CHAPTER 8

DISCUSSION OF RESULTS AND CONCLUSIONS

This chapter discusses the results of the two studies in relation to the research questions proposed in Chapter 2, from the literature review, and, later, from the observations in Study 1. There are five sections in this chapter. The first section, **Possible Limitations of the Study**, discusses limitations imposed by the design of the study and by the measurement procedures. Second, **Summary of Results**, summarizes the results of the two studies, relating them to the research questions. Next, **Post Study Developments**, outlines some research developments which have taken place since the commencement of this study. Then, **Implications of the Study**, considers the implications of the findings on teaching, the development of the SOLO Taxonomy and future research. The final section, **Conclusion**, relates aspects of this study to present standards for teaching statistics in the Australian mathematics curriculum profile.

Possible Limitations of the Study

The results of the two studies must be viewed in the light of possible limitations imposed by the measurement instruments and features of the study design. This section reviews limitations which may restrict the breadth of any conclusions made from the study.

Restrictions on the size, 180 students, and the breadth, only secondary students, of the sample limit the scope of the results obtained. As the testing was to take place in the school situation, causing as few disruptions as possible, and only one person, the researcher, was to apply the tests, the original sample was necessarily restricted. This was, however, considered a large enough sample for the analysis to be undertaken. Unfortunately, with natural attrition and the conscious decision of some students not to participate in phase 2 of the investigation, the sample size was reduced to 117 students which resulted in low cell numbers for some of the analyses for Study 2.

As far as the breadth of the study was concerned the design only allowed for secondary students, from Years 7 to Year 12, to be tested. This was thought to be a wide enough range. However, there proved to be not as great a range between the levels of these students as expected and more depth at the lower and upper levels of the hierarchy could be achieved by testing students at primary level, below the age of 12 years, and at tertiary level, above the age of 18 years.

Two other limitations of the design of the study are the restriction of each focus area to one question containing two items, and the restriction of the longitudinal aspect of the study to only twelve months. The test instrument was designed with four questions, one to cover each of the four focus areas of statistics, excluding probability. Each question contained two items which were related. A more thorough test instrument would contain more items for each focus area. However, due to the open-ended nature of the questions, completion of such a test would be very time consuming, hence such testing may need to be done concentrating on one focus area at a time. The longitudinal aspect of the study, restricted to a twelve month period, proved to be too short to highlight any significant change in the level of student understanding. The progress of students over a longer period of time needs to be studied to obtain a clearer picture of individual student growth.

Despite these limitations the study has produced some useful results and does have a number of design features which distinguish it from earlier research. The longitudinal aspect was used to measure growth in understanding and the sample was stratified over academic year, mathematical ability and gender to investigate the possible effects of these factors. The test was designed with open-format questions set in realistic contexts to give students the chance to write as much as possible about familiar activities. The test was administered in the normal school environment so that students felt less threatened about completing the test. Finally, the use of Rasch modelling, by way of the QUEST software, to analyze the levels of understanding is a feature of the design which has not been used in earlier research concerning investigation of levels of understanding in statistics.

Summary of Results

A number of research questions were proposed at the end of Chapter 2 and the study was designed to address these issues. First, a hierarchy was developed to code levels of statistical understanding and the SOLO Taxonomy was used to describe the hierarchy. Second, the effect of certain factors on the level of understanding were investigated. Third, the effect of time on the range of understanding of students and individual growth of student understanding were considered. Finally, interview testing was compared to written testing. Following is a discussion of the investigation into the creation of a hierarchy and an explanation of the structure using the SOLO Taxonomy.

The Hierarchy and the SOLO Taxonomy

As student responses were coded and ranked, it became evident that a hierarchy could be constructed which indicated increasing statistical understanding. There were nine levels recognized in the coding of responses spread over three groupings. Levels 0, 1 and 2, in the first group, included those responses which reproduced information in the question and did not refer specifically to data or data collection details. Levels 3, 4 and 5, in the second group, included those responses which were concerned with physical aspects of data collection or able to work with the data, but only describing features of the data. Levels 6, 7 and 8, in the third group, included those responses which considered qualitative aspects of data collection or explained how features of data could be used for a statistical task.

