Chapter Six

The Second Exploratory Study: Building up a Bigger Picture on Some Characteristics of Statistical Thinking

The focus of introductory statistics courses is increasingly on teaching students to comprehend and evaluate statistical information. In this chapter we investigated, using the analysis of statistically based reports as illustrations, the understandings and knowledge of a small group of students. These students were found to be developing interpretive skills. It is hypothesised that context knowledge and subject knowledge operate interdependently in the statistical reasoning process. It is also hypothesised that specific instruction may be needed to fully develop interpretive skills and statistical thinking. Furthermore the interpretation of statistically based information may involve the development of students' cognition, disposition and metacognition in this area. The role of variation in statistical thinking and the implications of the findings for teaching are also discussed.

6.1 Introduction

This second exploratory study was designed in the same format as the first except the interviewed students would participate in a course for five weeks (15 hours) as part of a class (>60 students), rather than have a one day course, before being interviewed again. Initially the focus of interest was on changes in students' probabilistic thinking, particularly in terms of their understanding of variation. However, as the study evolved these changes became less important in the resultant analysis than the exploration of the data for a better understanding of some characteristics of statistical thinking such as variation (see Section 9.3). Therefore a design decision was made at the end of this investigation to refocus the study on identifying some characteristics of statistical thinking from a broad perspective. The reasons for this were fourfold: (1) a university environment is not conducive to an intervention study; (2) such a study could not develop students' statistical thinking if the characteristics of such thinking were not known; (3) the data and issues that were being raised suggested that this avenue should be explored; and (4) the scope of the study should be narrowed.

According to Gal et al. (1995) and Friel et al. (1997) statistics educators have not been clear about what constituted sound statistical reasoning for the interpretation of statistically based information. Therefore I had no structure on which to base the types of solution I should have expected from the students. Therefore the data gathered from the two interviews in the second exploratory study were used: to raise issues about teaching; to explore aspects of interpretation; and to reflect on the nature of statistical thinking. In Chapter 10 a structure has been proposed for judging students' interpretation skills. Therefore some normative solutions, that could have been expected from the students in the second exploratory study, have been presented in Chapter 10, together with a re-analysis of group and individual student data.

6.2 How do students interpret some aspects of statistical information?

In order to explore and understand the statistical thinking that interpretive skills required, six tertiary students were interviewed. These students volunteered to take part in the study. The main criteria for choosing these students were that they seemed to offer me an opportunity to learn, and showed promise of advancing my understanding of statistical thinking in relation to interpretation of statistical information, and to the possible implications for teaching. Four students were enrolled in a first year statistics course, one had completed a stage two statistics course and one had never done formal statistics, yet used statistics including experimental design, statistical summaries and significance tests in other university courses and in his previous employment. There were four males and two females with ages ranging from 18 to 50. Their school mathematics ranged from Form 5 to Form 7 (15 to 18 year age group). At the time of the interview five of the students had completed formal course work in 'What is Statistics?' (includes polls and surveys, experimentation, observational studies, random sampling and non-sampling errors), 'Tools for Exploring Data' (includes numerical and graphical summaries), 'Probability' (includes probability rules, conditional probability, statistical independence), 'Discrete Random Variables' (includes probability functions, Binomial and Poisson distributions, expected values), 'Continuous Random Variables' (includes Normal distribution).

The students were individually interviewed in depth. Broad questions (see Appendix Two) were asked in order to simulate the situation of reacting to a statistically based item. The focus of interest was on eliciting and evaluating students' opinions and judgements. They were expected to refer to, or infer, statistical relationships for data in graphical displays or in text, to "*consider other information about the problem context or consult world knowledge they may have, to help in ascribing meaning to the data*" (Gal, 1997, p.

50). They were told that I was interested in their thinking and reasoning rather than in their getting the ‘correct’ answer. The questions were presented orally and on paper. Unplanned probes were used in order to clarify the student's thinking for me. The interviews were audio-taped.

The students responses to statistically based information were analysed. In the analysis and discussion (Section 6.5 and 6.6), their responses to all the items in the interview were drawn upon. The items included a newspaper text article on fitness, a television item on gambling, an adapted television item involving a clustering of events, a sports statistics scenario, a graph on race and genetics and a table of social data which the students were asked to explore on the computer. The first exploratory study (Pfannkuch & Brown, 1996) which explored five students' responses to similar, and in some cases the same, statistically based items was also used as a basis for the discussion.

The focus of the research was on statistics. Therefore probability questions included were interpreted from a statistical stance. Below are the questions used in the interviews with observations and potential issues raised. A consensus was formed by me and my supervisor about the interpretation of the data. The research design involved an ongoing analysis and interpretation of the data and hence was linked to the literature in an effort to understand the nature of the thinking, with the proviso that these understandings would be revised as more evidence was gathered (see Chapters 8 and 10).

6.3 First Interview

The first interview included the following items.

