CHAPTER 4

RESULTS - STUDY 1

The preceding chapter presented the design of Study 1, including aspects of both the pilot study and main study. The pilot study involved the development of the test instrument and testing procedures. The results of this study were also included as they influenced the design of the main study.

This chapter presents the results of the first study and some preliminary conclusions. The results for each of the four focus areas are dealt with separately. The first section is **Data Collection**, then **Data Tabulation and Representation,** next **Data Reduction** and finally, **Interpretation and Inference.** Information is presented in a similar format for each of the focus areas. Any tables considered necessary for secondary reference but not vital to the argument presented appear in the Appendix rather than in the body of the study. Following now are the results for the first of the four focus areas.

Data Collection

The first area of statistics investigated was Data Collection. As mentioned earlier, this topic involves the collecting of data in the form of a sample, suitable to answer the inquiry being investigated. This includes all forms of collection of sample and census data. There are two main aspects of concern here: the type of data that can be collected; and, the different sampling techniques.

The first question in the test, the data collection question, was designed to help clarify students' understanding of the process of data collection. The aim of the question was to present students with possible sampling situations and to allow them to decide on how the data collection should proceed. The sampling problem was presented in two different forms. In Part I, the problem for investigation was stated and students needed to decide on a sampling technique. This was an attempt to view what students perceived as possible means of collecting information to address the issue presented. Part II stated the method of sampling as well as the problem, leaving the task of deciding on the content of the sample up to the student. This was done in order to investigate students' understanding of data collection in general terms, and their appreciation of the actual process of sample selection.

The analysis of the responses to Question 1 is considered in this section. The responses to each part of the question are analyzed separately before being brought together for an overall analysis of performance. First, there is a discussion of possible levels into which the responses to Part I may be categorized. The performance of students over these

levels is presented and the implications of these observations are discussed. Then the results for Part II of the question are presented. Next, Parts I and II are compared to consider similarities and differences in the student performances, and the implications of these observations. Also, the overall performances of individual students for the two parts are considered. Finally, the SOLO Taxonomy is used to suggest a possible framework for explaining the responses and features of the various levels within modes.

Analysis of Responses to Part I

The first part, I, of Question 1 was a general data collection question. The question, as it was presented to the students, is given in Figure 4.1. The open-ended question was designed so that no reference was made specifically to any method of data collection and students were free to suggest any method that they felt was suitable.

When the responses have been recorded, in this section, other sections and the appendices, they have been reproduced as given by the students, including the use of capital letters or full stops. Misspelt words have been left as such, for example, *culd* for *could*, abbreviations often have the ' missing, for example *its* instead of *it's* and incorrect words have been left, for example *there* instead of *their*. First, the coding of the responses is described. Then, the statistical interpretation of this coding is outlined.

QUESTION 1
PART I

Radio stations have their own way of working out the most popular song on the radio and they often produce Top 40 charts. Imagine that you have been asked to do this independently of the radio station and answer the following questions :

(i) Describe the best way to find out what the most popular songs are on the local radio station.

(ii) Why did you decide to find out this way ?

Figure 4.1 - Data Collection - Part I Question

Coded Responses

Investigation of the students' responses showed that it was possible to divide them into a number of levels based on the statistical quality of the answer given. The levels

which were observed were grouped together based on the depth to which the response indicated how well the student was able to cope with the question. There are three major groups, and sample responses for the various levels within these groups are presented.

First Group

First are those responses which dealt with only the requirements of the question and within this group were two broad levels, coded as 1 and 2. These responses indicate a rationalizing of the requirements of the question with no real concern about the actual data collection. The following is a brief description of the levels within this group and sample responses from each. The responses have been recorded with the answer to part (i) indicated as such and the answer to part (ii) similarly (when given by the student).

Level 1 These responses indicate that not all aspects of the question have been considered. This results in a response which does not fully address the question, suggesting the use of data that have already been collected rather than collecting data. For example:

(7102)	(i)	*You could ring up the radio station, ask someone that works there, go in and ask them*
	(ii)	*It was the first thing that came into mind. It would be easier than worrying about it.*
(7209)	(i)	*Listen to some of the songs and pick the one that sounds the best or that you like.*
	(ii)	*Because if you have good taste other people might like the same music.*
(8207)	(i)	*Ask the radio announcer how they do it.*
	(ii)	*I don't know.*

Level 2 These responses indicate that all aspects of the question have been considered, and a reasonable answer has been produced. However, a suitable explanation could not be given as to why the answer was chosen, or, if an explanation was given, it is a personal comment. For example:

(8108)	(i)	*Whichever songs got requested the most.*
	(ii)	*because most people would want to hear the song.*
(9213)	(i)	*write down how many times they occur.*
	(ii)	*?*

(12102) (i) *Watch programs such as Rage or Video Smash Hits then maybe listen to the radio to see if it is right, or just ring up the station and ask.*

(ii) *So that I know what to expect if they ever play a song on the radio.*

Second Group

The second group of responses are those concerned with rationalizing the method of collection of the data. Three levels were observed within this group, coded as 3, 4 and 5. These responses attempt to describe suitable data collection, but are mainly concerned with physical aspects of the collection, such as, the time or cost involved. There is no evidence of concern for the quality of the resulting sample.

Level 3 These responses indicate that, in attempting to justify the suggested method of data collection, focus was directed back to the question and not to any specific aspect of the collection of the data. For example:

(7211) (i) *Have a piece of paper sent to all houses, get them to write their favourite songs on them and return them to the radio station.*

(ii) *I decided this way because I think it would be a good idea.*

(8208) (i) *Call people like 2AD does (Top 7 at 7).*

(ii) *Because I feel it to be the best way.*

(8215) (i) *I would listen to the radio and note the frequency (how often they are played) of certain songs. The most frequent would be those songs most popular on the radio.*

(ii) *I chose this way because, logically, the radio would play its most popular songs more often than others.*

(11108) (i) *Set up a request program and get the local public to request songs.*

(ii) *Because its the only way I could think of.*

Level 4 These responses give reasons, with an explanation for the method chosen, which focus on various physical aspects of the collection process. There is no real concern for the accuracy of the resulting sample. For example:

(7108) (i) *Do a survey and get people to ring the radio station and vote for a song they like and the one with the most votes will be number one do as it does on the T.V. show Rage.*

(ii) *I decided to find out this way because it would be easy for anybody and it is easy to play the songs they ask for.*

(9208)	(i)	*ask a certain amount of the population that listens to the radio station. Don't ask everybody.*
	(ii)	*Because any other way would be too expensive.*
(10213)	(i)	*Have a phone in census or a questionnaire that is put through the public for their forty favourite songs.*
	(ii)	*I decided to find out this way as you can get a larger amount of information in a relatively short amount of time.*
(12107)	(i)	*Surveys delivered to all people of all ages which have a list of the songs to be played in a rank of best to worst.*
	(ii)	*People can be totally honest without being influenced by any other person's decision. It is quick, cheap and anyone can participate.*

Level 5 These responses indicate that concern for the physical aspects of the data collection have been rationalized, but the only concern for the quality of the resulting sample is that data have been collected in such a way that the sample is fair or accurate. However, no indication is given as to how this is to be achieved. For example:

(8113)	(i)	*By finding what music is bought as singles most at the music stores.*
	(ii)	*Because it is the most accurate way of finding out this.*
(10112)	(i)	*I think that the best way would be to send out a survey to the local people and then add them all up to see what type of music is most popular.*
	(ii)	*I decided to find out this way because then you would have an overall view of what sort of music that the local people like to listen to.*
(12205)	(i)	*Have people ring in and vote for their favourite song. The song that is most popular will be no. 1, the second most popular no. 2 and so on.*
	(ii)	*To give everybody an equal chance of giving their opinion of their favourite song.*

Third Group

The final group of responses indicate that concern for the physical aspects of the method has been rationalized, and is now centred on the quality or accuracy of the data in the resulting sample. This group has two levels of responses coded as 6 and 7.

Level 6 These responses indicate the need for sample selection to be arranged so as to produce a range of data in the sample, which satisfies the needs of the investigation. There may be mention of a variable on which the sample selection is to be based. For example:

(7110) (i) *I would ask people for their top 40 songs and then compare them.*
(ii) *because then you would get a varied opinion.*

(9108) (i) *The best way would be to survey people and get their ideas which music they would like to hear and which is the most popular songs.*
(ii) *Because it would be the easiest and you would get different people's views on the type of music they like.*

(12104) (i) *Do a random survey on the radio turning it on at different times of the day at different intervals noting the songs that are being played. Also do a random survey of the listeners of the particular radio station, asking them what songs they request, and most frequently hear being played.*
(ii) *Because it gives you a less bias opinion and view on the popularity of certain songs. You can get a wider census area, making the results more realistic.*

Level 7 These responses indicate that selection of the sample has been made on more than one variable in an attempt to improve the range of the responses which are collected. For example:

(12114) (i) *Collect group of people of varying ages and background who listen to the radio and ask them their favourite songs.*
(ii) *Not biased to any group of people and asks people who are interested in music because they listen to the radio*

The responses presented in each of the levels are only a sample of all responses given, but were considered to be representative of the responses at the level indicated. It was possible to grade each of the responses into one of the seven levels outlined above. The check on coder reliability showed a 90% agreement. The three non-agreeing codings, out of the thirty, were all only one level from the researcher's coding and within the same group. The three groupings discussed are used in the development of a theoretical perspective in the SOLO Taxonomy Framework section to follow. However, before this is addressed, the results are analyzed according to academic year, mathematical ability and gender.

Statistical Interpretation

The results, arranged by academic year, are presented in Table 4.1.1. These data illustrate a number of interesting points. First, there are only three students (5%) from

Years 11 and 12 whose responses fall within the first group (Levels 1 and 2), whereas in Years 7 and 8 there are a number of students, nine (15%), in this group. Second, there are only three students, all in Year 12, whose responses were coded as Level 7. Last, there was an overall bulge at Levels 3 and 4. This bulge is consistent in all years, except Year 12 where the bulge shifts to Levels 4 and 5. This suggests there is a slight overall increase in the level of response from Year 7 through to Year 12.

Table 4.1.1 - Response Level by Academic Year - Question 1 Part I

Level	Year						Total
	7	8	9	10	11	12	
1	5	1	1	0	2	0	9
2	1	2	3	5	0	1	12
3	10	9	9	6	10	5	49
4	10	14	9	12	12	9	66
5	2	3	4	4	5	7	25
6	2	1	4	3	1	5	16
7	0	0	0	0	0	3	3
Total	30	30	30	30	30	30	180

Table 4.1.2 - Response Level by Mathematical Ability - Question 1 Part I

	Mathematical Ability			
Level	Low	Middle	High	Total
1-2	6	4	2	12
3	17	14	18	49
4	24	25	17	66
5	6	7	12	25
6-7	2	7	10	19
Total	60	60	60	180

The results, arranged by mathematical ability, are presented in Table 4.1.2 (on the previous page) and show that the coding level of a response is independent (χ^2 = 14.47, 8 d.f., $p > 0.05$) of the mathematical ability of the student who gave the response. The results, arranged by gender, are presented in Table 4.1.3. Although there appeared to be a larger number of females represented in Levels 6 and 7, testing showed that the coding level a response is independent (χ^2 = 10.13, 5 d.f, $p > 0.05$) of the gender of the student.

Table 4.1.3 - Response Level by Gender - Question 1 Part I

	Gender		
Level	**Female**	**Male**	**Total**
1	4	5	**9**
2	7	5	**12**
3	18	31	**49**
4	37	29	**66**
5	10	15	**25**
6-7	14	5	**19**
Total	**90**	**90**	**180**

It appears, that for this general data collection question, the level of the response improves slightly with academic year but is not significantly influenced by the mathematical ability or the gender of the student. These results are compared with those for Part II, following the discussion of the Part II responses.

Analysis of Responses to Part II

Part II of Question 1 was more specifically aimed at details of the selection of the members of a sample. The question was designed so that the concernof the size was already addressed but the question was still open-ended in that students were free to suggest which 30 students to use and their reasons for selecting those particular students. The question, as it was presented to the students, is given in Figure 4.2. A discussion of the coded responses is followed by a statistical interpretation of the coding.

QUESTION 1
PART II

There are often surveys of the community to see what T.V. programs they like to watch. The editor of the school magazine is interested in writing an article about the viewing habits of the students at A.H.S. and asked you to find out the information

(i) You are only able to ask 30 students from the school. Which students would you select to ask ? (Don't use names)

(ii) Why would you select these students ?

Figure 4.2 - Data Collection - Part II Question

Coded Responses

As with Part I, it was possible to divide the responses into a number of levels based on the statistical quality of the answer students gave, and these were divided into the same three major groups used for Part I. A selection of responses, worded in slightly different ways to those presented for Part I, for Levels 1 to 7 are included in Appendix D. Two levels which were observed for this part of the question which were not observed for Part I were coded as 0 and 8.

Level 0 This level was included in the First Group before Level 1. These responses indicate that no attempt at all has been made to answer the question. There was only one Year 7 response at this level and it only had the question number written.

Level 8 This level was included at the end of the Third Group after Level 7. These responses indicate that the selection of the sample is based on a number of variables and also on the composition of the population from which it is drawn. This is a more thorough attempt to make the sample representative.

(8215) (i) *I would try to select a broad spectrum of the populous, taking into account age and social groupings. I would keep the divisions proportionate to what they are in the school environment e.g. there are 80 people in one social group and 20 in another therefore i would take 4 people randomly from group 1 and 1 person from group 2.*

(ii) *I would use this method to be sure of getting the full range of viewing habits within the school, but so as not to overestimate the statistical effects of minority groups.*

It was possible to grade all responses into one of the nine levels, 0 though to 8, and these results are considered now according to academic year, mathematical ability and gender.

Statistical Interpretation

The results, arranged by academic year, are presented in Table 4.1.4. These data illustrate a number of interesting points. First, there are no students from the two senior years whose responses fall within the first group (Levels 0, 1 and 2), whereas in Years 7 and 8 there are a number of students, eight (13%). Second, there are only seven (12%) students from Year 7 and 8 whose responses were coded as Level 7 or 8 compared to eighteen (30%) Year 11 and 12 students. It was interesting though that the only Level 8 response came from

Table 4.1.4 - Response Level by Academic Year - Question 1 Part II

Level	Year						Total
	7	8	9	10	11	12	
0	1	0	0	0	0	0	**1**
1	1	0	0	0	0	0	**1**
2	3	3	0	3	0	0	**9**
3	1	0	0	0	1	0	**2**
4	12	7	8	3	1	3	**34**
5	7	9	10	5	7	11	**49**
6	2	7	10	13	13	6	**51**
7	3	3	2	6	8	10	**32**
8	0	1	0	0	0	0	**1**
Total	**30**	**30**	**30**	**30**	**30**	**30**	**180**

a Year 8 student, albeit, one who was very capable. Last, there is an overall bulge at Levels 5 and 6. This bulge, however, varies from year to year. It ranges from Levels 4 to 5 for Year 7 through Levels 4 to 6 for Year 9 to Levels 5 to 7 for Years 11 and 12. These observations

suggest that there is a slight improvement in the quality of the response from Year 7 through to Year 12.

The results, arranged by mathematical ability, are presented in Table 4.1.5. There is very strong evidence ($\chi^2 = 42.36$, 8 d.f, $p < 0.001$) to suggest that coding level is dependent on mathematical ability. There appears from Table 4.1.5 to be more students of low mathematical ability in the lower levels, and a greater number of high mathematical ability students responding at Levels 7 and 8. The results, arranged by gender, are presented in Table 4.1.6. There is some evidence ($\chi^2 = 9.92$, 4 d.f., $p < 0.05$) to suggest that coding level is dependent on gender, the performance of the responses from the females being slightly better.

Table 4.1.5 - Response Level by Mathematical Ability - Question 1 Part II

	Mathematical Ability			
Level	**Low**	**Middle**	**High**	**Total**
0-2	9	2	0	**11**
3-4	19	13	4	**36**
5	13	16	20	**49**
6	16	21	14	**51**
7-8	3	8	22	**33**
Total	**60**	**60**	**60**	**180**

Table 4.1.6 - Response Level by Gender - Question 1 Part II

	Gender		
Level	**Female**	**Male**	**Total**
0-2	2	9	**11**
3-4	15	21	**36**
5	24	25	**49**
6	33	18	**51**
7-8	16	17	**33**
Total	**90**	**90**	**180**

For this more specific data collection question there is an improvement in the level of the response with advancing academic year and both the mathematical ability and the gender of the student appears to affect the quality of the response. Females and students with higher mathematical ability tended to give responses at a higher level.

