 results. The performance of academic years was similar. The Year 8 results, rather than the Year 12 results as in the previous year, were significantly different to the results of all other years. This indicates that, within the context of the N.S.W. Syllabus, a difference of one academic year generally does not mean a difference in level of statistical understanding. The dependence on mathematical ability in the areas of data reduction and interpretation and inference was similar for both years. However, the Study 2 results questioned the dependence of performance on gender in data reduction and interpretation and inference, as had been observed in Study 1. The lack of dependence of the choice of path on the factors, academic year, mathematical ability and gender, was also confirmed. The benefits of prompting were again shown to depend on both mathematical ability and gender. However, Study 2 did not confirm the Study 1 indications that the presentation of data in raw form or as a graph influences the level of the response.

Considering the final research question, Q2.12, posed in Chapter 2, concerning growth in understanding over time, there proved to be little growth in the twelve-month period. The only significant growth took place in those students making the transition from Year 8 to Year 9. There was also a possibility that low ability females may be in danger of making a significantly lower increase in understanding than the low ability males over the twelve-month period. To study the growth of individual student's understanding, it would be necessary to conduct research over a much longer period than twelve months.

Interview Study

To gain more insight into results from the Longitudinal Study students were interviewed to investigate the skills of reading, comprehension and interpretation as well as the effects of probing and prompting. Time constraints restricted the number of students that could be interviewed but the results of the interview sessions were useful in supplementing earlier results. This section presents the results of the interviews and conclusions drawn.

Results

The analysis of the taped interview sessions was undertaken in four stages. First, any difficulties that arose in terms of reading, comprehension or interpretation are presented. Next, the differences between the interview responses and the written test responses are considered. Then, any changes in level which occurred as a result of probing are highlighted, and, finally, the changes which occurred as a result of prompting are discussed.

The interviews indicated that there were few problems with reading, comprehension and interpretation. During the written test, students were given the chance to ask about things not understood but no enquiries were made. Similarly, in the interview situation, students appeared to read and comprehend questions successfully. There was only one apparent interpretation problem. A Year 12 student misread the graph in Question 3 Part II, but the style of answer given indicated that had the graph been read successfully the answer would have been at the same level as for Part I.

The differences in levels between written test responses and interview responses are shown in Table 7.3.1. Most students responded at the same level or showed an increase in level. When comparing written test responses to initial interview responses some students gave different answers but at the same level. When asked to choose between these, most chose the interview response. In a few questions, the answer chosen was a combination of both responses or a decision could not be made at all. Whenever there was a difference in level between the interview and written responses, and students were asked to choose between them, the choice made was the higher level responses in all cases except in one question a low mathematical ability male Year 9 student chose the lower level. This suggests that generally students are able to select the better answer when given a choice between two answers at different levels.

Table 7.3.1 - Response Level Changes - Written Test versus Interview

		Decrease	No Change	Increase
Question 1	**Part I**	3	9	3
	Part II	2	10	3
Question 2	**Part I**	3	10	2
	Part II	4	4	7
Question 3	**Part I**	1	11	3
	Part II	0	9	6
Question 4	**Part I**	1	11	3
	Part II	3	5	7

More increases took place in Part II of the questions and this may be due to a confounding effect. It became apparent during the interviews that some students were able to provide a better response to Part II of a question because of probing and prompting which had improved the Part I answer. One such student was a high mathematical ability Year 8 female student who answered Question 2 Part I at Level 2 in the interview, even when probed but managed an answer at Level 4 when prompted. Then Part II, which had also been answered at Level 2 in the written test, was answered at Level 3 in the interview and then, with only probing, a Level 4 response was attained. No further increase was gained with prompting. This beneficial influence of feedback from working with responses does not occur during the written test. However, it is likely that if a student is interviewed, with no probing or prompting, then the level of response which is achieved is likely to be no better than in a written situation.