The SOLO Taxonomy was found to be a suitable framework for describing the hierarchy. The ikonic mode corresponded to the first group, while the concrete-symbolic mode accounted for the second and third groups. Descriptions of these modes and the levels within the modes were developed for each of the four focus areas. These descriptions are detailed in Chapter 4 and justified statistically in Chapter 5. Two cycles of the unistructural, multistructural, relational levels were recognized within the concrete-symbolic mode, corresponding to the second and third groups, respectively. The first cycle showing an awareness of data and the second demonstrating the use of data for specific statistical tasks.

The step of major importance in secondary schooling is from the ikonic to the concrete-symbolic mode. This transition is made in dealing with data when the data are no longer considered as an endpoint in statistical discussions but as the beginning of a process, able to be processed and used for statistical tasks. The transition is made in the data collection process once the data are considered as relevant to the collection process.

An interesting feature which evolved during the development of the levels of the hierarchy was the divergence of responses at Level 5 into two distinct paths. This continued through Levels 6 and 7 until a convergence again with the better quality Level 7 responses. Varying performance of students, as far as level and path are concerned, raises the question of what factors are influencing the responses.

Factors Affecting Understanding

The factors academic year, mathematical ability and gender were found to have varying influences on level of understanding and path chosen. The form of data presentation did not influence the level or path as much as may have been anticipated.

The higher the academic year, the higher the level of the response, although the differences between the performance of academic years are only significant when the academic years are at least four years apart. Each question had bulges of responses, for the different academic years, which were mostly in the first cycle of the concrete-symbolic mode, in Levels 3, 4 and 5. This indicates that the secondary students in this sample are progressing through the first cycle and into the second cycle of the concrete-symbolic mode during the six years of secondary schooling.

The more mathematically able students are able to produce higher level responses and there is slight evidence that females respond at a higher level than males. The path chosen by students did not appear to depend on academic year, mathematical ability or gender. In an attempt to find an explanation for this divergence, student's tendencies to process successively or simultaneously were checked. Although processing style influenced the level of statistical understanding, it had little influence on the path of the response, with those strong in simultaneous processing more likely to respond along Path A. These students preferred using statistical features rather than judgements to describe data and measures of central tendency rather than dispersion to reduce data.

Effect of Time

Two areas were of interest here, namely, the effect of time on the overall coding and related factors and, also, on individual student growth in understanding. When the range of responses and the influence of factors discussed above were retested twelve months later, similar results were obtained, except for a few discrepancies which may have been due to sample sizes which were low in some cells in the second study. This consistency indicates that the hierarchy description and the effect of factors are robust over time.

As far as individual student growth is concerned, the time span of twelve months was not long enough to show any significant change in the level of understanding. This is consistent with the fact that there were only small changes in the level of understanding between successive academic years. To study student growth in understanding over time, a study spanning more than one year is needed.

Interview and Prompting

The testing of students by interviewing was found to produce similar level responses to the written test. Probing showed a slight increase in the level of response. Prompting, in the form of providing some information towards the answer or feedback on previous responses, was found to increase the level of the response in both the written and the interview situation. To accurately assess a student's level of understanding, interviewing is not necessary but prompting, especially in an interview situation, could be useful in ascertaining whether a student is ready to make the progression to a higher level.

Post Study Developments

It has taken a number of years to develop the hierarchy used to code the responses and to analyze the results During this time there has been a growing trend to apply the SOLO Taxonomy to coding situations across a wide range of subject areas. Research which has been undertaken since the commencement of this study consolidates some of the findings presented in this report. These include the two cycles within the concrete-symbolic mode, multimodal functioning, the effect of academic year on the level of understanding and the description of levels within modes.