Comparison of Part I and Part II

Parts I and II were analyzed in the preceding sections and some features of the data discussed. Similar groupings and levels proved useful for both parts of the data collection question. In this section, the performance of the responses for the two parts of the question are compared. First, an overall comparison of the student numbers in each level for the two parts of the question, is undertaken. This is followed by a discussion of the implications of any trends and differences. Then, the performances of students are compared for both parts of the question.

Overall Comparison

The fact that it was possible to allocate all student responses to one of these levels suggests that this framework may be adequate in explaining students' understanding as far as the basic concepts of data collection are concerned. Comparing the number of students within each level, for the two parts of the question, assists in highlighting the trends and differences between the two sets of data.

The number of responses that were coded in each level for Parts I and II appear in Table 4.1.7. There is strong evidence ($\chi^2 = 109.5$, 6 d.f., $p < 0.001$) to suggest that the there is an association between the level of the coding and the part of the question being answered. Referring to Table 4.1.7 suggests far less responses are coded into the Levels 0 to 3 and far more responses in the Levels 6 to 8 for Part II of the question than for Part I. Even the large bulge of responses at Levels 3 and 4 for Part I has shifted up to Levels 4, 5 and 6 for Part II. This suggests that, given a sampling situation, students are able to discuss the practicalities of the collection of data, but find it difficult to rationalize this sufficiently to consider the consequences of the sample choice on the data produced. However, once students have been prompted with some information about the sample, they are able to concentrate on which members of the population to choose. This has allowed an increase in the level of the responses, with consideration now being given to the aspects of selection which affect the sample selected and hence the quality of the data collected.

Table 4.1.7 - Level Comparison for Parts I and II - Question 1

	Question 1		
Level	Part I	Part II	Total
0-1	9	2	11
2	12	9	21
3	49	2	51
4	66	34	100
5	25	49	74
6	16	51	67
7-8	3	33	36
Total	180	180	360

In terms of the influence of academic year, both parts of the question yielded similar results. Comparing the results presented in Tables 4.1.1 and 4.1.4, there are three noticeable trends. First, there are at least twice as many Year 7 and 8 students in the first group (Levels 0, 1 and 2) as there are Year 11 and 12 students. Second, there are more Year 11 and 12 students than Year 7 and 8 students in the final two levels of the hierarchy. Third, there is a very large bulge in the numbers in each case. However, there is a difference in how this bulge is distributed through the academic years for the two parts. For Part I, the bulge is in Levels 3 and 4 for all years, except perhaps for an increase to Level 5 in Year 12. However, for Part II, the bulge shifts progressively upwards from Level 4 to Level 7 as the years progress from 7 through to 12. This indicates that there are more differences in ability between the academic years when it comes to describing the details of sample selection than when formulating a general sampling situation.

It appears that mathematical ability is more likely to affect performance in Part II than Part I. Comparison of results in Tables 4.1.2 and 4.1.5 shows that the level of response was only dependent on mathematical ability when responding to Part II. The gender analysis results, in Tables 4.1.3 and 4.1.6, also showed a significant dependence was found only in the Part II analysis, where the females appear to have performed at a better level than the males.

The differences in levels of response for Part II suggests that females and those who achieve at a higher mathematical level were more adept at taking into account theoretical aspects of the sampling situation. These students were more likely to consider the resulting data and the population, which was being sampled, when suggesting which

members of the population should be included in the sample. It appears that under prompting, in the question, the level of response has improved from Part I to Part II but this prompting appears to have been more beneficial for females and the higher mathematical achievers.

Although there were some similarities in the levels of the responses for Parts I and II, there were differences in the quality of the responses, as would have been expected, from the design of the two parts of the question. It appears that it is possible to develop a hierarchy of levels to categorize responses and the responses fall into the categories in an explicable fashion. These levels are used in a later section to construct a theoretical framework for describing student understanding with respect to data collection.

Comparison of Student Performances

To determine whether there is consistency in the way in which any particular student responds, with and without prompting, the performance of students over the two parts of the question is now considered. A comparison of the level of each student's response to Part I and Part II are presented in Table H.1.1 (Appendix H). It would appear from this table that there is a link between the level at which a student responds for the two Parts, I and II, of the question. The table has too many cells to allow analysis and hence the information has been condensed using each student's grouping rather than level. These groupings are given in Table 4.1.8. For Part II the first and second groups had to be combined because of the very small number of responses which were in the first group. There is strong evidence ($\chi^2 = 9.11$, 2 d.f., $p < 0.02$) of an association between the response grouping and the part of the question being answered. This suggests that a student's coding on Part I can be used as an indicator of performance on Part II. This relationship between the two parts of the question is worth investigating further.

Table 4.1.8 - Group Comparison within Students - Question 1

		Q1 Part I			
		First	Second	Third	Total
Q1 Part II	First & Second	13	79	4	96
	Third	8	61	15	84
	Total	21	140	19	180

Looking at the numbers of responses at particular levels for each part, rather than just groupings, in Table H.1.1 suggests an increase in the ranking of the responses from Part I to Part II. The Nonparametric Sign Test applied to the paired data for the students' responses gave 40 differences of 0 which were ignored. Of the remaining 140 differences, there were 126 which were negative, which gives strong evidence ($Z = 9.38$, $p < 0.01$) of a student's coding levels not being the same for the two parts of the question. As there were more negatives than positives, the students tended to attain a higher coding on Part II of the question than on Part I. In fact, there was a marked increase in the level of the responses over the two parts of the question. Only 22% of students were coded as achieving the same level of response for the two parts and a mere 8% actually had a reduced coding level for Part II. This indicates that not only was there a general improvement in the level of the responses, but, individually, more than two thirds of the students showed more understanding in their response when given some prompting in Part II.

In both parts of the question, data collection tasks were presented but in the second part prompting was given as to some of the physical details of the sample. The significantly higher level attained in Part II suggests that, with prompting, students were able to demonstrate a deeper understanding. Most noticeably, those students who were concerned with the practical aspects of the sample collection in Part I were able to free themselves of these concerns in Part II and take into consideration variables affecting the population members selected, and the effect that this has on the resulting data that are collected. The comparison of the grading of responses to Parts I and II indicate that the levels used to categorize the responses are sufficient to show trends which are not totally unexpected. These levels are used in the next section as a basis for describing a SOLO style framework.

SOLO Taxonomy Framework

The levels established for the classification of the responses are now used along with the structure of the SOLO Taxonomy to create a framework which could be used in future to code student responses. The first group of three levels contains responses which were ikonic in character. The responses in the second and third groups represent two different cycles in the concrete-symbolic mode. Examples of responses in the following discussion are referred to by the subject's research number and the responses can be found either earlier in the chapter in the appropriate level or, if followed by a D, in Appendix D.

In the ikonic mode, the responses suggest that the required task could not be linked with any sort of symbolic representation. Such responses were mainly from students in the junior secondary years and were more numerous for Part I than Part II. There is a framework of growth within this mode. The responses which appear as Level 0 are prestructural, where no attempt is made to answer the question.

The responses in Level 1 are a mixture of unistructural and multistructural responses. It was difficult to separate these responses because of the small overall number of responses in the ikonic mode. Also, because the task had not been addressed, it is difficult to determine how many visual cues from the question are in focus or what personal beliefs and experiences have been drawn upon, without further investigation. Students would need to be interviewed to determine which aspects of the question, if any, they have drawn upon. These responses range from those students (such as, 8207) who chose to ask others how to collect the data and could not give a reason, to those who discuss song popularity but only consider their personal choice (such as, 7209).

Finally, Level 2 responses correspond to the relational level within this mode. Here the responses indicate that the various aspects of the question have been incorporated. The question has been understood sufficiently to enable a reasonable answer to be proposed. However, attempts to justify the answer are unsuccessful or result in a personal comment. These range from a student (8108) who would choose the most requested songs because people would want to hear them, to a student (10203,D) who chose to ask Year 10 so that questions could be directed at people known personally, to another student (12102) who justified her decision to watch programs or ring up the radio station by explaining that she would then know what to expect to hear on the radio.

The responses in the second and third groups have been able to link the concepts in the question to concrete experience. Every indication is that the question has been understood. The reasons given are linked directly to the practical aspects of data collection or to concerns about the accuracy of the sample. These responses are in the concrete-symbolic mode. Within this mode, two cycles became evident. Each cycle contains three levels, unistructural (U), multistructural (M) and relational (R).

The first cycle involves consideration of physical aspects of the data collection. The elements in the first cycle are the practicalities which need to be taken into consideration when data are to be collected. Typical considerations are the number in the sample and the type of data to be collected. These are influenced by things, such as, the cost and time involved. A relational response in the first cycle is not achieved until the student is able to consider all physical considerations as a functioning set, and hence come to the realization that there is more than this to be considered. In this cycle, the unistructural, multistructural and relational levels correspond to the Levels 3, 4 and 5 as identified earlier in the section. A description of the three levels follows.

U1 - These responses indicate that consideration of practical aspects has focused back to the question, and the reasons given are usually that the method chosen is the best. A variety of suggestions are made as to the method of collection, such as, ring-in requests or surveys, but no practical aspects of the collection are considered.

Typical responses are those of a student (7211) who chose to survey houses to find out the most popular songs because it would be "a good idea" and a student (7213,D) who suggested choosing four children from each year in the sample of 30 because it was one way of doing it.

M1 - These responses indicate that as well as deciding on a method, consideration is given to other physical aspects of the data collection, such as, time taken or ease of collection. Typical responses are those of a student (7108,D) who chose particular people to question because they would be "sensible and responsible", another (9208) who was concerned about expense and a student (12107) who wanted people to answer honestly as well as ensuring that collection was quick and cheap.

R1 - These responses indicate that all considerations of the physical collection of data have been combined and some concern is shown as to the accuracy of the data collection. However, no indication is given as to how the sampling technique should be applied in order to achieve this accuracy. A typical response from a student (8113) suggested collecting data from music stores because this would be the "most accurate" method. Often this desire for accuracy was expressed as a desire to be fair, by giving everyone an equal chance to give an opinion (student (12205)) or survey people who are different as suggested by a student (7104,D) who wanted to select students from each of the Years 7 to 12 so that the sample contains people of different ages. There is, at this stage, no suggestion of what this accuracy really involves, that is, having the sample contain a good representation of the possible variation that could occur in responses from the population.

The second cycle involves appreciating that the method of selection of the sample influences the quality of the responses. Reasons given in responses now indicate that some attention has been focused on ensuring that the data collected presents a suitable range of opinions. The elements in the second cycle are the various considerations which affect the variety of data in the sample, such as, consideration of variables that may be used to select which members of the population are to be included in the sample. A relational response in the second cycle is not achieved until a variety of variables have been considered as concern is centred on making the sample as representative of the population as possible. In this cycle, the unistructural, multistructural and relational levels correspond to the Levels 6, 7 and 8, as outlined earlier. A description of the three levels follows.

U2 - These responses indicate that data collection has been based on selecting a sample according to qualitative considerations. At this early stage, the major concern is that the sample, once collected, exhibits suitable variety. Typical responses range from a student (7110) who strived to achieve varied opinions, to a student (12104) who

wanted a selection based on a random technique so that there would be less bias, to another student (11213,D) who selected five students from each year to see if "TV shows people watch differed with age".

M2 - These responses indicate that more detailed consideration is given to these qualitative aspects of data collecting, which usually involves selection of a sample based on more than one variable. Typical responses are those of a student (12114) who chose people of varying ages and backgrounds to ask about favourite songs so that the group would not be biased, another student (7111,D) who selected from each year and smart kids and those not so smart so as to ensure a "wider viewing range", and one (12215,D) who suggested selecting the sample based on academic year, sex and background (economic, social and racial) to achieve average viewing habits with "no emphasis placed upon a certain group".

R2 - These responses indicate that the sample selection is not only based on consideration of a number of variables, but concern is also shown as to the composition of the original population and making the sample as representative as possible of this population. There was only one response in this group, a Year 8 student (8215) who suggested selection of the sample based on a number of variables and recommending that the selection be based on the proportions of these groups in the population. The reason given was that a full range of viewing habits was needed in such a way as "not to overestimate the statistical effect of minority groups".

The framework presented above could be used to grade future responses to help determine the stage a student has reached in understanding the data collection process. The main feature which distinguishes the concrete-symbolic mode responses from the ikonic mode responses is evidence of the recognition that data need to be collected to address an issue. Ikonic mode responses either do not address the required task, or address the issue without collecting data. Instead, they suggest using information collected by others or considering a personal judgement sufficient. The concrete-symbolic mode responses discuss one or more aspects of the data collection process. Within this mode, the first cycle responses are only concerned with physical aspects of data collection, while second cycle responses consider the influence of the method of selection on the quality of the data. Within each of these modes, more detailed descriptions were given to assist with the classification of any response to a particular level. This framework is designed specifically for use with classifying responses to questions concerning data collection. Frameworks for the other three focus areas are developed now, following the analysis of the responses to each of the other three questions. The next section presents the results of the analysis of Question 2.

Data Tabulation and Representation

The second area of statistics investigated was Data Tabulation and Representation. This involves two main areas, the tabulation of data and the pictorial representation of data. Question 2, the data tabulation and representation question, was designed to help clarify students' understanding of data presented in two basic forms, frequency table and frequency histogram. The question was not testing students' ability to arrange data into a table or a graph, instead it aimed at presenting students with some data and having them describe what information they were able to gather from the representation. The question was left open with no reference to any specific facts to investigate what students perceived when presented with data. In Part I, the data were presented in table form and in Part II as a graph. This was done in order to investigate whether the form in which the data were presented influenced what information the students were able to gather from the data. This section mirrors the structure of the Question 1 analysis.

Analysis of Responses to Part I

Part I of Question 2 presented data to the students in the form of a simple frequency table. The question is given in Figure 4.3. No specific reference was made to what sort of information students were required to gather from the table. This meant that they were free to respond with as much information as they felt necessary. Response codings are followed by statistical interpretation.

Coded Responses

As with Question 1, it was possible to divide the responses into a number of levels based on the statistical quality of the answer given. Again, three major groups were found and sample responses for the various levels within these groups are presented.

First Group

Responses in this group dealt with only the requirements of the question and there were three broad levels, coded as 0,1 and 2. These responses rationalize the requirements of the question with no real concern for the processing of the data.

Level 0 These responses indicate that the question has not been considered or the requirements were not clear. For example:

(7201) *Its like a graph.*

QUESTION 2
PART I

A class teacher wanted students to practice collecting data. One Year 8 student decided to collect data concerning the number of icecreams that she ate during a week for a seven week period. The table the student came up with is given below.

Week 1	3
Week 2	5
Week 3	7
Week 4	4
Week 5	2
Week 6	7
Week 7	5

What does the table tell you ?

Figure 4.3 - Data Tabulation and Representation - Part I Question

Level 1 These responses indicate that not all aspects of the question have been considered sufficiently to produce an answer. Usually, some key fact from the wording of the question is reproduced in the response. For example:

(7110) *The amount of icecreams this girl ate.*
(7111) *The table tells us about the icecream eating habits of a year 8 girl.*
(12203) *She ate icecreams.*
(12211) *The table tells me how many icecreams she eats.*

Level 2 These responses indicate that the question has been considered and a reasonable answer produced. No use is made of information from the data in the table. At this stage, only some key fact(s) from the wording of the question are mentioned. For example:

(7206) *How many icecreams the student ate in 7 weeks.*
(8213) *The table tells you in column one: what week it was and in column two: how many icecreams were eaten in that week.*

(10110) *The table tells me how many icecreams were consumed (that was close nearly got tricked) - the table (as it has no names for the columns) only tells me what week it is.*

(11108) *The table shows me how many icecreams were eaten per week.*

Second Group

These responses are concerned with attempting to understand the data. Again, three levels were observed within this group, coded as 3, 4 and 5. These responses attempt to describe the data, but only manage to do so in non-statistical terms.