<u>LOTTO Question</u>

On the "Money" programme on TV1 on Friday 28 April 1995 the advice for playing LOTTO was to spread the numbers that you chose. Comment on this advice.

Observations

Responses to this question were twofold. On thinking logically it did not matter whether the numbers were spread or not but on thinking intuitively they would spread the numbers rather than choose, for example, the numbers 1 through 6. One student had reconciled the two ways of thinking by saying that logically it did not matter whether the numbers were spread or not but in practice he would use 'personal numbers' such as family birthdays. Another student was adamant that it did not matter what combination was chosen, whereas the other four students had conflicts in their thinking. It would appear that three of them were interpreting the question from the perspective that non-consecutive sequences were more likely than consecutive sequences: “*they never come in a run of numbers*.” Therefore it was better to spread the numbers in a LOTTO game. The conflict in their thinking resulted in the fact that their subject knowledge suggested that the probability of all possible combinations was the same. Both of these ways of thinking about the question were correct.

The other student said he would not spread his numbers but would pick numbers randomly by closing his eyes and letting his pen fall on a number, or would choose the first number that came into his head. When asked if he would choose the numbers 1 through 6, he said that those numbers would not come randomly into his head but that he might get that if he used a random device such as a spinning wheel. The more he talked the more he realised that there were contradictions in his explanations. *"I just contradicted myself, didn't I? By saying now that - yes - not really relevant whether you spread the numbers or not. Each number has as much probability of being chosen as the last, so it doesn't really matter, no."*

Potential Issues

The student quoted above held beliefs about randomness and how to pick a random number which were possibly a product of childhood games. He had learned through instruction that a random sample meant all combinations were equally likely. He did not realise that his methods of randomly picking a number were biased until he was forced to think of an extreme example and hence he gained an insight into what randomness meant. His learned knowledge was divorced from his 'own' knowledge and the two had never been connected in the teaching/learning process.

This might demonstrate the importance of listening to student ideas and reconciling those ideas with the subject knowledge. Perhaps instruction had failed to address the intuitive models (Fischbein, 1987) held by these students and to logically resolve them. Too few links with their primary intuitions might have been established (Borovcnik, 1990) in the teaching process.

Map Question

Every year in New Zealand approximately seven children are born with a limb missing. Last year the children born with this abnormality were located in New Zealand as shown on the map (Fig. 6.1). What do you think?

(This item was adapted from a British television news item about the clustering of cases of a flesh-eating disease in several regions in England. The media and people were highly anxious and were searching for causes whereas a scientist was saying that there was nothing to worry about - it was variation.)

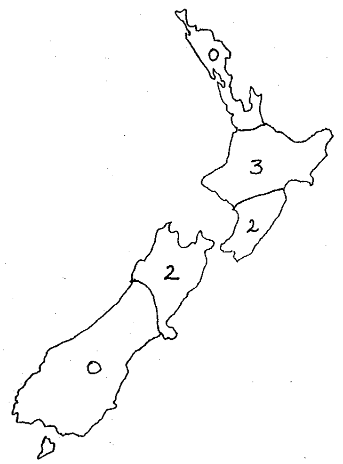

Figure 6.1 Map Question

Observations

The combination of text and diagram produced an explicit reading and understanding phase as all students gave a *description of the data* first. Three students were aware that the numbers of births should perhaps reflect population proportions and they mused that they would have expected one-third of the births to be in the Northland/Auckland region. The next phase of interpreting the data produced deterministic explanations such as pesticides, genetic mutations and the ozone layer. One student wanted to see the data over a period of time, perhaps a recognition of variation, and another student thought that the data were ‘fair’: “*it seems fair at first glance*” but wanted more information because the data were not what she expected. Not one student articulated the notion of small sample variation and expectations in the long run. These observations were similar to the first exploratory study.

Potential Issues

Context knowledge, such as population distribution in this problem, has not always been recognised in statistics teaching as an important facet of solving statistical problems yet this clustering of events required such knowledge to interpret what could be a typical media story. The conjecturing of possible causes has also been an important part in the search for patterns in the data and might be an essential component of statistical thinking. When dealing with human problems, probabilistic thinking might be deemed as unrealistic since humans would be adept at 'seeing' patterns in randomness. One way of overcoming that propensity would be to check for stability of patterns over time. Instruction has not seemed to have elucidated such issues, nor small sample versus large sample variation, for this particular type of problem.

Error Rate Question
In a firm in Wellington the management was concerned at the number of errors that office staff were making in transactions. The four office staff were audited every day over a month and the box-and-whisker plots shown were obtained. If you were the manager and had been presented with this graph (Fig. 6.2) what would you think?

This item was chosen to put the students in the role of a manager, or in an occupation where they did not collect the data but needed data to guide their actions. (It was based on a true story reported to me by a statistician (Thompson, personal communication, 1994), who was concerned about how managers were interpreting statistical information.)