Since the first recognition of more than one U-M-R cycle within the concrete-symbolic mode this phenomena has been observed by a number of researchers including Pegg (1992) in geometry and algebra items, Levins and Pegg (1993) in science items and Watson and Collis (1993) in probability and statistics. The most recent research into the application of the SOLO Taxonomy to levels of understanding in statistics has been undertaken in pilot stages of a large project to follow the implementation of the Chance and Data curriculum in Tasmania. Watson and Collis (1993) assessed items and the potential of the SOLO model as a tool for classification. Items devised were closed and so assessment using the SOLO Taxonomy could be in two different forms. Either the item could be classified by the level required to reach a correct answer, or the responses could be classified by the level attained by the student. In the present study, the items were open-ended and so only the responses could be classified, not the items themselves. Classification of the responses, in the Watson and Collis (1993) study, was only discussed for a limited number of questions, nevertheless they did clearly demonstrate two cycles of unistructural, multistructural and relational levels within the concrete-symbolic mode. The concept of two cycles in the concrete-symbolic mode was also observed in the coding of responses to interview items related to probability in yet another pilot stage of the Tasmanian chance and data project (Lidster, Pereira-Mendoza, Watson and Collis, 1995).

Early investigations in the project on developing interview protocol and using the theoretical framework of the SOLO Taxonomy to discuss the outcomes achieved by students are presented in Watson, Collis, Callingham and Moritz (1994). A report on the pilot study to develop three interview schedules for the project by Watson and Collis (1994) presented evidence of multimodal functioning in students in Years 3, 6 and 9 when responding to the interview items. Not only was the multimodal functioning considered important, but also the type of ikonic support given. Interview responses in fact showed strong support for the problem-solving path proposed by Collis and Romberg (1991), see Figure 2.8 in Chapter 2. The tendency toward multimodal functioning was also observed by Watson, Collis and Moritz (1994) in a report on the pilot study to develop a twenty item multiple choice/short answer questionnaire for the project. Another interesting feature of this report was that with the change in academic year from Year 6 to Year 9 little change was observed in the level of understanding. This is supportive of the present study which found that successive academic years did not differ significantly in mean level of understanding.

The present study investigated the longitudinal change over one year, finding no significant change, but the Watson, Collis and Moritz study has been funded to investigate the growth in level of understanding over three years. This has the potential to illuminate significant growth in understanding, especially in the students progressing through the period of schooling which includes the study of aspects of statistics.

More recently, as part of the Tasmanian project, Watson, Collis and Moritz (1995) tested, using both the questionnaire and interviews, a sample of 171 girls in Years 3, 5, 7 and 9 at a private school to develop a model for coding responses to sampling problems. The report includes the description of a model with levels and modes which reflects the model suggested in Chapter 4 for the data collection focus area. Again, the results include recognition of multimodal functioning and two cycles in the concrete-symbolic mode but, even more pertinent to the present study, are the descriptions of the features of responses at particular levels, the performance of the Year 9 group, the influence of prompting and the hypothesized convergence of ideas.

Analysis of responses uncovered concrete-symbolic responses which made use of ikonic support such as the Year 7 student who responded at U1 level but included with the answer "to impress people" (Watson, Collis & Moritz, 1995, p.11). This demonstrates that students engaged in multimodal functioning, as has been suggested in earlier research and was observed in the present study. Despite only testing students to Year 9 level, coding of responses uncovered two cycles in the concrete-symbolic mode as observed in the present study. Closer examination of the descriptions of the various levels in these two cycles (p.25) and the earlier ikonic mode shows very close correspondence to the level descriptions suggested in the present study. The first cycle, in the concrete-symbolic mode, has levels U1 and M1 concerned with aspects related to the sampling process and culminates with the basic idea of the sample as part of the representative measuring process. The research reported in Study 1 identified the first cycle as a concern with the physical aspects of the sampling process culminating with a desire for accuracy of the sample often expressed in an 'unstatistical fashion'. The second cycle has levels U2 and M2 concerned with the representativeness of the sample, as well as the notion of reliability and validity, culminating in R2 where these ideas are integrated. Similarly, in the present study a second cycle, which was concerned with the consideration of variables, was reported with R2 responses considering not only variables but concern with making the sample as representative as possible of the original population.