Level 3 These responses indicate that attempts to describe the data have failed and focus has been directed back to the question. All key facts from the question are mentioned (that is, icecreams eaten, per week and seven weeks). For example:

(7209) *How many icecreams she ate each week for seven weeks.*

(11105) *It tells me how many icecreams she ate in 7 weeks, how many she ate each week.*

(12202) *The table specifies how many icecreams were consumed by a student during a weekly period over 7 weeks.*

Level 4 These responses indicate an attempt to understand the data, and an awareness that features of the data need to be mentioned in the answer. However, restricted experiences at data description result in the information in the table being quoted verbatum. For example:

(10106) *The table tells me how many icecreams a year 8 student ate every week e.g. W1 she ate 3 icecreams; W2 she ate 5 icecreams and so on .....*

(12207) *The table tells me that for week 1 the student ate 3 icecreams, in week 2 she ate 5 icecreams, week 3 she ate 7 icecreams etc.*

At this stage there is a divergence of the responses into **two** distinct paths which appear to develop at seemingly parallel rates. These are labelled:

Path A for responses which describe statistical features of the data; and

Path B for responses which make judgements about the data.

Level 5 These responses indicate the data have been understood and features of the data are described by making a simple observation. They suggest readiness to engage in data

description but a lack of experience and appropriate tools to produce a statistically sophisticated response. Responses in Path A describe the data by using a simple numerical description, implying some statistic, and those in Path B make some simple judgement about the data. For example:

Path A

(8214) *She has no strict number of icecreams per week and has a range of numbers.*
(9210) *The table tells me that there is no set pattern in which she ate icecreams.*
(11111) *Some weeks she ate more than other weeks.*
(11215) *That the girl's consumption of icecreams fluctuates.*
(12112) *The number of icecreams that she consumes per week during the 7 weeks varies considerably. She never eats the same number week after week.*

Path B

(7203) *She got hunger off and on.*
(8108) *That the girl is very unhealthy.*
(8203) *She's fat and hungry and likes icecream.*
(10103) *She likes eating a lot of icecream.*
(11210) *The table tells me that the student ate icecreams whenever she felt like it.*
(12102) *That person is addicted to icecreams.*

Third Group

The last group of responses indicate a readiness to describe the information contained in the data and concern at expressing this information in an acceptable statistical form. Only two levels of responses, coded as 6 and 7, were identified with both split into A and B paths. Level 7 also has some responses incorporating elements from both paths.

Level 6 These responses indicate the use of data from the table to make one detailed observation. The Path A responses use a simple statistic to describe the data while Path B responses make a simple judgement linked to the data. These responses show more sophistication than those at Level 5 linking the observations to features of the data, rather than making broad statements. For example:

Path A

(7112) *The table tells you how many icecreams the year 8 student ate each week in the 7 weeks. She ate more than any other week in weeks 3 and 6 etc.*

(8109) *The table shows us how many icecreams the student ate in a weekend we could probably work out roughly how many she ate in a year.*

(8110) *That she ate 33 icecreams in the 7 weeks.*

(10109) *The table tells me that the amount of icecreams eaten varies from 2 - 7 over the 7 weeks.*

(11214) *The number of icecreams eaten each week by the girl and the total number over seven weeks.*

Path B

(8210) *She likes very much icecream. The number of icecreams fluctuates per week so there isn't a standard number per week.*

(9107) *She had a craving for icecreams on the 3rd and 6th week.*

(9201) *The table tells you that she likes icecream for a couple of weeks then she gets back into them again.*

(10105) *That she got a craving in the middle of the week and near the end she likes icecream.*

(11106) *The table tells us that the student enjoys eating icecream on a regular weekly basis.*

(12107) *The year 8 girl recorded how many icecreams she ate in 7 weeks and by the numbers given, she really likes icecream.*

Level 7 These responses indicate a more in-depth understanding by presenting more than one observation related to the data. Those in Path A contain more than one statistic in the description while those in Path B present more than one judgement. Some descriptions containing both statistic(s) and judgement(s) are also represented, after sample responses from Paths A and B. For example:

Path A

(9112) *She ate the most icecreams in week 3 and 6. And the least in week 5. On average she ate 4.7 or 5 icecreams a week.*

(12104) *The data tell you that on average the student eats 4.7 icecreams per week, with a total of 33 icecreams in the seven week period.*

(12108) *The student has a minimum of 2 icecreams a week and a maximum of 7 icecreams a week.*

Path B

(9106) *Week 3, 6 were fairly hot days and he/she felt the need for something cooling.*

(12103) *Sometimes she had a taste for icecreams more than other weeks it could be because of weather.*

(12212) *The student likes icecream or it is summer and she wants to keep cool.*

Responses which are not in Path A or B, data description involving both statistic(s) and judgement(s).

(10212) *The table tells you that the girl ate 33 icecreams in 7 weeks and that she must have liked icecreams.*

(11109) *This table shows you that this student from year 8 eats a considerable amount of icecream weekly. But some weeks the amount eaten is higher. This could also show us that how many eaten depends on the weather.*

(12215) *The data show us how many ice-creams that she ate week by week for a seven week period. It shows us that she ate icecream every week with the highest being 7 and the lowest being 2. It also shows us that their is no regular eating habits when it comes to icecream in this survey.*

These responses are only a sample of all responses given but are representative of the responses at each level. A third level in group three, if it had been found, would have been coded as 8 and would have contained responses which not only mentioned statistic(s) and judgement(s) but used the statistics as evidence for the judgements made. The Levels 5, 6 and 7 have responses split into the two paths, A and B, with Level 7 also containing a third group of students who have combined aspects of both paths. The check on coder reliability showed 100% agreement. Before developing a theoretical perspective in the SOLO Taxonomy Framework section, results of the coding are analyzed according to academic year, mathematical ability and gender.

Statistical Interpretation

The results, arranged by academic year, appear in Table 4.2.1 and a number of interesting points become apparent. First, there are only eleven students (18%) from the

two senior years whose responses fall within the first group, compared to twenty five (42%) from Year 7 and 8. Second, there is a larger number of senior students compared to junior, noticeably twelve in Year 12, whose responses were coded as level 7. Third, there is a large bulge in most years at Level 2. Fourth, there appears to be a larger number of responses in the last level of each of the first two groups (that is Levels 2 and 5), than in the previous two levels of the group. This is more noticeable in the junior years. Last, there appears to be a balance in the number of students whose responses reflect Path A and Path B.

Table 4.2.1 - Response Level and Path by Academic Year - Question 2 Part I

Level	Year												Total
	7		8		9		10		11		12		
0	1		0		0		0		0		0		1
1	4		1		0		1		0		2		8
2	14		5		10		11		6		3		49
3	3		3		3		2		7		2		20
4	0		5		0		3		0		2		10
	A	B	A	B	A	B	A	B	A	B	A	B	
5	1	5	3	6	4	3	2	4	4	2	3	3	40
6	1	0	3	2	1	4	2	2	2	3	1	2	23
7	1	0	1	0	3	1	1	0	5	0	4	3	
	0		1		1		2		1		5		29
Total	30		30		30		30		30		30		180

Generally, there appears to be a plateau reached in each of the first two groups and, although there is overall a large bulge at Level 2, there is an improvement in level in the senior years. There are similar numbers of students giving responses in Path A and Path B indicating that students appear equally likely to choose to describe the data by using statistical features or making judgements.

The results, arranged by mathematical ability, are presented in Table 4.2.2 for response level and Table 4.2.3 for response path. These show that both the coding level ($\chi^2 = 7.21$, 10 d.f., $p > 0.50$) and the path ($\chi^2 = 4.85$, 2 d.f., $p > 0.05$) are independent of mathematical ability.

Table 4.2.2 - Response Level by Mathematical Ability - Question 2 Part I

	Mathematical Ability			
Level	Low	Middle	High	Total
0-1	4	1	4	9
2	18	16	15	49
3-4	11	8	11	30
5	15	12	13	40
6	5	12	6	23
7	7	11	11	19
Total	60	60	60	180

Table 4.2.3 - Response Path by Mathematical Ability - Question 2 Part I

	Mathematical Ability			
Path	Low	Middle	High	Total
A	8	18	16	42
B	16	15	9	40
Total	24	33	25	82

The results, arranged by gender, are presented in Table 4.2.4 for response level and Table 4.2.5 for response path. Both the coding level ($\chi^2 = 7.73$, 6 d.f., $p > 0.20$) and path ($\chi^2 = 1.23$, 1 d.f., $p > 0.20$) are independent of gender.

It appears that when attempting to understand data presented in a frequency table the level of response improves slightly with academic year, but similar numbers of students tend to use statistics or judgements to describe the data. Neither the mathematical ability nor the gender of the student appears to significantly affect the performance of this task.

Table 4.2.4 - Response Level by Gender - Question 2 Part I

	Gender		
Level	Female	Male	Total
0-1	5	4	9
2	25	24	49
3	8	12	20
4	8	2	10
5	15	25	40
6	13	10	23
7	16	13	29
Total	90	90	180

Table 4.2.5 - Response Path by Gender - Question 2 Part I

	Gender		
Path	Female	Male	Total
A	23	19	42
B	17	23	40
Total	40	42	82

Analysis of Responses to Part II

The data are presented in graphical form in the second part, II, of Question 2. The question was designed to determine whether the form of presentation of the data affected students' understanding when describing data. The question, is given in Figure 4.4. Statistical interpretation of the codings follows.

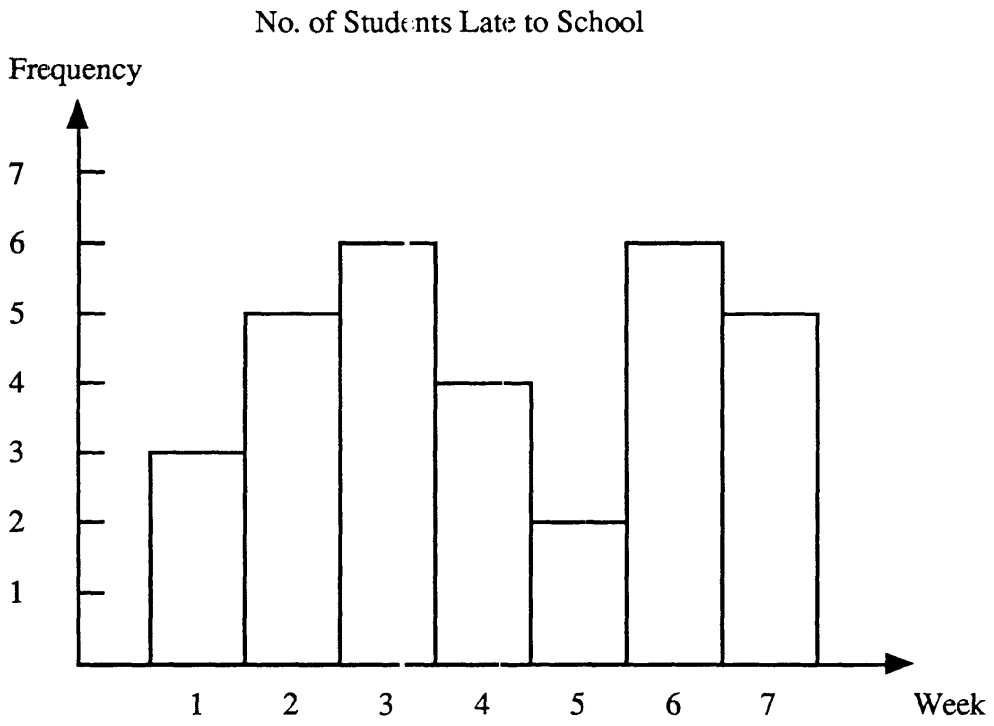
Coded Responses

As with Part I, the question was designed with no specific references. However, answering this question meant that students needed to be able to understand and

interpret the graph before they were able to describe the data. Responses to Part II fell into similar levels to those found for Part I. Some responses expressed differently to those of Part I are presented in Appendix E.

QUESTION 2
PART II

The deputy in the school kept a record of the number of students who were late to school each week. He decided it would be useful to draw a graph to illustrate the information. The graph is presented below.



What does the graph tell you ?

Figure 4.4 - Data Tabulation and Representation - Part II Question

Statistical Interpretation

The results, arranged by academic year, are presented in Table 4.2.6 and some interesting features emerge. First, while twenty one (35%) from the two senior years gave responses in the first group, having not apparently understood the question, there were thirty four (57%) from the Years 7 and 8. Second, Level 7 contained thirteen Year 11 and 12

students but only three Year 7 and 8 students. Third, there appears to be a large number of responses in the last level of each of the first two groups (Levels 2 and 5) than in the previous two levels of each group. Fourth, there were no students in Level 0 or in Level 6 in Path B. Last, there are almost twice as many students coded in Path A as Path B, with Path A more popular in all but Year 7 and Year 10. These observations suggest a slight improvement in the quality of responses for Years 7 through to 12 and a preference for Path A.

Table 4.2.6 - Response Level and Path by Academic Year - Question 2 Part II

Level	Year												Total
	7		8		9		10		11		12		
0	0		0		0		0		0		0		0
1	8		7		1		0		2		3		21
2	11		8		11		14		12		4		60
3	2		1		3		4		4		1		15
4	0		0		0		1		0		2		3
	A	B	A	B	A	B	A	B	A	B	A	B	
5	0	7	3	3	5	3	0	7	4	2	7	3	44
6	1	0	6	0	5	0	4	0	1	0	2	0	19
7	1	0	1	0	0	2	0	0	4	0	6	2	
	0		1		0		0		1		0		18
Total	30		30		30		30		30		30		180

The results, arranged by mathematical ability, are presented in Tables H.2.1 and H.2.2 (Appendix H) for response level and path, respectively. Coding level proved to be independent ($\chi^2 = 12.41$, 10 d.f., $p > 0.20$) of mathematical ability. However, there was strong evidence to support a dependence ($\chi^2 = 9.60$, 2 d.f., $p < 0.01$) of coding path on mathematical ability. Low mathematical ability students gave more Path B responses than expected, and for the middle and high mathematical ability students Path A was more popular. The results, arranged by gender, are presented in Tables H.2.3 and H.2.4 for response level and path. Both coding level ($\chi^2 = 7.24$, 5 d.f., $p > 0.20$) and path ($\chi^2 = 0.001$, 1 d.f., $p > 0.95$) are independent of gender.

For the tabulation and representation question, with data presented in graphical form, the level of response improves slightly with academic year and more students

elect Path A than Path B in most years. The gender of the student does not appear to influence the level of the response, but students with a lower mathematical ability have more of a tendency to describe the data using judgements rather than statistics.

Comparison of Part I and Part II

In the preceding sections, the analysis and discussion of Part I and II clearly showed that the groupings and the levels are similar for both parts of the data tabulation and representation question. The performance of the responses for the two parts of the question are compared in this section, mirroring the structure established previously.

Overall Comparison

All responses could be allocated to one of the levels suggesting that the framework is adequate for explaining students' understanding, as far as the basic concepts of data tabulation and representation are concerned. There is a slight upward shift in the trend of the responses over the academic years as would be expected. All but one student felt that they understood the question sufficiently to attempt an answer. Even at Level 1, which included students who had misinterpreted the graph, there were not many students, most of them being in Years 7 and 8. Comparing the results for the two parts of the question, assists with highlighting trends and differences.

Table 4.2.7 shows the number of responses that were graded in each level for Parts I and II. There is evidence ($\chi^2 = 13.53$, 6 d.f., $p < 0.05$) that the level of a response is associated with the part of the question being answered. Referring back to Table 4.2.7 suggests that many more responses than expected were coded at the lower levels in Part II, while Part I had more responses than expected in the uppermost levels. This suggests that students exhibit a higher level of understanding when the information is presented in a table rather than as a graph. Table 4.2.8 shows the number of responses that were graded into each path for Parts I and II. Evidence ($\chi^2 = 2.39$, 1 d.f., $p > 0.10$) suggests that the choice of path is not associated for the part of the question being answered. This means that the paths used by students in Part I do not indicate which paths will be used in Part II.

Four noticeable trends emerge when comparing the results of the analysis of academic year, presented in Tables 4.2.1 and 4.2.6. First, there are more Year 7 and 8 students in the first group (Levels 0, 1 and 2) than there are Year 11 and 12 students. Second, there are more senior students than junior students in Level 7. Third, there is a large bulge in the numbers at Level 2 in many years. Fourth, there is a lack of responses at Level 0, in fact only one, a Year 7 student in Part I. The general trend is for a slight increase in the level of performance of the students over the six academic years.