Most students added more to the initial response when probed and prompted, but in many instances this gave a more detailed answer at the same level rather than increasing the level of the response. For example, a middle mathematical ability male Year 12 student, when answering Question 3 Part I, chose an estimate for the left hand size and gave as the reason "it's the most popular" and under probing kept giving more responses but could only add that it was "the mode" and "the most common". The student was only expressing the same idea in different ways and could not go beyond the level of this response. Although low sample numbers prevented conclusions being drawn concerning a gender influence on the quality of responses, it was noticed that generally the male students were inclined to give more detail in the oral response than a written one. Results in the written test suggesting males demonstrated a lower level of understanding may have been influenced by their reluctance to produce a written answer to an open-format question.

The changes in level due to probing are shown in Table 7.3.2. The number of students who increased the level of response when probed was less than a third for all questions, except one, suggesting that many students attain the highest possible level of response in the initial attempt at an answer.

Table 7.3.2 - Response Level Changes due to Probing

		No Change	Increase
Question 1	**Part I**	12	3
	Part II	15	0
Question 2	**Part I**	10	5
	Part II	10	5
Question 3	**Part I**	12	3
	Part II	12	3
Question 4	**Part I**	9	6
	Part II	11	4

The level changes due to prompting, presented in Table 7.3.3, show that this had more of an influence on improving the quality of the answer than probing. If the student responded positively to prompting, then prompting was continued until resistance was met. However, if there was no response to first attempts at prompting then the question was left. Some students actually began to give responses at a lower level under prompting, as though they felt compelled to give more. These situations were recorded as no change, as initially under prompting the response was unchanged, and it was only when forced that students resorted to lower level responses. One male Year 10 student with low mathematical ability answered Question 2 Part I with a simple subjective opinion on the data and did no better under probing. Then, when prompted he felt pressured to give more and resorted to irrelevant discussion concerning boys riding bikes because girls need time to comb their hair. Eventually, he discussed his own bike-riding habits.

In Question 1, the number who increased the level of response under prompting is only small. This could possibly be due to the fact that not much data collection work is studied at school and for many students these were new ideas. Also, in the focus area of data reduction, there is not much improvement in level of response with prompting. This is the area most studied at school so students are most likely to feel confident about presenting ideas without too much prompting. Prompting helped most in the areas of data representation and tabulation, and interpretation and inference.

Table 7.3.3 - Response Level Changes due to Prompting

		No Change	Increase
Question 1	Part I	10	5
	Part II	12	3
Question 2	Part I	7	8
	Part II	10	5
Question 3	Part I	13	2
	Part II	11	4
Question 4	Part I	8	7
	Part II	10	5

An unexpected and noticeable effect of prompting was a path change in the new response. Changes from Path A to Path B responses, or Path B to Path A responses, indicated that the students have the facility available to respond along either path but will use their preferred path unless pressured to delve further. One such student is a Year 9 girl with high mathematical ability who answered Question 2 Part I at Level 5B, making subjective judgements of the data. Under probing she was able to elaborate on these which culminated in a Level 7B response. It was not until she was prompted into explaining how she made such judgements that she responded using more statistical assessments of the data, giving a Level 7A response, and so indicating that both paths of reasoning were now available for responding. Eventually, some students felt confident enough that both paths were combined in the response, thus attaining a higher Level 7. A number of students made path changes but they were all during prompting, no path changes occurred when students were probed. This indicates that students need encouragement to make use of the path which is not natural for them so that they are able to present a fuller view of data analysis.

It was noticeable during the interviewing that, with prompting, students could be encouraged to make use of data when they had previously not done so. One high mathematical ability male Year 11 student who had only given personal reasons for prediction in the Question 4 Part II written answer used the data in the interview response. When asked which response was better and why he chose the interview answer because "it's got more to do with the graph probably". He had finally come to realize the need to refer to data to justify answers.