Not only are the descriptions of the levels very similar but the performance of Year 9 students on sampling items are similar in the two studies. Watson, Collis and Moritz (1995, p.16) observed that of the 64 students in Year 9 who responded, there were 73% within the four levels from U1 in the first cycle of the concrete-symbolic mode to M2 of the second cycle, inclusive. The present study observed that, of the 30 Year 9 students who responded in Study 1, there were, once again, 73% in these four levels. Also, similarly observed in both studies was the positive effect of prompting on the level of the response in both the written (p.12) and the interview situation (p.17). Finally, at the end of the report a convergence of ideas is hypothesized (p.29) to give two parallel cycles of U1, M1 and R1 in the areas of the random process and sampling which culminate in a second cycle dealing with the application of the sampling construct. The present study observed splits into two distinct paths in three of the focus areas but not in data collection. Also, the splits tended to be at different stages, namely, from R1 through to M2. More research is needed to determine whether there is in fact a similar split in the area of data collection and where the split occurs.

Some findings of the present study were pre-empted with the presentation of the results of the analysis of Question 3 from Study 1 (Reading & Pegg, 1995) at the Annual Conference of the Australian Association for Research in Education in 1995. At this conference, the above report from Watson, Collis and Moritz (1995) was presented also. Confidence in the hierarchy proposed in the present study is boosted by the fact that a number of independent research projects are working within a consistent framework.

Implications of the Study

Recent research has verified many of the results of this study with far reaching implications. Three important areas are addressed below. First, suggestions for modifications to the approach to the teaching statistics in secondary schools are outlined. Second, reinforcement of current ideas and new approaches to testing the structure of the SOLO Taxonomy are presented. Third, some ideas for future research in statistical understanding are suggested.

Implications for the Curriculum and Teaching

As suggested earlier, SOLO levels are useful for curriculum development as well as for student assessment. Students' levels of understanding, as observed in research, should be taken into consideration when constructing curriculum standards. The present study has indicated a development, although slow, in the level of sophistication of student understanding. Curriculum designers should take into consideration identified levels and develop activities, discussions and questions that aim to encourage the transition from one level to another.

In light of the results of the present study, some care would need to be exercised as far as the level of understanding expected of students in tasks in the Standards for School Mathematics (1989) in America. For example, part of Standard 10: Statistics (pp.105-108), for Years 5 to 8, suggests that students can "make inferences and convincing arguments that are based on data analysis". In the present study, results in Chapter 4 of the analysis for Question 4, on interpretation and inference, suggest that the majority of similar aged students in Australia do not exceed Level 5 on such inference tasks. This indicates that the students are not yet able to use data to justify patterns observed or inferences drawn.

Current research into SOLO type levels has been taken into consideration in Australia in the production of the N.S.W. Advanced Course Mathematics Syllabus (Pegg, 1995). Various aspects of teaching should be reassessed and restructured in the light of the findings on SOLO levels. In particular, new forms of student assessment, such as more open-ended tasks, should be considered and teachers may need to restructure their approach to teaching statistics, perhaps concentrating on a more student-centred learning environment.

With the changing curriculum, there is a need to review assessment procedures. Traditional norm-referenced tests measure the knowledge of a student relative to other students but interest should now be centred on assessing the "status and growth in each student's ability" (Romberg, 1992, p.43). Assessment of students should involve less decisions about whether an answer is right or wrong and more about the quality of the answer. Romberg (1992) suggested the need for a common assessment framework rather than a national examination system. This study suggests a hierarchy for such assessment, a student being graded as to whether responses have demonstrated the requirements of a particular level. The stage reached then determines those new concepts to which the student will be receptive in a learning situation.

Prompting has been shown to influence the level at which a student responds, so care in the teacher's approach in the classroom could assist the student with the transition from one level to the next, provided they are ready to make the move. Also, teachers need to be aware of the diverging paths along which students are operating. Teachers, just as students do, have a preferred path. One needs to be careful not to present new information to students always using one particular style because students who have a different preferred path may be disadvantaged. Teachers need to structure lessons and present material so that students are encouraged to make use of features from both paths. Data reduction should not be restricted to the discussion of central tendency, even at the simplest of levels the dispersion of the data should be discussed also. When describing data, as well as using statistics, qualitative assessments of data should be encouraged. If students are encouraged to operate in both spheres then they are more inclined to develop thinking along their less preferred path.