Table 4.2.7 - Level Comparison for Parts I and II - Question 2

	Question 2		
Level	Part I	Part II	Total
0-1	9	21	30
2	49	60	109
3	20	15	35
4	10	3	13
5	40	44	84
6	23	19	42
7	29	18	47
Total	180	180	360

Table 4.2.8 - Path Comparison for Parts I and II - Question 2

	Question 2		
	Part I	Part II	Total
Path A	42	50	92
Path B	40	29	69
Total	82	79	160

The unexpectedly large bulge of responses at Level 2 is the point at which students repeat a number of key facts from the question but do not use the data in any way in their description. This occurs in Years 7, 9 and 10 in Part I and Years 7, 9, 10 and 11 in Part II. More Year 8 students chose to produce a better answer by referring to the data and this may have been due to the simple statistic and graph work that is covered in the mathematics course and so is fresh in their minds The better results for Year 12 could mean that by this stage most students are really ready to describe the data in more detail, despite the fact that little statistical work is done with senior students, except in Mathematics in Society and Mathematics in Practice Courses.

Comparing the results presented in Tables 4.2.1 and 4.2.6 there are four noticeable differences. First, the number of Year 11 and 12 students on Level 0, 1 or 2 is

much larger for Part II (21) than for Part I (11). Second, the overall number of responses at Level 7 is much larger for Part I, (29), than for Part II (18). Third, the Year 11s performance on Part II of the question is much poorer than for Part I. These differences suggest that although there is an increasing level with academic year, the overall range of performance is better when data are presented in table form than as a graph. Lastly, there is a slight difference in the way that the choice of path varies with academic year. For Part I, similar numbers of responses are categorized as Path A or Path B in all years. However, in Part II in Year 7 there appear to be many more responses in Path B while in the other years there is a predominance of Path A type responses. This suggests that when Year 7 students are presented with data in graphical form they are more likely to use judgements than statistics in their descriptions. When older students are presented with graphical data they are more likely to use statistics in their descriptions.

Varying mathematical ability appears to produce similar affects in both parts of the question. Comparing Tables 4.2.2 and H.2.1 indicates that in neither part of the question does the mathematical ability of the student affect the level of the response. However, comparing Tables 4.2.3 and H.2.2 suggests that mathematical ability does affect the path into which a response is coded for Part II but not for Part I. When the data are presented in graphical form, more middle and high mathematical ability students prefer to describe the data using statistics than judgements. The gender of the student does not appear to influence the level or the path of the response for either part of the question. This is evident from a comparison of Tables 4.2.4 and H.2.3 and Tables 4.2.5 and H.2.4.

It appears that presenting data in graphical form has influenced the students' ability to describe the data. Not only is there a general reduction in the level of understanding of the data, but there is also a difference in the way that students approach the problem of describing data. Perhaps students need to have more experiences which involve describing data, especially data which are presented in graphical form.

From the above, it appears possible that the hierarchy of levels developed can be used to categorize responses. There appears to be two possible paths of reasoning in the second and third groups and the form in which the data are presented may influence a student's choice of path. These levels are used in a later section to construct a theoretical framework for describing student understanding with respect to data tabulation and representation.

Comparison of Student Performances

The performance of students over the two parts of the question is now considered. First, the levels at which the student responded are compared. Then, the paths along which the students' responses diverged are compared.

Part I and Part II student response levels are presented in Table H.2.5. As with Question 1, information had to be condensed using each student's grouping rather than level, see Table 4.2.9. There is strong evidence (χ^2 = 49.84, 4 d.f., $p < 0.001$) of an association, for each student, between the response grouping and the part of the question being answered. The data in Table H.2.5 suggest that there is a decrease in level from Part I to Part II. Using the Nonparametric Sign Test, of the 116 non-zero differences, 71 were positive. There is evidence ($Z = 2.32$, $p < 0.02$) to suggest that the coding levels are different for the two parts of the question and as there were more positives than negatives it can be assumed that the students tended to attain a higher coding on Part I of the question than on Part II. This suggests that a student is likely to demonstrate a lower level of understanding when data are presented in graphical form.

Table 4.2.9 - Group Comparison within Students - Question 2

		Q2 Part I			
		First	Second	Third	Total
Q2 Part II	First	46	24	11	81
	Second	9	33	20	62
	Third	3	13	21	37
	Total	58	70	52	180

Although not anticipated, it appears that there is a divergence of responses into two distinct paths. This leads to the consideration of whether there is a link between the Path, A or B, taken for the two parts of the question. For each part of the question the students were categorized as to whether their responses fell into one of the following four categories:

Before - Responses in Levels 0, 1, 2, 3 or 4 (that is before the split into paths);

A - Responses in Levels 5, 6 or 7 which followed Path A ;

B - Responses in Levels 5, 6 or 7 which followed Path B; and

After - Responses in Level 8 or in level 7 but not showing any indication of being specifically in Path A or B.

The results are presented in Table 4.2 10.

Table 4.2.10 - Path Comparison within Students - Question 2

		Q2 Part I				
		Before	A	B	After	Total
	Before	72	13	10	4	99
Q2	A	7	25	13	5	50
Pt.	B	8	4	15	1	28
II	After	1	0	2	0	3
	Total	88	42	40	10	180

There is strong evidence (χ^2 = 72.27, 9 d.f., $p < 0.001$) to support an association, for each student, between the path chosen and the part of the question being answered. This suggests that a student's choice of path in Part I can be used as an indicator of the path which will be chosen in Part II. In fact, of the 57 students who were in one of the response paths for both parts of the question, 40 (70%) students followed the same path in both parts, while only 17 (30%) followed different paths. It appears that a student is likely to give responses in the same path for both parts of the question. If a student describes data, presented in a table, using statistics then a similar approach is expected if the data are presented as a graph. Similarly, if a student is inclined to describe data using judgements, then this approach is used whether the data are presented in a table or in a graph.

Students are not as good at describing data when it is presented as a graph rather than as raw data. However, the inclination to use statistical features or make judgements when describing data is similar, irrespective of the form of presentation.

SOLO Taxonomy Framework

Having established that the responses could be classified using these levels, the next step was to use these levels and the structure of the SOLO Taxonomy in an analysis which mirrors that of Question 1. Response codes followed by an E, can be found in Appendix E.

Ikonic mode responses show no evidence of linking the required task with any sort of symbolic representation. Prestructural responses appear as Level 0, only one Year 7 student was unable to attempt an answer. Level 1 responses are a mixture of unistructural and multistructural responses. For the same reasons as with Question 1, it was difficult to separate these responses into separate unistructural and multistructural levels. These

responses range from those students (such as 7201,E) who misinterpret the graph, to those students who reiterate a key fact from the question, for example, one student (7110) stated that the table tells the amount of icecreams the girl ate. Finally, the relational level within this mode, corresponds to Level 2 responses. These responses indicate that the various aspects of the question have been incorporated and the question understood sufficiently to enable the proposition of a reasonable answer. These responses only present key facts from the question and do not relate to the data. They range from a student (11108) who said that the table shows how many icecreams were eaten per week, to a student (8213,E) who specified what was along each axis of the graph (that is what week it was and the frequency of students late to school).

Second and third group responses have been able to link the concepts in the question to concrete experience. The answer, which suggests that the question has been understood, links directly to aspects of the question or the data. These responses are in the concrete-symbolic mode, with two cycles, each containing the U, M and R levels.

The first cycle involves appreciating that it is possible to describe data. The elements in the first cycle are the actual pieces of data themselves (data items). A relational response in the first cycle is not achieved until the student is able to consider all data items as a functioning set, and the features as capable of being described in another form. As with Question 1, the unistructural, multistructural and relational levels correspond to the Levels 3, 4 and 5.

U1 - These responses indicate that the table or graph has been considered and all the key aspects of the data understood. A concise description is given, which included all key features of the question. Typical responses are from a student (7209) who mentioned "how many icecreams she ate, each week and for seven weeks" and another (9205,E) who stated 'the number of students late to school, each week and over a seven week period".

M1 - These responses indicate that attempts to focus in on features of the data have failed and all data are repeated as a description. Typical responses are those of a student (10106) who listed how many icecreams were eaten in weeks 1 and 2 and then adds "and so on" and another student (10210,E) who detailed every single week with the number of students late.

R1 - These responses present a simple observation which has combined various pieces of data in the description. This is the stage where there appears to be a split into two separate Paths, A and B, for the preferred styles of processing. Processing along Path A gives an observation which is a numerical description of the data while processing along Path B tends to make a personal judgement of the data as an observation. Typical responses in Path A are those of a student (12205,E) who said

there were always people late to school and another student (11215) who stated that the girl's consumption of icecreams fluctuates. Typical responses in Path B are those of a student (7203) who said the girl "got hunger off and on" and another (9102,E) who classified as "big" the number of students late each week.

Second cycle responses show an appreciation for the need to refer to the features or behaviour of the data as part of the description. The elements in the second cycle are the various features (or properties) of the data which a statistical description should include. A relational response in the second cycle is not achieved until the student is able to present statistical facts and judgements and also some relation between them in an the overall data description. In this cycle, the unistructural and multistructural levels correspond to the Levels 6 and 7 as outlined earlier. A relational response would be equivalent to a Level 8 response, but no such responses were observed in the data. The first two levels still contain the separate A and B processing paths. However, there are also responses at the M2 level which show evidence of elements from both paths.

U2 - These responses indicate a simple observation, made either in the form of a statistic or a judgement, which demonstrates some feature of the data. Path A responses mention one simple statistic and range from a student(10108,E) who stated what weeks most people were away, to a student (8110) who gave the total number of icecreams eaten, to a student (10109) who stated that the icecreams eaten vary from 2 - 7 in the seven weeks. Path B responses mention one simple judgement and range from a student (10105) who described a "craving" for icecream in the middle and near the end, to another student (11106) who said that the girl enjoys eating icecream on a regular weekly basis.

M2 - These responses indicate that more than one aspect of the data has been processed and a number of statistic(s) and/or judgement(s) are given in an attempt to more accurately describe the data. Responses in Path A mention a number of statistics, for example a student (12108) gave the minimum and maximum number of icecreams eaten and another (12101,E) discussed low values, when these doubled and when increases and decreases occur. Responses in Path B mention a number of judgements, for example, a student (12103) stated that the girl has more taste for icecreams on some weeks and it may be due to the weather, and another student (9110,E) stated that there is a wide range of people being late and that a lot of students are late quite often. Responses at the M2 level which incorporate elements from both paths are typified by a student (8113,E) who said that a lot of students come late and also that the most were away in weeks 3 and 7, and another student

(10212) who stated that the girls ate 33 icecreams in the week and also that she must have liked icecreams.

R2 - These responses present information concerning statistic(s) and judgement(s) and some attempt to use the statistic to qualify the judgement. It is at this stage that processing is no longer on either Path A or Path B, but integrates the features of the data. No response was coded at this level but some in the combined path of the R2 level were very close to being relational. These responses would need to indicate that the statistics created using the data have been used to justify any judgements that have been made.

This framework is useful in determining the stage a student has reached in understanding data representation and tabulation. The main feature which distinguishes the concrete-symbolic mode responses from those in the ikonic mode is the retrieval of facts from the recorded data. Ikonic mode responses go no further than recognizing the variables which are being measured. Concrete-symbolic mode responses show that the data items have been considered. Within this mode, the first cycle responses suggest that the data items have been considered as separate items while second cycle responses indicate an overview of the data, in the form of a statistic or judgement.

Data Reduction

Data Reduction was the third area of statistics investigated. This area involves the reducing of data into one of four types of statistical form, namely: measures of central tendency, including mean, mode and median; measures of dispersion, including range, quartiles, interquartile range, mean deviation, variance, standard deviation and dispersion; other summarizing statistics including, index, crude rates, death rates and birth rates; and, regression and correlation.

The data reduction question, Question 3, was designed to help clarify students' understanding of the condensing of data into a statistical form. This is one of the earlier steps necessary in the process of analysing data. The aim of the question was to present students with some data and then have them reduce that data into a more usable form. No reference to any specific statistics was made in order to view what students perceived as necessary steps in data reduction, and to allow them to use whatever facilities they had available and felt were suitable for the task. As with Question 2, the data were presented in two different forms, as raw data in Part I and as a graph in Part II. This was done to investigate whether the form of data presentation influenced the way in which students reduced the data. The analysis of the responses to Question 3, considered in this section, mirrors the detail provided for Question 2.

Analysis of Responses to Part I

The data were presented in a raw form for the first part of Question 3. This open-format question made no reference to any measure of central tendency or dispersion, allowing students to make use of any measure they felt was appropriate. The question is given in Figure 4.5. Response coding description is followed by an analysis of the codings.

QUESTION 3
PART I

As part of a large project which had to do with measuring and discussing the human body, one of the tasks was to measure the lengths of peoples' feet to the nearest centimetre. The results of the 29 students in the class are as follows:

26	26	26	27	27	27	27	28	28	28	28	28	28
29	29	29	29	29	30	30	30	30	30	30	30	31
32	32	33										

(i) If you were asked to give a number, or numbers, (to the nearest cm.) which could be best used to represent the size of the left feet in that class, what numbers, or numbers, would you select ?
(ii) Give reasons for your selection.

Figure 4.5 - Data Reduction - Part I Question

Coded Responses

Upon investigating students' responses it was possible to divide them into a number of levels based on the statistical quality of the answers given. As before, three major groupings were identified.

First Group

In this group, with three levels were coded as 0, 1 and 2, contains responses which dealt with only the requirements of the question. These responses indicate consideration only of the requirements of the question with no use of the data in formulating

the response. The responses have been recorded with the answer to part (i) indicated as such and similarly part (ii).

Level 0 These responses indicate that the requirements of the question were not considered or an answer could not be attempted. For example:

(7101)	(i)	*I don't know.*
(9209)	(i)	*I haven't got the faintest idea. I haven't learnt it.*

Level 1 These responses indicate an attempt at answering the question but either it is a nonsense answer or a reason is given which does not answer the question asked. For example:

(7102)	(i)	*I would select all of the odd numbers because the left foot is the same as the left hand most people are right handed.*
(7214)	(i)	*About 4 meters (sic).*
	(ii)	*don't know*
(8201)	(i)	*Numbers.*
	(ii)	*Because it is a lot of numbers to choose from.*
(12204)	(i)	*I would take cm and always round up when it is X,5 and down when it is X,4.*
	(ii)	*I would say it is the most common way to do it so people all over the world would understand it.*

Level 2 These responses indicate a reasonable attempt at answering the question but there is either no explanation of how the answer was obtained, or an explanation is given which is not related to the data or question. At this stage, explanations, if given, resort to personal experience. For example:

(8115)	(i)	*I would be particular, but close my eyes and put my pen down on a number.*
	(ii)	*It is a random selection and easy to do.*
(8202)	(i)	*26, 27, 28, 29*
	(ii)	*you don't see many people with that sized feet.*
(8207)	(i)	*29 cm*
	(ii)	*I don't know.*
(9205)	(i)	*26 to 30*
	(ii)	*Because these are the two numbers I felt like selecting.*

Second Group

Responses in this group show an understanding of the question and attempt to rationalize the reduction of the data. These attempts to process the data are hampered by the lack of experiences and tools for reducing data.

Level 3 These responses indicate that, in attempting to justify a reasonable estimate or estimates, the reason did not refer to a feature of the data. Instead, the focus was on some feature of the question itself. For example:

(7213)	(i)	*30 cm*
	(ii)	*If you rounded them up to the nearest 10 cm you would get 30 cm.*
(10201)	(i)	*28 and 30*
	(ii)	*Because that would tell you how big the foot was to the closest cm.*

Level 4 These responses indicate that the data were used to obtain a reasonable estimate or estimates and an awareness that the data needed to be used to justify this answer. However, restricted experiences at data reduction result in all data being quoted as necessary in the reason. For example:

(11111)	(i)	*26, 27, 28, 29, 30, 31, 32, 33*
	(ii)	*every childs foot is the length of one of these numbers*
(11201)	(i)	*I would use the numbers 26 to 33.*
	(ii)	*This covers all the sizes of the left feet.*

At this stage, as with Question 2, there is a divergence of the responses into two distinct paths which appear to develop at seemingly parallel rates. These are labelled:

Path A for responses which reduce data based on measures of central tendency; and

Path B for responses which reduce data based on measures of dispersion.