Conclusion

The interviews proved useful for clarifying a number of issues. There were no problems with reading and comprehension and only one with interpretation. In fact, none of these aspects were considered to have influenced the overall results in any significant way. There were two major concerns in the Interview Study. The first concern was whether the written form of assessment could measure understanding accurately, compared to an oral test, as alluded to in research question Q2.10 in Chapter 2. The second concern was the effect of prompting, as queried in research question Q2.11.

As the interview responses were usually at a similar level to those of the written test responses, and probing had little effect on the level of the response, the written test can be considered to be a reasonable form of assessment of the student's level of understanding. Increases occurring in Part II of questions could be attributed to the benefits of feedback from discussion of Part I Probing encouraged some students to add more detail to the response but often this did not lift the level of the response. Prompting did appear to affect the response, sometimes increasing the level, occasionally reducing the level and at times changing the path. Some students were able to increase the level of their response with a prompt indicating that they are probably in a transitional stage where they are confident with the level they are working at, and ready to experiment with new ideas. Other students, however, resisted all attempts to increase the level of response and some even, feeling the pressure of having to produce more information, reverted to a lower-level response. This is an indication of multimodal functioning as some students responded at a level in the concrete-symbolic mode but then used ikonic level responses when prompted for more information. Under prompting, some students altered the path of their response from A to B or vice versa. This indicates that most students have both paths available but tend to respond along a preferred path unless prompted to make use of the other.

Although only a small number of students were interviewed, the results suggest that a written test is useful for assessing a student's level of understanding. The interview situation, or oral test, is more likely to be useful in assessing students to determine those who are best prepared for progression to the next level of understanding.

Overview - Study 2

Study 2 was designed to investigate an explanation for the divergence of responses into two distinct paths, the robustness of the hierarchy, the growth in understanding over time and the accuracy of the written test compared to the oral.

The preference for a student to process successively or simultaneously affects the level of statistical understanding of a student. However, the processing preference does not have an influence on the path of a response, although the result was close to being significant. Students' preference for successive processing did not appear to be related to the path chosen. However, those who processed simultaneously showed a preference for Path A, this being particularly evident in the data tabulation and representation focus area.

Investigating the robustness of the hierarchy was necessary to determine whether the levels of the hierarchy and the influence of any factors related to the level of understanding would be the same, irrespective of the chronological year of the measurement. The hierarchy developed in Study 1 was found to be sufficient to code all responses observed. Understanding was influenced by the academic year and mathematical ability of the student. Increased academic year or mathematical ability produced higher level responses. There were only slight differences in level of understanding between academic years with a steady increase in level with increasing academic year, as observed in Study 1. Unusually high results, relative to other years, were observed for Year 9 in Study 2 but this may be accounted for by higher student numbers in the sample in that Year than any other. However, both the academic year and the mathematical ability of a student did not influence the path of the response. There was also a gender influence, with females achieving higher levels than males, although this was not consistent across all levels of mathematical ability. A student's gender did not influence the path of the response, though. Provided a student was able to understand a graph, the level of a student's understanding in other areas did not depend greatly on whether the data are presented in raw form or as a graph.

Over a period of a year students did not undergo a significant change in level of statistical understanding. This observation applies irrespective of mathematical ability or gender. However, there were slight differences depending on academic year with those students progressing from Year 8 to Year 9 showing a significant increase.

As a measurement of understanding, the written test was sufficient, as students provide similar level responses in an interview situation, even with probing. An important difference with the interview situation is the benefit to students of immediate feedback which has helped some students to produce higher level responses than those of which they may be normally capable. An advantage of oral testing or interviewing was that many students, under prompting, were able to improve the level of the response and also to change response path. This could be a useful indicator of whether a student is ready to progress to the next level in the hierarchy.

The results of Study 2 verified and extended the results of Study 1, as well as bringing to light other interesting observations. A discussion of the observed results and conclusions, recent research developments and the implications of this study follows in the next chapter.