Curriculum and assessment procedure changes are long term tasks but changes to teacher strategies can be introduced immediately. As well as re-assessing present teaching strategies, teachers should be constantly on the look out for new teaching ideas which will assist in improving understanding. One recent example is the development of stem-and-leaf plots to improve understanding of graphs (Pereira-Mendoza & Dunkels, 1989). As well as the implications for teaching, the results of the study also have implications for the development of the SOLO Taxonomy.

Implications for SOLO Taxonomy Development

Some findings of the present study reinforce past and present research into the development of the SOLO Taxonomy, while others add to the knowledge of the application of the SOLO Taxonomy to statistical understanding. The features reinforced are the presence of two cycles in the concrete-symbolic mode and the use of ikonic support for concrete-symbolic responses. In all four focus areas covered by the questions, two cycles of unistrictural, multistructural and relational levels were found within the concrete-symbolic mode. This phenomena has been observed recently in studies of the application of the SOLO Taxonomy to a number of areas including statistics (Watson, Collis & Moritz, 1995), probability (Lidster, Pereira-Mendoza, Watson & Collis, 1995), and geometry (Davey & Pegg, 1992). Multimodal functioning was most obvious in the interview situation where students were prompted, and, as the pressure to produce reasons for answers or better answers became too intense, some students reverted to ikonic support for the initial concrete-symbolic level response. This dependence on earlier levels for support has been observed in other research including Watson and Collis (1994), Watson, Collis and Moritz (1994, 1995) and Redden (1995).

More specifically, the present study has developed descriptions for various levels when the SOLO Taxonomy framework is applied to statistics, and suggests the analysis of tau and threshold values to assist in confirming where cycles appear. As far as quantitatively recognizing the cycles of levels is concerned, no other study to date has used tau values and threshold values from the Rasch analysis for this purpose, as has been done in the present study. High tau values, that is, a high degree of difficulty in moving up to a level from the one below, for Levels 3 and 6, indicate that these are the levels which mark the beginnings of new cycles. Threshold values, that is estimates of understanding needed to attain a level, were similar for levels within a cycle or mode, such as Levels 3, 4 and 5 in the first cycle of the concrete-symbolic mode, and then show larger jumps in value for the moves between the modes or cycles within the modes, that is the moves from Levels 2 to 3 and from Levels 5 to 6. The analysis shows also a big step from Level 6 to Level 7 which, although considered to be within a cycle, was found to be a difficult move for secondary students. It must also be remembered that Level 7, also included the few Level 8 responses. Since the descriptions for the ikonic mode and the two cycles of levels in the concrete-symbolic mode were developed for this study, the chance and data project in Tasmania has also been developing descriptions for levels of understanding in statistics (Watson, Collis & Moritz, 1995) which are consistent with the present study.

All these findings help to consolidate features of the SOLO Taxonomy and to expand on the knowledge of the application of the taxonomy to the hierarchy of understanding in statistics.

Implications for Future Research

From the findings of the first study, there are four major areas which need further research, namely: a more detailed description of levels; a development of assessment instruments; an explanation of the split of response coding into paths; and, an investigation of the growth in understanding over time. More research is needed within each of the four focus areas of statistics, and also probability, to produce more detailed descriptions of the SOLO levels in the hierarchy of statistical understanding. In particular, more information is needed with respect to the description of the upper and lower ends of the hierarchy. Primary students' responses need to be assessed to better describe the ikonic mode and upper secondary and tertiary students' responses could provide more information about the second cycle of the concrete-symbolic mode and the formal mode.

Different researchers are developing various questionnaires and interview protocols to procure responses to develop descriptions of the levels of the hierarchy. Now an assessment instrument is needed which will allow non researchers, especially teachers, to assess the level of a student's understanding of statistics. Care needs to be taken with the wording of questions or tasks. In the present study, even though deliberate probing was presented at times, there may have been implied probing in some questions. For example, Question 3 was not meant ot contain probing but may have guided students by asking for a "number or numbers" to represent data. There may be value in replicating this study with subtle changes to the questions.