Level 5 These responses indicate that the data could be reduced to a simple statistic. However, attempts to justify the answer usually result in an unsophisticated description of the statistic constructed. This indicates a readiness to engage in data reduction but a lack of experience with the tools needed to produce a statistically sophisticated response. For example:

Path A

(8209)	(i)	*I would select numbers 28, 29, 30*
	(ii)	*They are approximately the size of the students feet in the class.*
(10202)	(i)	*28 to 30*
	(ii)	*It is the average size for the students.*
(11215)	(i)	*29 cm*
	(ii)	*Because 29 is the average foot size.*
(12208)	(i)	*29*
	(ii)	*It is the average number (middle number)*

Path B

(7203)	(i)	*26 - 33*
	(ii)	*so you don't have to write so much*
(11211)	(i)	*The numbers, used to represent the size of the left feet in the class, that I would choose would be 26, 27, 28, 29, 30, 31, 32, 33.*
	(ii)	*The reason I chose these numbers is that if I put down one of every size measured it would give you an indication of the range (from shortest to longest).*

Third Group

These responses indicate a readiness to reduce data into an acceptable statistical form and attempt to relate explanations back to the data. The first two levels, 6 and 7, are split into Paths A and B. Some responses in Level 7 also incorporate elements from both paths.

Level 6 These responses indicate that the data could be reduced to the form of a simple statistic and the reason for making such a selection related back to the data. However, the reason stated usually reflects a rote-learnt definition rather than a true appreciation of the data. For example:

Path A

Mode -

(10205)	(i)	*28*
	(ii)	*Reason is because 28 happens the most.*
(11107)	(i)	*28 & 30*

	(ii)	*Because these 2 sizes are the 2 most common sizes in the results.*
(11206)	(i)	*30 cm because this is the mode left foot size.*
	(ii)	*It is the mode.*
Median -		
(10209)	(i)	*29*
	(ii)	*It is the median*
(11208)	(i)	*30*
	(ii)	*In the middle.*
Mean -		
(10214)	(i)	*28 and 29*
	(ii)	*The mean was 28.8*
(10215)	(i)	*28.86*
	(ii)	*because it is the mean of all the feet*
(12113)	(i)	*I would give all the numbers and divide by the number given to give an average (to nearest cm) size of the students feet.*
	(ii)	*I would do it this way because I believe it would be the fairest way to find the average size because the more numbers you take the closer and more accurate the survey will be.*

Path B

(8210)	(i)	*29 to 30 cm*
	(ii)	*because most feet are around that size.*
(11103)	(i)	*26 & 33*
	(ii)	*I gave the shortest & longest numbers to show the variation.*
(12201)	(i)	*26 - 33*
	(ii)	*That is the range.*

Level 7 These responses indicate the desire to present more information by discussing more than one statistic. Reasons given still reflect more the definitions of the statistics rather than discussions of properties of the data. Some responses discuss concepts which relate to statistical measures from the other path, for example, a discussion of measures of central tendency may also mention some aspect of measures of dispersion. This is the stage where the two paths are beginning to converge again. For example:

Path A

(8106)	(i)	*28 & 29*

	(ii)	*Because the no 28 is the most frequent one & also it is ruffly halfway between 26 & 33.*
(9108)	(i)	*I would chose the median or the mode of the feet*
	(ii)	*Because thats all I could think of*
(9114)	(i)	*30, 29*
	(ii)	*29, 30 is approximately the mean, median and mode of the numbers.*
(11106)	(i)	*The numbers I would select to represent the size of the left feet in that class would be 28 and 30*
	(ii)	*because they have the highest number in the class and those two are the average of them all.*
(11213)	(i)	*29 cm.*
	(ii)	*I would select this because it seems to be the average length of the people in the class. Also it is nearly the middle number.*
(12207)	(i)	*30 cm*
	(ii)	*because it worked out to be the average and also the most number of people in the class had their left foot 30 cm.*

Path B

(9101)	(i)	*I don't really understand this question but I'll say the top one the bottom one (length) and the middle one*
	(ii)	*it gives a fair idea of the range in sizes.*

Responses which are not in Path A or B, data reduction involves the use of measures of both central tendency and dispersion.

(10106)	(i)	*I would use 28 and 30.*
	(ii)	*Because most of the classes(sic) left feet were around that size and plus the sizes of 28 and 30 were most common.*
(10213)	(i)	*I would give the numbers 28 and 29*
	(ii)	*I would give these numbers as there are about the middle of the class and you can have a deviation of 3.*
(11108)	(i)	*26, 27, 28, 29 30, 31, 32, 33 or an average*
	(ii)	*To show the range and the average.*
(12210)	(i)	*The mean or the average size of the left feet would represent the whole class.*

(ii) *The average gives an indication of the size foot of every student in the class. The standard deviation from the mean would also be useful in representing the class as this takes into consideration every student.*

Level 8 These responses mention both measures of central tendency and dispersion and use features of the data in an attempt to establish a link between the two. For example:

(12209) (i) *29*

(ii) *It is the middle number. The answers fluctuate a little to the negative and the positive sides of the number 29 (but mostly range around it).*

(8215) (i) *I would select either the mode 30 or the median 29*

(ii) *The mode occurs most often and would be the most common foot length in the class, 29 is the middle of the range in foot sizes and as such should come close to most of the non-extreme values.*

It was possible to grade all responses into one of the nine levels presented with some levels splitting into two paths, as occurred in Question 2. The check on coder reliability showed a 97% agreement. The one non-agreeing coding, out of thirty, was only one level from the researcher's coding and within the same grouping.

Statistical Interpretation

The results, arranged by academic year, are presented in Table 4.3.1 and illustrate a number of interesting points. First, there are only two students (3%) from Years 11 and 12 whose responses fall within the first group (Levels 0, 1 and 2), whereas in Years 7 and 8 there are twelve students (27%). Second, of the four Level 8 responses, three were in Year 12 and only one in Year 8. Third, there is a large bulge in all years at Level 6. Last, there are at least three times as many students whose responses reflect Part A than Path B, in each academic year. These observations suggest a slight improvement in response level with increasing academic year. Perhaps there is a tendency to feel satisfied once a simple statistic has been used in the data reduction process. There is a definite preference for measures of central tendency rather than dispersion when reducing data, irrespective of academic year.

The results, arranged by mathematical ability, for response level and path, are presented in Tables H.3.1 and H.3.2. respectively. Strong evidence ($\chi^2 = 22.15$, 6 d.f., $p < 0.01$) suggests that coding level depends on mathematical ability. In the first group there are more low mathematical ability and less middle and high mathematical ability students than expected, while in the Levels 7 to 8 there are far less low mathematical ability students than expected. However, the coding path of any particular response is independent ($\chi^2 = 3.00$, 2

d.f., $p > 0.20$) of mathematical ability. The results, arranged by gender, for response level and path, are presented in Tables H.3.3 and H.3.4, respectively. Both the coding level ($\chi^2 = 3.42$, 3 d.f., $p > 0.30$) and path ($\chi^2 = 1.23$, 1 d.f., $p > 0.10$) are independent of gender.

Table 4.3.1 - Response Level and Path by Academic Year - Question 3 Part I

Level	Year												Total
	7		8		9		10		11		12		
0	1		2		2		0		0		0		5
1	2		2		0		0		0		1		5
2	1		4		1		0		0		1		7
3	1		0		0		1		0		0		2
4	1		0		2		2		3		1		9
	A	B	A	B	A	B	A	B	A	B	A	B	
5	2	1	6	0	4	0	5	0	2	1	5	1	27
6	16	2	10	2	13	3	13	3	14	3	9	4	92
7	2	0	3	0	3	1	3	0	5	0	2	0	
	1		0				3		2		3		29
8	0		1		0		0		0		3		4
Total	30		30		30		30		30		30		180

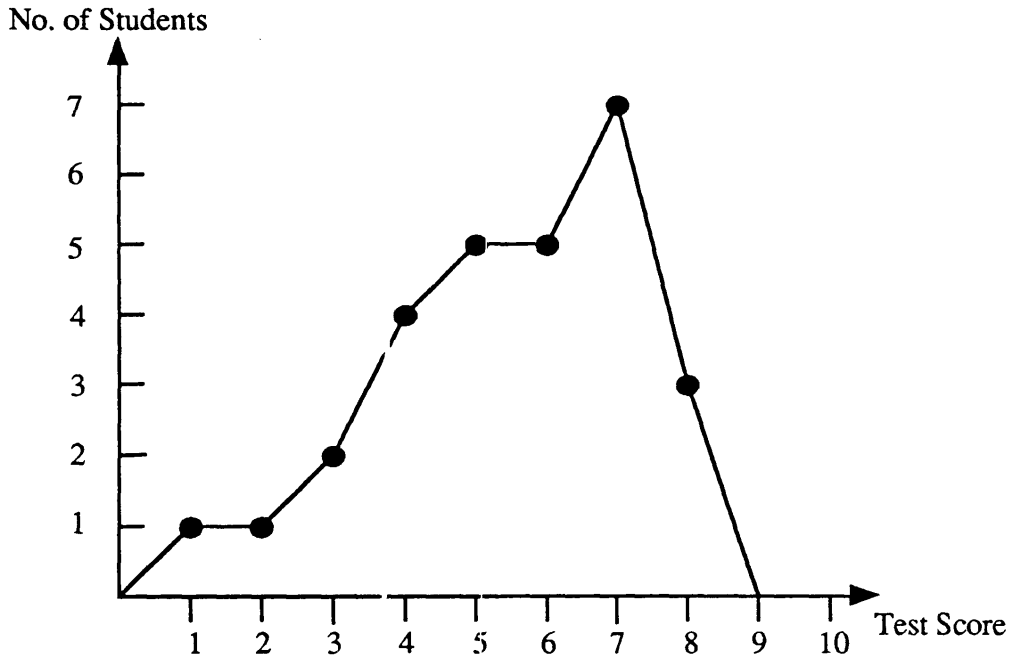
These results suggest that for the process of data reduction, given data in raw form, the level of response improves progressively with academic year and many more students use measures of central tendency rather than dispersion to describe the data. The gender of a student does not appear to affect the response given but students with a higher mathematical ability are more likely to give a higher level response.

Analysis of Responses to Part II

Part II of Question 3, with the data presented in a graphical form, was also designed so that no reference was made specifically to any measure of central tendency or dispersion. However, answering this question meant that students needed to be able to understand and interpret the graph, before they were able to engage in data reduction. The question is given in Figure 4.6. Next is an analysis of the response codings.

QUESTION 3
PART II

A teacher was interested in how students performed in a spelling test to decide whether they needed extra help. The graph below represents the scores out of 10 achieved by the 28 year 9 students in the class in the spelling test.



(i) If you were asked to give a number, or numbers, which could be best used to represent the score in the spelling test of students in that class, what numbers, or numbers, would you select ?

(ii) Give reasons for your selection.

Figure 4.6 - Data Reduction - Part II Question

Coded Responses

Part II responses fell into similar levels as for Part I, except that no Level 3 responses were identified. Sample responses, appear in Appendix F.

Statistical Interpretation

The results, arranged by academic year, are presented in Table 4.3.2 and some interesting points can be observed. First, there are less students, only two (1%), from the two senior years who responses fall within the first group, than from Years 7 and 8 where there are thirteen students (7%). Second, there was only one student, from Year 10, whose response was coded as Level 8. Third, there were no students in Level 3 and only one in Level 4. Fourth, there is a large bulge at Level 6 in all years. Last, there are approximately three times as many responses in Path A as in Path B. Although there are less seniors in the lower groups, there is not much indication of an increase of level with academic year and, irrespective of academic year, many more students favour Path A than Path B.

Table 4.3.2 - Response Level and Path by Academic Year - Question 3 Part II

Level	Year												Total
	7		8		9		10		11		12		
0	1		4		3		2		0		0		10
1	2		3		2		2		0		1		10
2	2		1		0		5		1		0		9
3	0		0		0		0		0		0		0
4	0		0		1		0		0		0		1
	A	B	A	B	A	B	A	B	A	B	A	B	
5	2	0	6	0	5	3	2	1	4	4	4	3	34
6	16	3	13	0	11	4	13	3	18	1	8	7	97
7	1	3	1	1	1	0	1	0	1	0	3	0	
	0		1		0		0		1		4		18
8	0		0		0		1		0		0		1
Total	30		30		30		30		30		30		180

The results arranged, by mathematical ability, are presented in Tables H.3.5 and H.3.6. There is strong evidence ($\chi^2 = 21.90$, 6 d.f., $p < 0.01$) to suggest that coding level depends on mathematical ability. In the first group there are more low mathematical ability and less middle and high mathematical ability students than expected, while in the Levels 7 to 8 there are far less low ability students than expected. However, the coding path

of any particular response is independent (χ^2 = 1.45, 2 d.f., p > 0.30) of mathematical ability.

The results, arranged by gender, are presented in Tables H.3.7 and H.3.8. There is evidence (χ^2 = 10.37, 3 d.f., p < 0.02) to suggest that the level of the response depends on gender, with more males giving lower level responses and more females in the higher levels. As before, the path of any particular response is independent (χ^2 = 0.01, 1 d.f., p > 0.90) of gender.

For the process of data reduction, given data in a graphical form, the level of response improves progressively with academic year and many more students use measures of central tendency rather than dispersion to describe the data. This preference for the use of measures of central tendency does not appear to be affected by either the gender or the mathematical ability of the student. However, higher level responses tend to be given by females and students with a higher mathematical ability.

Comparison of Part I and Part II

In the preceding sections, the analysis of Parts I and II showed that the groupings and the levels are similar for both parts of the data reduction question. The performance of the responses for the two parts of the question are compared in this section and the analysis mirrors the structure for Question 2.

Overall Comparison

The framework appears to be adequate in explaining students' understanding as far as the basic concepts of data reduction are concerned, as it was possible to allocate all responses to a level. The slight upward shift in level over the academic years, as expected, suggests the suitability of the levels. Some students appear to have found this question difficult with many students being able to provide either no answer or a personal response. However, this group includes those who could not interpret, or misinterpreted, the graph. Trends and differences are now considered by comparing the numbers of students within each level and path.

Response levels for Parts I and II appear in Table H.3.9. There is evidence (χ^2 = 16.62, 6 d.f., p < 0.02) that the response level is associated with the part of the question being answered. More responses than expected were coded at the lower levels in Part II, while Part I had more responses than expected in the uppermost levels. The students exhibited a higher level of understanding when the information was presented in a table rather than as a graph. Table H.3.10 shows the number of responses that were graded into each path for Parts I and II. The paths into which responses are coded are not associated (χ^2 = 3.28, 1 d.f., p > 0.05) with the part of the question being answered.

Analysis with respect to academic year, presented in Tables 4.3.1 and 4.3.2 shows four noticeable trends. First, there are more Year 7 and 8 students in the first group (Levels 0, 1 and 2) than there are Year 11 and 12 students. Second, there are mainly senior students in Levels 7 and 8. Third, there is a large bulge in the numbers at Level 6 in every year. Fourth, there are many more Path A responses than Path B in every year. The general trend is for a slight increase in the level of performance of the students over the six academic years and a preference for measures of central tendency in the data reduction process.

The greater number of students who have responded by describing measures of central tendency could perhaps be due to the heavy emphasis that many teachers place on mean, mode and median. This gives the students a restricted set of experiences on which to base their data reduction, and in fact may force students, who would naturally be inclined to follow Path B processing, to follow what to them may be a less natural reasoning pattern. The large bulge of responses at Level 6 is the point at which students are using one simple statistic to describe the data. This may be due to the fact that once students have been presented with simple statistical facts at school, there are limited opportunities for data exploration and so they do not have the chance to develop more advanced data reduction skills.

Despite the similarities, there are some differences between the sets of data for the two parts of the question. Two in particular show up in a comparison of the results presented in Tables 4.3.1 and 4.3.2. First, the overall number of responses at Level 7 or 8 is much larger for Part I, (33), than for Part II (19). Second, Levels 3 and 4 have less responses for Part II (1) than for Part I (11).

The small number of students in Levels 3 and 4 in Part II could be due to the fact that there were more responses in the lower group for that part. There were a number of students who misinterpreted the graph and this made it impossible for them to formulate a sensible answer for the question. These responses were graded in the first group and so would help to account for the larger number of students in the lower levels for Part II. This is a problem with understanding the graph rather than the reduction of data and appeared to be a particular problem with Year 10 students. The larger number of higher level responses for Part I suggests that students are better able to make detailed statistical descriptions of data when presented in raw form rather than as a graph.

The performance of students appears to be affected in a similar fashion by varying mathematical ability for both parts of the question. This is apparent after comparing Tables H.3.1 and H.3.5 and Tables H.3.2 and H.3.6. Higher ability students are better able to understand the process of data reduction but the strong preference for the use of measures of central tendency, rather than dispersion, appears to apply across all mathematical abilities. This is the case whether the data are presented in raw form or as a graph.

A comparison of Tables H.3.4 and H.3.8 shows that the path chosen is not influenced by the gender of the student in either part of the question. There is a preference, irrespective of gender, for a description based on measures of central tendency. However, gender influences the level of the response, for Part II of the question and not Part I, as can be seen by comparing Tables H.3.3 and H.3.7. When the data are presented as a graph, the females are able to give a data reduction which is statistically superior to that given by males. However, when presented with raw data, the performance of males and females is similar.

These results suggest that the form in which the data are presented influence a student's reduction of data.

Comparison of Student Performances

To determine whether there is consistency in the way any particular student responds when the data are presented in different forms, the performance of students over the two parts of the question is considered. Both levels and paths are analyzed.

The levels of each student's response to Part I and Part II are presented in Table H.3.11. The groupings, of these levels, over the two parts of the question are given in Table H.3.12. There is strong evidence ($\chi^2 = 54.38$, 4 d.f., $p < 0.001$) that the grouping of a student's response is associated with the part of the question being answered. Investigating the association further, of the 91 non-zero differences, 57 were positive. This produced strong evidence ($Z = 2.31$, $p < 0.02$) that there are not equal proportions of positive and negative differences. As there were more positives than negatives it can be assumed that the students tended to attain a higher level on Part I than on Part II. A student is thus likely to show more understanding of the data reduction process when the data are presented in raw form.

Students were then compared to determine whether there is a link between the Path, A or B, taken for the two parts of the question, with the same categories as defined for Question 2. The results are presented in Table H.3.13. This is strong evidence ($\chi^2 = 71.79$, 9 d.f., $p < 0.001$) of an association between the path chosen and the part of the question being answered. Table H.3.13 suggests that there is consistency in the processing path over the two parts of the question. In fact, of the 117 students who were in a response path for both parts of the question, 101 (86%) students followed the same path in both parts (that is, both Path A or both Path B) while only 16 (14%) students followed different paths.

Overall, these results show that data reduction is performed at a lower level when the data are presented as a graph rather than in raw form. The ability to read and interpret the graph is expected to have some bearing on this result. However, in attaining a particular level of response, the processing path taken by a student does not vary with the form of data presentation, for example those students who have a natural inclination to

consider measures of dispersion when reducing data are inclined to do so whether the data are presented in raw form or as a graph.

SOLO Taxonomy Framework

The levels described earlier are now used, along with the SOLO Taxonomy, to create a framework which could be used in future to code student responses. Response codes followed by an F can be found in Appendix F.

Responses in the ikonic mode suggest no link could be made between the required task and any sort of symbolic representation. Such responses were mainly from students in the junior years. Within this mode, there is a framework of growth with prestructural responses appearing as Level 0, when no attempt is made to provide an answer.

Unistructural and multistructural responses, both appear in Level 1, and as with earlier questions, separation was difficult. These responses range from those students (such as 7214) who answer the question by literally choosing "numbers" when asked to choose a number or numbers, to those who focus on irrelevant aspects of the question, for example, a student (12204) discussed rounding off the measurements or those who misinterpret the graph, for example, a student (8204,F) discussed the marks going up by 5s, to those who discuss foot size but do not choose relevant data, for example a student (7102) wanted to use the odd numbers to represent the left feet because most people are right handed.

The relational level responses, corresponding to Level 2, indicate that the various aspects of the question have been considered sufficiently to produce a reasonable answer. However, attempts to justify the answer were either unsuccessful or resulted in a personal comment. These range from a student (8207) who did not know why 29 cm would represent the foot size, to a student (9205) who chose 26 to 30 because these were numbers he felt like selecting, to another student (10203,F) who chose 9 to represent the mark because it was a good mark.

As before, responses in the other two groups have been able to link the concepts in the question to concrete experience. Reasons given link directly to the question or the data. These responses are in the concrete-symbolic mode and can be interpreted by two cycles, each of three levels.

The first cycle involves appreciating that it is possible to reduce data into a more usable form. The elements in this cycle are the actual pieces of data themselves (data items). A relational response in the first cycle is achieved when all data items have been considered and represented in a concise form. Again, the unistructural, multistructural and relational levels correspond to the Levels 3, 4 and 5.

U1 - These responses present a simple answer as a reduction of the data but focus remains on the answer with the only reason given relating back to the question rather than to the data. A typical response is that of a student (7213) who used 28 and 30 cm to represent the data but could only give as the reason "that would tell you how big the foot was".

M1 - These responses still indicate a simple answer, but attempts to focus away from this and back to the data fail when attempting to combine the information. This results in quoting that all the data items are needed to represent the data. Typical responses are those of a student (11201) who chose 26 to 33 to cover all the sizes of the left feet, and another (9103,F) who chose 1, 2, 4, 5, 7 and 3 as the marks because it represents each numeral.

R1 - These responses indicate that the answer and reason have combined the various pieces of data in the reduction process. This is the stage where there appears to be a split into two separate Paths, A and B, for the preferred styles of processing. A student who processes along Path A tends to focus in on the central tendency of the data during reduction, while a student who processes along Path B focuses on the overall spread of the data. Responses in Path A range from a student (8209) who selected 28, 29 & 30 as these are approximately the size of students feet, to another student (11215) who chose 29 as it is the average foot size. Responses in Path B range from a student (7203) who chose 26-30 for the foot size so as not to have to write too much, to another student (11103,F) who chose 1 & 9 for the marks to show the variation between the numbers.

Responses in the second cycle show an appreciation of the fact that the reduction of the data involves the use of one or more statistics to describe the features or behaviour of the data. The elements in the second cycle are the various features (or properties) of the data which statistical data reduction are trying to describe. A relational response in the second cycle is not achieved until the student is able to consider various data features, and the fact that these are related when it comes to considering the overall data description. As before, the U, M and R levels correspond to Levels 6, 7 and 8, respectively.

U2 - These responses use a simple statistic to reduce the data but the reason given is usually just restating the definition of the statistic rather than relating it to the data itself. Path A responses mention one measure of central tendency either by a description, for example a student (10205) chose 28 for the foot size as it happens the most, or by its correct statistical term, for example a student (10209) chose 29 because it is the median. Path B responses mention one measure of dispersion either by a description, for example a student (12206,F) chose 5-7 as the mark because the

majority of the class fell into that category, or by its correct statistical term, for example, a student (12201) chose 26-33 as the foot size because it is the range.

M2 - These responses consider more than one aspect of the data and give a number of statistics in an attempt to give a more accurate impression in the reduction process. Responses in Path A mention a number of measures of central tendency, for example, a student (7215,F) chose 7, 1 and 4 because these show the most, the least and the average. Path B responses mention a number of measures of dispersion or give a more detailed description of the dispersion, for example, a student (7204,F) gave 4-6 as a mark because there are an equal number of scores from 10 to 6 and from 1 to 4. Responses at the M2 level which incorporate elements from both paths are typified by one student (11108) who chose both range and average to describe the foot size, and another student (11113,F) who selected 5 or 6 to represent the marks because they are in the middle and have fairly even number of results on either side.

R2 - These responses reduce the data by presenting information concerning both central tendency and dispersion. Some attempt is also made to use one style of statistic to qualify the other. At this stage the processing integrates the features of both paths. A typical response is from a student (12209) who chose 29 to represent the foot size as it is the middle but also discussed how the other numbers fluctuate either side of the 29.

The above framework could be useful in determining the stage a student has reached in understanding data reduction. To distinguish a concrete-symbolic mode response from a ikonic mode response there must be a realization that it is possible to condense data into a more usable form. Ikonic mode responses grasp at some feature of the question or personal experience of the quantity being measured, rather than the data itself, to decide on a number which is representative of the data. Concrete-symbolic responses indicate attempts to use the data to decide on a representative value. First cycle responses, within this mode, grapple with the realization and acceptance of the fact that a number or numbers can be used to represent the data. Second cycle responses condense the data into more statistically sophisticated forms.

Interpretation and Inference

The fourth area of statistics investigated was Interpretation and Inference. This area involves the interpreting of results and making of inferences from data. There are four main aspects of concern, namely: the reading information from tables; detailed interpretation and estimation; designing experiments including randomness, errors, test statistics and

critical regions; and, inference including diagram and table inferences, risks, trends and projections.

The fourth question was designed to help clarify students' understanding of the interpretation of data and the use of data to make inferences. This is an important step in the analysis of data, especially when producing scientifically acceptable arguments. The question aimed to present students with some data and then have them consider features of that data and make predictions using the data. Although no reference was made to any specific features, two inferences were asked for so as to encourage students to make use of the data. This was an attempt to view what students perceived as the important features of data and to allow them to use the data as they thought best to make predictions. As with Questions 2 and 3, the data were presented in two different forms, raw data in Part I and as a graph in Part II. The analysis of the responses to Question 4, considered in this section, mirrors the detail provided for Questions 2 and 3.

Analysis of Responses to Part I

The first part of Question 4, given in Figure 4.7, presented data in a raw form. No reference was made specifically to any features of the data or to any possible trends or projections for the future which meant that students were free to suggest and use any patterns or trends that they felt appropriate. In the analysis below, statistical interpretation follows the coding of the responses.

QUESTION 4
PART I

A well known intersection in Armidale has had a number of serious accidents. The number of serious accidents was recorded for the last ten years.

1981	1982	1983	1984	1985	1986	1987	1988	1989	1990
2	4	5	9	3	6	7	10	4	6

(i) Describe any pattern that you can see in the data.
(ii) Approximately how many accidents would you expect in 1991? Why?
(iii) Suggest some years in the future (after 1990) when you think the number of accidents might exceed 8. Why did you select those years?

Figure 4.7 - Interpretation and Inference - Part I Question

Coded Responses

As with the other three questions, investigation of the students' responses showed that it was possible to divide them into a number of levels based on the statistical quality of the answer given. The responses were coded on the basis of what was included in all three parts, (i), (ii) and (iii), of the question. For example, some students said that they could see no pattern in (i) but then went on to describe features of a pattern when answering (ii) and (iii). The coding of the response tried to capture the essence of the answers given over all three parts of the question.

First Group

The first group of responses dealt with only the requirements of the question and within this group there were three Levels, 0, 1 and 2. These responses indicate a rationalization of the requirements of the question but no consideration for, or processing of, the data.

Level 0 These responses indicate that an answer could not be attempted. Either a decision could not be made as to what was needed or having decided on what was needed, the answer was not known. For example:

(8205)	(i)	*pass*
	(ii)	*?*
	(iii)	*?*

Level 1 These responses indicate that some aspects of the question have been considered and an answer attempted. The responses can include predictions not substantiated using the given data. In some cases, a comment was given as justification, but the only link with the prediction was of a personal nature. For example:

(7212)	(i)	*My minds a blank*
	(ii)	*2001 No reason*
	(iii)	
(8201)	(i)	*that people should be more careful*
	(ii)	*8 because people aren't careful drivers*
	(iii)	*1992 1994 because there will be more traffic in those years*
(9109)	(i)	*I cant see a pattern*
	(ii)	*I dont know*
	(iii)	

(9207)	(i)	*Don't know*
	(ii)	*7*
	(iii)	*1994*

Level 2 These responses indicate that all aspects of the question, but not the data, have been considered. Often pattern features were only mentioned when making predictions. Sometimes trends were found in a set of numbers, for example, the years, rather than in the data as a whole. At this stage, explanations for predictions still resort to personal experience, or knowledge, rather than using the data. Any explanation given was consistent with the prediction. For example:

(7101)	(i)	*The last two numbers of the date go in the write (sic) order like 1981, 1982, 1983.*
	(ii)	*8 because there is 6 in 1990 and so there could be some more between 1990 and 1991*
	(iii)	*1992 Because of travellers* *1993 Because of speeding driver* *1994 Because of animals jump out in front* *1995 There might be none.*
(7102)	(i)	*I can't see any pattern in the data*
	(ii)	*a fare few like any other year because of easter and Christmas holidays and things like that (About 153)*
	(iii)	*every year because of the same reason as above*
(9115)	(i)	*double it, add one, -6, sort of*
	(ii)	*Probably 7*
	(iii)	*1992, 1996 it sort of fits the pattern*
(10201)	(i)	*can't answer*
	(ii)	*8 accidents*
	(iii)	*1991 Because it hasn't said the numbers 8 and the year 1991*

Second Group

In the second group are responses which, having considered the question, attempt to understand the data. These responses attempt to find a trend, but when trying to use the data to make predictions they do so in a non-statistical fashion.

Level 3 These responses indicate that when describing the pattern, or making the predictions, the data have been used to explain the answer but the focus has been made on one data item only. For example:

(8105)	(i)	*I don't see any pattern*
	(ii)	*about 4. I randomly picked a number from my head.*
	(iii)	*1998 Because in 1987 7 people died so in 1998 8 people died.*
(9203)	(i)	*? cant find a pattern*
	(ii)	*10 it was a guess but it was close to the other numbers*
	(iii)	*?*
(10205)	(i)	*In 1981 they had less accidents than in 1990*
	(ii)	*10 Don't know why.*
	(iii)	

Level 4 These responses indicate that all the data were considered in attempting an answer, but the pattern is only described as an overall impression. Explanations for predictions may still rely on personal experiences or just state that the pattern has been followed without explaining how. For example:

(7204)	(i)	*The number of accidents since 1981 has increased in a big number*
	(ii)	*3, I think people are much more aware of the dangers of careless driving*
	(iii)	*2015, 2020, 2030 because I recon more young people will take driving as a game.*
(9207)	(i)	*There has been increases in accidents*
	(ii)	*There would be about 8 accidents*
	(iii)	*1992, 1996 because of what the statistics say*
(11201)	(i)	*on the information the number of accidents should go down*
	(ii)	*about 5 because everyone in our year is getting their P plates*
	(iii)	*about 92 because everyone in our year is getting their P plates.*
(12201)	(i)	*it gradually gets worse*
	(ii)	*I'll guess about 5*
	(iii)	*1992 it wouldn't be far away*

At this stage there is a divergence of the responses into two distinct paths which appear to develop at seemingly parallel rates. These are labelled:

Path A those who use specific features based on central tendency to describe the pattern, for example, peaks or troughs in the graph are commented on; and

Path B those who discuss the overall spread of the data, for example, using an analogy with waves

Level 5 These responses indicate that the data have been understood and a trend recognized. Attempts to describe the pattern usually result in describing one feature which incorporates all the data. There is a readiness to engage in descriptions of data patterns but not sufficient experience, or tools available, to produce a statistically sophisticated response. Data are not used to justify the predictions. Path A responses describe the pattern using one feature based on central tendency, such as, maximums, while Path B responses use one feature of dispersion, such as, fluctuations. Some responses could succinctly describe the pattern, for example identify a four yearly cycle, but were unable to use the observed pattern to justify their predictions. Others were unable to provide as sophisticated a description of the pattern but made comments that suggested the pattern needed to be used to justify the predictions. For example:

Path A

(8108)	(i)	*there are all accidents over 2 a year*
	(ii)	*about 5 because they would be starting to fix it up*
	(iii)	*1992, 1996 because there Olympic years and people will be travelling a lot maybe through Armidale if the location is in Australia*
(8113)	(i)	*I can not see any pattern in this data except that it builds up until after 1988 when a lot of accidents occurred and then it went down low again*
	(ii)	*I would expect about 4 or 5 accidents in 1991 because there were so many accidents in 1988 that people will be more careful*
	(iii)	*I think that there will be more accidents in 1993 because people will be careless again*
(9113)	(i)	*The number is at it's(sic) highest every 4 years*
	(ii)	*5*
	(iii)	*1992, 1996*
(10104)	(i)	*The only pattern that I can see is that its below 10*
	(ii)	*7. People aren't listening to the laws.*
	(iii)	*1992, 1995, 2000. They are putting the laws on harder and more people will disobey them*

Path B

(7202)	(i)	*low to high then to low*
	(ii)	*5*
	(iii)	*1999 because it's good*
(7215)	(i)	*it is going up and down*
	(ii)	*8, because it is going up*
	(iii)	*2000 - 10 more cars due to population*
		2200 - 0 cars will be flying
(8204)	(i)	*it went in a wavy line, up and down*
	(ii)	*8 - 10*
	(iii)	*1991 Christmas holidays there's a lot more traffic on the roads*
(10207)	(i)	*it rises and falls but the amount of accidents are basically becoming more frequent*
	(ii)	*about 9*
	(iii)	*1995*

Third Group

This final group of responses indicate a readiness to describe the pattern in the data in a more acceptable statistical form. Also predictions are related to the data.