The two parallel paths into which responses were split, from the relational level in cycle one to the relational level in cycle two in the concrete-symbolic mode, need to be investigated further. The split was observed in three of the four focus areas and more research is needed to determine whether such a split exists in the data collection area as well. Many students showed a tendency to process along one path or the other. Further research into this tendency may be able to explain why students make particular responses, as the preference for successive or simultaneous processing was found to provide little explanation for the phenomena. Such information may be helpful in assisting teachers to plan the strategies, for presenting statistics, to be used in the classroom.

Finally, very little research has focused on the growth, over time, of statistical understanding. This study found that over the period of twelve months there was not an appreciable increase in the level of understanding of students. This growth needs to be studied over a longer period of time. The chance and data project in Tasmania plans to study growth over a three-year period. This research and other work planned by the project may help to address aspects of the future research suggested. Continuing study of the levels of understanding in statistics should enable educators to improve the teaching and assessment of statistics for secondary students.

Conclusion

The importance of statistics has been recognized with the inclusion of Chance and Data as one of the five basic strands in the Australian Education Council's (1991) *A National Statement on Mathematics for Australian Schools*. Also published, to assist teachers in their implementation of the national statement, were a curriculum profile (Australian Education Council, 1994a) and worked sample questions (Australian Education Council, 1994b). The profile for the curriculum describes Levels, 1 to 7, for each of the five basic strands. Within the Chance and Data strand, these levels are described for five areas, the four focus areas covered in the present study and probability. In some ways, these levels correspond to the levels observed in the research reported in this study but there are also instances where there are differences in expectations.

Close correspondence of levels exists in the data collection focus area. This study described Level 6, the entry level of the second cycle of the concrete-symbolic mode, attained by the majority of students by the end of Year 12, as involving consideration of the quality of the data when planning sampling. The curriculum profile, also, first mentions concern for data consistency in Level 5. Previous to that, concern is focused on taking "care in collecting data", introduced in Level 3. This corresponds to the concern with physical aspects of the data collection process which is characteristic of responses in Levels 3, 4 and 5 in the first cycle of the concrete-symbolic mode in the present study. However, not all aspects of the levels described in the curriculum profile and this study coincide so precisely. For example, in data reduction, Level 6 of the curriculum profile has students summarizing data using both location and variability and in the present study Level 6 responses only used one style of measure or the other. It was only the better Level 7 responses which combined aspects of both location and variability into the summarizing of data. Clearly, much more research is needed in this area if curriculum documents are to be more than some pragmatic list of content and reflect accurately students' understanding.

In general, this study reinforces the level structure presented, if not the absolute description of each level, in the curriculum profile. However, care should be taken when using the curriculum profile. The great range in the quality of the responses observed in the present study for students in any particular academic year indicates that it would be dangerous for a teacher to assume that all students in a particular year would attain the same level. The levels should be used as a guide to assess the capabilities of individual students. Activities similar to those in the present study or the *Mathematics - Work Samples* (Australian Education Council, 1994b) could be used for such assessment. The learning situation should be constructed so as to allow students to expand within their own capabilities. In presenting activities, teachers need to be mindful of the academic year, mathematical ability and gender of the students, all of which this study has shown to influence, to varying degrees, the students' understanding of statistics.

Continued research should help to refine measurement instruments which will allow grading of students' responses and the assessment of students' level of understanding. As a student progresses through the schooling process, these assessments should build up to produce a profile of the growth of student understanding in statistics. Similar work on responses in other academic areas, would ultimately allow the comparison of a student's level of understanding over a large variety of tasks. The present study, although useful, only undertook a small proportion of the research which is needed.

Most students' views of statistics are wrapped up in aspects of calculation. These views need to be broadened and a deeper understanding encouraged. To do this more attention needs to be paid to students' thoughts and their methods of expression. Perhaps we all need to heed the words of Henry Clay who said "Statistics are no substitute for judgement".