Level 6 These responses indicate that the pattern in the data could be described and one feature of the pattern used to justify the prediction. For example:

Path A

(7112)	(i)	*the numbers go lower in the odd numbered years than the even numbered years*
	(ii)	*I would say 3 accidents because it is an odd numbered year*
	(iii)	*1992, 1994 because they are even numbered years*
(9102)	(i)	*In 1984 & 1988 there were more accidents than in any other years.*
	(ii)	*about 5. Because I think that in that year there might be that many.*
	(iii)	*1994, 1998. Because in every 4 years there is a fair few accidents.*
(12205)	(i)	*Every 4 years the number of accidents boom and then the next year drop dramatically*
	(ii)	*Higher than 6 because its only the third year.*
	(iii)	*2000, 1996, 1992 only then it will because it fits the pattern*

Path B

(9108)	(i)	*it increases then it falls over and over*
	(ii)	*7 or 8 judging on the pattern and if it continues*
	(iii)	*1992, 1996 approx every 4 years. I selected those years because of the pattern how it rises up over 8 every 4 years.*
(9111)	(i)	*the number start off low then gradually rise then suddenly drop*
	(ii)	*about 7*
	(iii)	*1992, 1997, 2002 because they are 5 years apart and it gives the numbers time to drop and rise again*
(11102)	(i)	*The data goes up and down in different years*
	(ii)	*about 10 why? because it is repeating and adding one*
	(iii)	*1991 because it's going up*

Level 7 These responses indicate either the need to present more information about the pattern or, in the justification of the prediction, discuss more than one feature of the data. Some responses discuss concepts which were related to both paths. For example, those discussing central tendency features of the pattern also mention some aspect of dispersion. This is the stage where the two paths are beginning to converge again. For example:

Path A

(9112)	(i)	*The number was highest every 4 years and its lowest every 4 years starting from 1*
	(ii)	*2 because its an average and its written there*
	(iii)	*1991 because every 4 years there is an increase of accidents*
(11104)	(i)	*the pattern is very irregular but there is always an accident at that intersection making it a trouble spot*
	(ii)	*In 1991 there would be about 7 accidents according to the table the accident rate is on an increase again*
	(iii)	*In 1992 the accident level would exceed 8 as that would be when the highest out of 4 years would be due. There is 3 years between each high rate.*
(12113)	(i)	*a pattern in the data is that every 4th year the number of accidents decreases remarkably*
	(ii)	*7 accidents because from the data the pattern seems to have in the 3rd year 1 more accident than in year 2.*

(iii) *They may exceed 8 accidents in 1992, 1993 because the accidents follow a pattern and usually every 3rd and 4th year they exceed 8 accidents, so I just followed the pattern.*

Path B

(9214) (i) *It spends 4 years going up and then it drops again & the 2 yrs after only 1 difference*

(ii) *7, well 1981, 1985 & 1989 have been the lowest years in the groups of 4 and the 2 years after there is always 1 different and 1990 & 1991 are 2 after a low.*

(iii) *1992, 1996, 2000 because every 4 years has a higher rate.*

(11107) (i) *In 1981 there were only 2 it increased until 84 dropped at 85 and increased to 88 dropped at 89 etc*

(ii) *Around 8 or 9 because that is how the pattern continues*

(iii) *1992, 1996, 2000 etc because the accidents usually occur the most every 4 years*

(12214) (i) *There is no real pattern, but it seems to be increasing in a 4 year spell each 4 years.*

(ii) *7, in the pattern after the second year it goes up by 2.*

(iii) *1992, 1996, 2000 etc. because at the end of 4 year pattern it seems to always exceed 8.*

Responses which are not in Path A or B, pattern description and prediction involves students using features of both central tendency and dispersion.

(8215) (i) *The number of accidents appears to be rising with fluctuations occurring at various places. After a fluctuation brings the number below 4 the number in the next year doubles and the year after it goes up by 1 e.g. 3=6=7, 2=4=5. The number peaks every 4th year in the 4 year cycle and troughs every 1st year the number rises on years 2, 3 and 4.*

(ii) *I would think about 7*

(iii) *1992, 1996, 2000, 2004 etc. This is because the pattern peaks every 4th year.*

(12104) (i) *Steady increase from '81 to '84. Drop in '85. Steady increase from '86 to '88. Drop in '89. A predicted increase from '90 to '92.*

(ii) *7 to 8 the pattern would suggest another steady rise in accident occurrence on the intersection.*

(iii) *1992 and 1996 and 2000 because once again we see the cycle taking four years from the trough to the peak.*

Level 8 These responses indicate the use of features of both central tendency and dispersion. Attempts are made to establish a link between them when describing the pattern and justifying predictions. For example:

(12111) (i) *It starts small and increases for 4 years till it reaches a peak of 9, when it goes down again and increases for 4 years until it reaches a peak of 10 and then it starts small again. The numbers overall are gradually increasing*

(ii) *8 or 9 because it goes up if it is to follow the pattern and it usually increases by more than 1.*

(iii) *1991, 1992, 1995, 1996 because if the pattern continues it should be at its largest in 1992, then at a low in 1993 (4 or 5 perhaps) increase in 1994 and be above 8 in 1995 and above 10 in 1996. And the pattern continues like that.*

(12209) (i) *They have 4 year cycles. They start low reach a peak, at the end of this drop off once again and begin to peak*

(ii) *between 7 to 9. This follows the pattern of the previous years.*

(iii) *1992. The accidents would possibly follow the previous years and peak at 4 years*

All responses given by students could be graded into one of these nine levels. The check on coder reliability showed a 97% agreement. The one non-agreeing coding, out of thirty, was in the same group and only one level from the researcher's coding.

Statistical Interpretation

Five interesting points arise from the results, arranged by academic year, in Table 4.4.1. First, there are only two students (3%) from Year 11 and 12 whose responses fall within the first group, whereas in Years 7 and 8 there are nineteen students (32%). Second, there is a larger number of senior, compared to junior, students whose responses are coded in the highest two levels. The only students whose responses were coded as Level 8 are three from Year 12, while at Level 7 there were only three (5%) juniors compared to twenty four (40%) seniors. Third, there is a marked difference in performance between Year 7 and Year 12, with no Year 7 response above Level 6 and no Year 12 response below Level 4. Fourth, there is a large bulge in all years at Levels 5 and 6 with over half of the responses coded into these two levels. Last, approximately twice as many student responses reflect

Path B than Path A in every Year, except 10, where there are equal numbers in each path. This suggests that there is a marked improvement in understanding from Year 7 to Year 12 and many more students prefer to make inferences using measures of dispersion.

Table 4.4.1 - Response Level and Path by Academic Year - Question 4 Part I

Level	Year												Total
	7		8		9		10		11		12		
0	1		1		0		0		0		0		2
1	6		5		7		2		2		0		22
2	4		2		2		3		0		0		11
3	0		1		1		1		0		0		3
4	3		1		1		0		2		3		10
	A	B	A	B	A	B	A	B	A	B	A	B	
5	3	8	4	5	3	5	5	2	1	1	1	3	41
6	2	3	2	6	2	6	5	7	3	9	3	8	56
7	0	0	2	0	1	0	1	2	2	9	1	5	
	0		1		2		2		1		3		32
8	0		0		0		0		0		3		3
Total	30		30		30		30		30		30		180

The results, arranged by mathematical ability, are presented in Tables H.4.1 and H.4.2. Strong evidence (χ^2 = 26.17, 6 d.f., $p < 0.001$) suggests that coding level depends on mathematical ability. In Levels 0 to 5 there are more low mathematical ability and less middle and high mathematical ability students than expected, while in the Levels 6 to 8 there are far more middle and high mathematical ability students. However, there is strong evidence (χ^2 = 0.48, 2 d.f., $p > 0.70$) to suggest that coding path is independent of mathematical ability. The results, arranged by gender, are presented in Tables H.4.3 and H.4.4. This significant result (χ^2 = 13.12, 5 d.f., $p < 0.05$) suggests that the level of the response depends on gender. Referring to Table H.4.3 suggests that there are more responses given by females than expected in the Levels 6, 7 and 8 but the coding path is independent (χ^2 = 1.15, 1 d.f., $p > 0.20$) of gender.

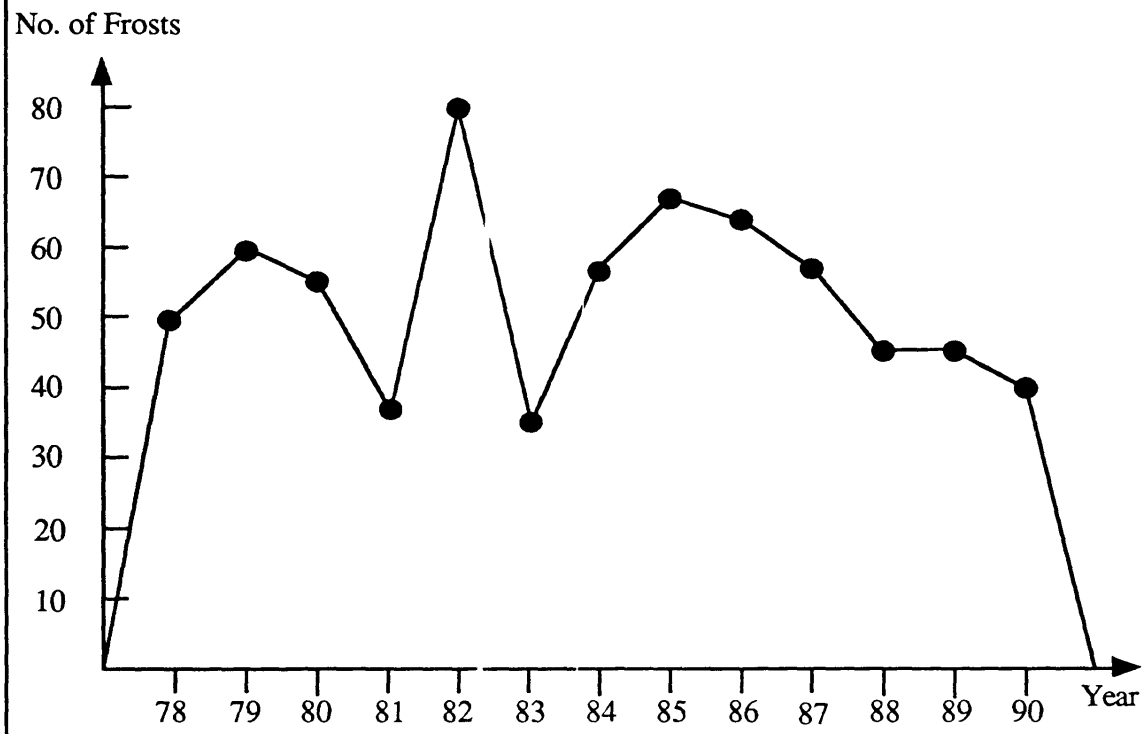
These results above suggest that for the process of describing data patterns and making predictions given raw data, the level of response improves progressively with

academic year and many more students use measures of dispersion, rather than central tendency, as the basis of the description. The gender and mathematical ability of a student do not appear to affect the path of a response given, but females and students with a higher mathematical ability are more likely to give a higher level response.

Analysis of Responses to Part II

QUESTION 4
PART II

Armidale is fairly well known throughout N.S.W. for its cold weather. The number of actual frosts per year have been recorded for the last 13 years in Armidale and the results are presented in the graph below.



(i) Describe any pattern that you can see in the data.
(ii) Approximately how many frosts would you expect in 1991? Why?
(iii) Suggest some years (after 1990) when you think the number of frosts might exceed 60. Why did you select these years?

Figure 4.8 - Interpretation and Inference - Part II Question

The data were presented in graphical form in the second part of Question 4, given in Figure 4.8 (on the previous page). As with Part I, no specific reference was made to any features of the data or to any possible trends or projections for the future. However, answering this question meant that students needed also to be able to understand and interpret the graph before they were able to describe the data or make predictions. An analysis of the response codings follows.

Coded Responses

When the responses to Part II were analyzed, similar levels to Part I were obtained, except that no Level 8 responses were identified. Some responses, which differ from those for Part I, are presented in Appendix G.

Statistical Interpretation

Four interesting points arise from the results, arranged by academic year, in Table 4.4.2. First, there are less students, nine (15%), from Years 11 and 12 whose responses fall within the first group, than from Years 7 and 8, where there are sixteen students (27%). Second, there are seventeen senior students (28%) coded into the highest level, 7, but only seven juniors (12%) Third, there is a bulge at Level 6 but in Year 7 the bulge is lower and in Year 12 it is higher. Last, approximately twice as many student responses reflect Path B than Path A, except in Year 10 where there are more Path A responses. There appears to be a reasonably clear increase in level with increasing academic year and in most years, students prefer to base predictions on overall spread.

The results, arranged by mathematical ability, are presented in Tables H.4.5 and H.4.6. There is strong evidence ($\chi^2 = 20.49$, 6 d.f., $p < 0.001$) to suggest that coding level depends on mathematical ability. In Levels 0 to 5 there are far more low mathematical ability than expected, while in the Levels 6 and 7 there are far more middle and high mathematical ability students than expected. As with the other questions, coding path is independent ($\chi^2 = 4.31$, 2 d.f., $p > 0.30$) of mathematical ability. The results, arranged by gender, are presented in Tables H.4.7 and H.4.8. Both coding level ($\chi^2 = 7.12$, 5 d.f., $p > 0.20$) and path ($\chi^2 = 3.44$, 1 d.f., $p > 0.05$) are independent of gender.

For the process of interpretation and inference, given data in a graphical form, the level of response improves noticeably with academic year, and many more students use the overall spread of the data rather than specific features to justify predictions. This preference, for the discussion of overall spread, does not appear to be influenced by either gender or mathematical ability. The level of the response, although not influenced by the gender of the student, is higher for students with higher mathematical ability.

Table 4.4.2 - Response Level and Path by Academic Year - Question 4 Part II

Level	Year 7		8		9		10		11		12		Total
0	1		0		0		1		0		3		5
1	6		5		5		3		1		2		22
2	2		2		1		1		3		0		9
3	2		0		1		1		0		0		4
4	3		0		1		1		2		1		8
	A	B	A	B	A	B	A	B	A	B	A	B	
5	0	6	3	4	1	4	5	0	0	2	0	2	27
6	4	3	4	11	5	9	7	8	5	12	4	6	78
7	1	0	0	0	2	1	1	1	2	2	2	6	
	2		1		0		1		1		4		27
Total	30		30		30		30		30		30		180

Comparison of Part I and Part II

Similar groupings and levels were found for both parts of the interpretation and inference question. The performance of the responses for the two parts of the question are now compared, using a similar structured analysis to that for Questions 2 and 3.

Overall Comparison

All student responses could be allocated to one of these levels suggesting that the framework may be adequate in explaining students' understanding, as far as the basic concepts of data interpretation and inference are concerned. As with the other questions, the upward shift in the trend of the responses over the academic years suggests the suitability of the levels. A large number of junior students found it difficult to formulate an interpretation of the data.

To better highlight the trends and differences between the two sets of data, the student numbers within each group and each path for the two parts of the question are now compared. The number of responses that were graded into each level for Parts I and II appear in Table H.4.9. These results ($\chi^2 = 7.94$, 5 d.f., $P > 0.10$) indicate that there is no association between the coding level and the part of the question being answered. Response

paths for Parts I and II appear in Table H.4.10. These results ($\chi^2 = 0.02$, 1 d.f, $p > 0.80$) indicate, also, that the path chosen is not associated with the part of the question being answered. Thus, knowledge of which part of the question was being answered could not be used to give any indication of the expected level or path of a response.

The analysis based on academic year shows, by comparing Tables 4.4.1 and 4.4.2, four noticeable trends. First, there are considerably more Year 7 and 8 students in the first group than there are Year 11 and 12 students. Second, there are more senior students than junior students in Levels 7 and 8 of the hierarchy. Third, there is a large bulge in the numbers at Level 6. Fourth, there are more responses in Path B than Path A in every year, except Year 10.

The unexpectedly large bulge of responses at Level 6 is the point at which students are using one simple feature to describe the pattern in the data. Once they have decided on this feature, they explore no further and are prepared to use this feature to help with any predictions to be made. Lack of exposure to interpretation and inference situations may mean that students are not aware of the possibility of allowing the pattern to have more than one feature or characteristic. Even if students are aware that this is possible, the lack of exposure may limit the students' confidence. The greater number of students who have responded by describing patterns using features based on dispersion or spread (Path B) could perhaps be due to the limited tools for description available to the students. Students who would normally prefer to process in an environment based on central tendency (Path A) may not feel that this is suitable in such a situation and so resort to a general description of repeated increases and decreases, which was classified as a description of spread.

Comparing the results presented in Tables 4.4.1 and 4.4.2 there are three noticeable differences. First, the number of Year 11 and 12 students on Level 0, 1 or 2 is much larger for Part II (9) than for Part I (2). Second, Part II has five Year 12s below Level 4 whereas Part I has none. Third, the overall number of responses at Level 7 or 8 is larger for Part I, (35), than for Part II (27).

It is difficult to explain why there were five Year 12s at such low levels in Part II, except to note that three of these did not answer the question at all and their answers to Part I showed reasonable understanding of the concepts. It may be that these students actually missed that part of the question. There were less students in the top two levels for Part II of the question but not really enough to suggest that the graph made it any harder to answer the question. In fact, the task of describing the pattern for Part II may not have been quite as simple as for Part I because the changes in the data were not so even and, maybe, not so easily described. This difficulty may have been offset by any help that may have been gained by observing the data as a graph rather than in a table.

Varying mathematical ability had a similar influence on results in both parts of the question. Tables H.4.1 and H.4.5 both show a significant influence of mathematical

ability on the level of the response, with higher ability students giving responses which demonstrate greater understanding. The path of the response is not influenced by varying mathematical ability for either part of the question. The results in Tables H.4.2 and H.4.6 show a definite preference for Path B with students using descriptions of the overall spread of the data to interpret and make inferences. Comparing Tables H.4.4 and H.4.8 shows that the gender of the student did not influence the path of the response in either part of the question. However, the analysis of the variation of levels with the gender of the student showed different results for the two parts of the question. Comparing Tables H.4.3 and H.4.7 shows that in Part I the females tended to give higher level responses than the males but for Part II there was no significant difference in performance. This suggests that when the data are presented in a raw form the females have more understanding, than males, of the interpretation and inference of data.

The above analysis suggests that the hierarchy of levels, with two parallel paths, used to categorize the responses is suitable. The form in which the data are presented does not appear to influence either the preference for using overall spread to describe data, or the tendency for higher mathematical ability students to respond at a higher level. However, females appear to perform better than males when the data are expressed in raw form rather than as a graph.

Comparison of Student Performances

To consider whether there is consistency in the way in which any particular student responds when the data are presented in different forms, the levels at which the students responded on each part of the question are compared.

The level of each student's response to Part I and Part II are presented in Table H.4.11, with groupings in Table H.4.12. There is strong evidence ($\chi^2 = 35.21$, 4 d.f., $p < 0.001$) that the grouping of a student's responses is associated with the part of the question being answered. The figures in Table H.4.12 suggest similar groupings on Parts I and II of the question. In fact, responses for 57% of students fall into the same groupings for Parts I and II of the question. Investigation of the nature of the relationship between the levels on the two parts of the question, shows that, of the 108 non-zero differences, there were 55 positives. There is not enough evidence ($Z = 0.29$, $p > 0.70$) to suspect that there are unequal proportions of positive and negative differences and so the student's levels can not be considered to have significantly increased or decreased from Part I to Part II.

Next, student results were compared to determine whether there is a link between the Path, A or B, taken for the two parts of the question. The results are presented in Table H.4.13. There is strong evidence ($\chi^2 = 34.69$, 9 d.f, $p < 0.001$) to support the existence of an association between the path chosen and the part of the question being

answered. The figures in Tables H.4.15 suggest that the same processing path is selected for the two parts of the question. In fact, of the 90 students who were in a response path for both parts of the question, 60 students (67%) followed the same path in both parts (that is, both A or both B) while only 30 students (33%) followed different paths.

These results show that the performance of a student in interpretation and inference is at a similar level whether the data are presented in raw form or as a graph. If a student attains a coding in a particular grouping on one part of the question, then they are more likely to be coded in the same grouping, than any other, for the second part of the question. Similarly, the processing path taken by students does not vary with the form of presentation, for example, students who process along Path B for Part I are likely to process along the same path for the second part of the question. It appears that presenting the data in graphical form has not affected the individual student's abilities to interpret the data and make inferences.

SOLO Taxonomy Framework

The SOLO Taxonomy, along with the levels described earlier, is now used to create a framework. Response codes followed by a G appear in Appendix G.

The ikonic mode responses suggest that the students were unable to link the required task with any sort of symbolic representation and were mainly from students in the junior years, as with the other questions. The prestructural responses, appearing as Level 0, are those where there is no attempt made to answer.

Level 1 responses are a mixture of unistructural and multistructural responses. It is possible that the unistructural responses give no reason, while the multistructural responses are those that give reasons which are not related to the predictions. These responses range from those of students (such as, 9109) who can not find a pattern and give no answers to (ii) and (iii), making it difficult to determine what aspects of the question have been considered, to those students who focus on personal observations triggered by aspects of the question, for example, a student (7204,G) discussed the greenhouse effect, to those students who attempt predictions which appear to have no relation to the data, for example, a student (7212) predicted the year 2001 but gave no reason why, and, another student (8105,G) predicted 60 because it is a random number.

Relational level responses, correspond to Level 2, and indicate that various aspects of the question have been considered when producing an answer. However, attempts to justify the predictions are either nonexistent or rely on personal comments rather than relating to the data. These range from a student (10201) who predicted 8 accidents for 1991 but gave no reason, to a student (7209,G) who predicted 70 frosts in 1991 because winter is going to be cold.

As with the framework for earlier questions, the responses in the second and third groups linked the concepts in the question to concrete experience. These responses fall into two U-M-R cycles within the concrete-symbolic mode. The first of these cycles involves describing the pattern, but usually in a form which makes prediction a difficult task. As with Questions 2 and 3, the elements for consideration are the actual pieces of data themselves (data items). Once all data items have been considered, and the pattern described and used for prediction, a relational response in the first cycle has been achieved.

U1 - These responses present either a simple pattern as an answer or infer a pattern from the explanation of the prediction. Focus remains on the answer given and does not relate back to the question or the data. This makes it very difficult for predictions to be made. A typical response is that of a student (9203) who could not find a pattern but predicted 10 frosts in 1991 because it was close to the other numbers.

M1 - These responses still indicate a simple pattern as an answer and, when focusing back to the data, attempts to combine more information into the description fail. This results in only an overall impression of the pattern being given. Predictions still have personal comments as justification, rather than using the data or pattern. A typical response is that of a student (1201) who said the number of accidents should go down, but "everyone in our year is getting P plates" was given as the reason for the prediction of 5 accidents for 1991. It is also possible that in attempting to describe the pattern all data are quoted as a student (10113,G) has done for the frost figures.

R1 - These responses present a simple description of the pattern with discernible features which encompass all of the data. The predictions show evidence of the pattern having been used to produce them, but an explanation of how the pattern was used could not be made. This is the stage where the coding splits into two separate Paths, A and B, for the preferred styles of processing. Processing along Path A tends to describe the pattern using a feature based on central tendency, while processing along Path B uses the overall spread of the data to describe features of the pattern. Responses in Path A are similar to those of a student (8201) who pinpointed a lot of frosts in 1982 and another student (10104) who stated that the number of accidents is below ten. Responses in Path B range from those of a student (7202) who described the accidents as going from "low to high then low", to those of a student (7208,G) who described the number of frosts as "going up by great numbers and going down moderately". Responses obviously transitional to the next mode were recognized. Some students (such as 9113) succinctly described the pattern as a four-year cycle and although it is obvious that the pattern was used as a basis for prediction, this could not be explained. Other students (such as 7215), although providing a less

sophisticated description of the pattern, were able to suggest in explanation that the pattern had been used as a basis for the prediction.

Responses from the second cycle involve appreciating that the pattern in the data can be used to justify the predictions made. The elements to be combined in the second cycle are the various features (or properties) of the data which constitute the pattern. If the various data features have been considered when describing the pattern, and the pattern found is used to justify predictions, then the response is considered to be relational.

U2 - These responses give a simple description of the pattern in the data and use this pattern to make predictions. At this level, difficulty is encountered in justifying the prediction using the pattern, and the explanation may just quote that the prediction follows the pattern. Path A responses mention a simple pattern based on central tendency, for example, a student (9102) noted there are more accidents in 1984 and 1988 and then predicts more than 8 accidents will occur again in 1994 and 1998 because "every few years there is a fair few accidents". Path B responses are based on dispersion, for example a student (10207,G) described the frost data as looking like waves with some years having higher frosts than others and predicted more than 60 frosts in 1993 and 1996 "because of the pattern on the graph".

M2 - These responses indicate that more than one feature of the data have been considered when describing the pattern, and attempts to justify the predictions use features of the pattern. Responses in Path A mention a number of features based on central tendency, for example a student (7104,G) pinpointed the highest and lowest number of frosts as well as the average and then predicted 40 to 50 frosts in 1991 because it is the average. Responses in Path B mention a number of features based on dispersion, for example, a student (12214) described the pattern as increasing in four year spells every four years but then predicted seven frosts in 1991 because "in the second year it goes up by 2", thus describing another feature of the pattern. Responses at the M2 level which incorporate elements from both paths are typified by a student (12104) who detailed the exact years when there are steady increases and decreases in the number of accidents and then when predicting that there would be more than 8 accidents in 1992, 1996 and 2000 mentioned that there is a four-year cycle from the trough to the peak.

R2 - These responses present features of the pattern concerning both central tendency and dispersion and accurately justify the predictions using the pattern described. It is at this stage that there is integration of the features of the data. A typical response is that of a student (12209) who described the pattern as having four-year cycles, as well as mentioning lows and peaks. He then justified a prediction of 1992 for the year when

accidents will exceed eight by stating that following on from previous years the number will peak at four years.

The levels and modes identified could be useful in determining the stage a student has reached in understanding the process of interpretation and inference. Once a response exhibits the realization that predictions need to be substantiated, it indicates that the transition to the concrete-symbolic mode has been made. Ikonic mode responses tend to rely on personal experience, rather than data, to make predictions. Inability to process the data usually results in no pattern being discussed. Concrete-symbolic mode responses indicate that the data have been taken into consideration in the pattern description or the prediction. First cycle responses, in the concrete-symbolic mode, give pattern descriptions, or predictions, based on the data but exp anations do not connect the pattern to the prediction. Second cycle responses, not only describe a pattern in the data and give reasonable predictions, but are able to explain how the pattern was used to make the prediction.

Conclusion

The analysis of student responses to the questions in the four major focus areas allowed the research questions Q 2.1 to Q 2.8, posed at the end of Chapter 2, to be addressed. The initial classification of responses showed that it was possible to arrange students' answers so that there was a hierarchy showing increasing statistical understanding. The number of main steps, or stages, identified ranged from 7 to 9 over the four questions. The lowest of these levels, coded as 0 was where no response was attempted. The top level, coded as 8, contained integrated responses which showed a good understanding of statistics. There were not many responses at this level and these were identified only in three of the focus areas; no responses to the data tabulation and representation question were considered to exhibit sufficient understanding to be coded at this level. This classification of responses demonstrates that it is possible to devise a hierarchy of growth of students' understanding within each of the four major focus areas.

As well as a hierarchy of levels a divergence into two distinct approaches was observed in the responses in all focus areas except data collection. Some students tended to focus in on particular aspects. These responses were classified as Path A and included those who represented data by using a simple numerical desciption or based reduction, interpretation and inference on measures of central tendancy. Other students tended to stand back from the data for a wider view and take a global approach. These responses were classified as Path B and included those who represented data by using a simple judgement or based reduction, interpretation and inference on measures of dispersion.

Not only was an hierarchy identifiable, but it was possible, as hypothesized in question Q 2.2, to use the SOLO Taxonomy to offer a framework, to explain this hierarchy, in each of the main focus areas. Responses were identified from both the ikonic and concrete-symbolic modes. The number of responses from the ikonic mode were not sufficient to clearly distinguish all levels within the mode. However, as a greater proportion of the responses were coded within the concrete-symbolic mode it was possible to clearly describe a number of levels. The description of the levels identified can be found in the appropriate section of this chapter for each of the focus areas. The levels found in the concrete-symbolic mode fell into two distinct cycles which is consistent with results obtained by other researchers, as discussed in Chapter 2. A number of responses in the concrete-symbolic mode exhibited the use of ikonic support as illustrated in path D in Figure 2.7. For example, a student (8204), when answering Question 4 Part I, brings in the personal experience of Christmas traffic to qualify a concrete-symbolic response.

Of major importance in secondary education is the difference between the ikonic and concrete-symbolic modes, as queried in question Q 2.5. Students who are still operating solely within the ikonic mode need to be encouraged to improve their understanding so as to be able to process within the concrete-symbolic mode, where most secondary learning is expected to take place. To assist students, it is necessary to know the major differences between responses in the two modes. The major difference appears to be the ability to identify that the numbers, or data, actually contain information which could prove to be useful. In the data collection aspect, the step from ikonic to concrete-symbolic mode is marked by the focusing on specific details of the data collection process. This shows an awareness of the data as information which needs to be actually collected rather than something which already exists. The other three major aspects of statistics all involve dealing with data. The major difference between the ikonic and concrete-symbolic modes in these areas is in the ability to be able to focus in on the data items in detail. In the ikonic mode the data are recognized but not processed in any way as individual items which carry information. Most processing which occurs is related to the question or personal experiences rather than information contained in the data. Responses in the concrete-symbolic mode attempt, with varying degrees of success, to use the data items and the information they contain, to answer the question. This is an essential aspect of understanding in the areas of tabulation, representation, analysis, interpretation, and, inference processes.

The final three research questions, Q 2.6, Q 2.7 and Q 2.8, addressed in Study 1, are related to the effect of the factors, academic year, mathematical ability and gender of the student, on the level of the response. Response paths were not mentioned in the research questions as their existence was only recognized during the analysis but the influence of the three factors on response paths is also mentioned here for completeness.

There appears to be a slight increase in the level of the response with increasing academic year. However, the bulk of responses were coded within Levels 4 to 6, with the location of the bulge being slightly lower for Question 1 than for the other three questions. As far as response paths are concerned, Path A was slightly more popular in Question 2 and much more in Question 3, with Path B being more popular in Question 4. This pattern did not vary greatly with changing academic year.

Mathematical ability had some influence, either in one or both parts of the question, on the level of the response in all major areas except data tabulation and representation. The students with a lower mathematical ability were more likely to give a low level response than those with higher mathematical ability. However, mathematical ability did not affect the path into which a response was categorized, except in the data tabulation and representation when more low mathematical ability students used Path B responses than was expected when the data were presented in graphical form.

The influence of the gender of the student on the level of the response was significant in three of the four focus areas. In data reduction, the females performed at a higher level than the males when the data were presented in graphical form, while in the interpretation and inference question the females performed better than expected when the data were in raw form. In the data collection question, the females exhibited better understanding than the males when outlining the details of the sampling. However, the gender of the student did not appear to influence at all the path, A or B, of the response.

Also of interest, are the effects of prompting and the form of the presentation of the data. In Question 1, Part II involved students being asked for more detailed information after having been already given some of the sampling details. A greater level of understanding was demonstrated in this part suggesting that, with prompting, students may be able to show more understanding. In the other three questions, Part II of the question presented the data as a graph rather than in raw form. The level of response for both Parts I and II was similar, except in Question 2 where a lower level of understanding was demonstrated when the data were presented in graphical form. The difference was more noticeable in the number of responses in the first cycle than the second of the concrete-symbolic mode. Thus some students find it more difficult to understand data presented as a graph but this does not appear to greatly affect their ability to reduce, interpret and make inferences. The path into which a response was coded does not appear to be influenced by whether the data are presented in raw form or as a graph.

The analysis so far has considered the results separately for the four focus areas and compared them. Combining these results allows a discussion of understanding about statistics in general. It also allows the difficulty of each of the questions to be compared. The results and conclusions of the combined analysis are presented in Chapter 5.