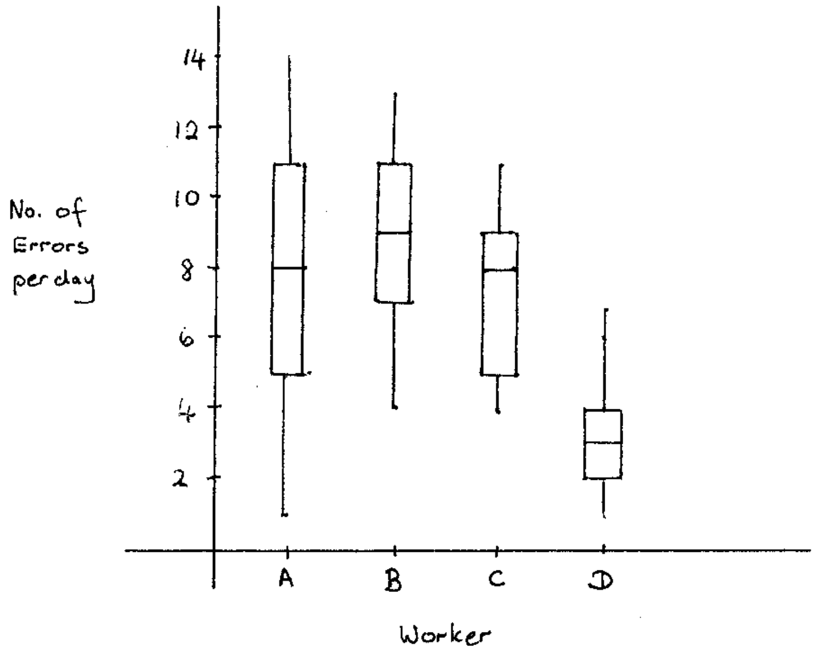

Figure 6.2 Error Rate Question

Observations

The first effect of the graphical information, on all the students, was to produce a detailed reading and description of the graphical features, such as the medians and the upper quartiles. The interpretation phase then followed with the graph being interpreted at face value. Four students suggested D was the best worker and as manager they would either sack the worst three workers or help them to improve. There was an assumption that the graph represented the truth. Superficial attention was paid to factors that could be causing the difference. One student wanted to check on the number of transactions and their complexity before interpreting the graph. The oldest student, whose background in a work environment would suggest context knowledge of such a situation, was interested in other factors that might explain the graph. The first exploratory study observations were closest to that observed for this oldest student. This might be because this question was given to the first study students after their intervention course, or because they too had been in work environments.

Potential Issues

The orientation of students towards reading exact values from graphs was also noted by Biehler (1996) who regarded this tendency as a problem area facing statistics students. The explanation of the variation in the graph raised some issues. In school mathematics, students have been taught that graphs tell stories. For example, they could be asked to describe the journey of a car using a distance-time graph. Perhaps this is why some students' *attention* was on giving a description of the statistical graph. A statistical graph might tell a different story from a mathematical graph. Have students not hypothesised why a graph could be demonstrating systematic variation and then followed that hypothesis through? Has the graph been the end point in instruction rather than the beginning of an investigation? Ben-Zvi and Friedlander (1996) also found that some topics effected higher modes of thought, whereas other topics invoked description only. They conjectured that preconceptions related to the context might lead students to ignore statistical ideas.

Or was the reason for the students' *attention* being on a description of the graph, a lack of context knowledge? In contexts such as medical matters or disasters it might be natural to probe for reasons while in other contexts it might not be, as ignorance of the situation might prevail. For example, when workers have produced errors, the workers have often been blamed. It has been well known in TQM (Total Quality Management) that the system was usually the problem rather than the workers. Hence cultural conditioning of expectations in such a context might have prevented the conjecturing of causes in working conditions. If the context knowledge was not present then interpretation might not take place.

Curcio (1987) suggested that graphicacy had three components involving reading the data, reading between the data and reading beyond the data. The third component was considered a higher level thinking skill as it involved extrapolation, elaboration of what was given and the making of inferences beyond what was explicitly presented (Curcio, 1987; Resnick, 1987). If this developing theory of graphicacy was applied to this question, then the higher order thinking skills, needed to interpret the information, seemed to be the conjecturing of some explanatory variables for the possible systematic variation in the graph. The students also had no systematic thinking tool, such as the fishbone diagram, on which to base their response. They could only respond at an intuitive level based on their personal experience. Resnick (1987, p. 48) stated that *"students must come to think of themselves as able and obligated to engage in critical analysis."* If such a learning culture could be developed then students would develop higher order cognitive abilities.

Die Question
A fair die was tossed with the resultant sequence of numbers: 3 4 4 3 5 3 5. What do you think of these results?

Observations

Two students interpreted the data literally and said they were not surprised. *"This dice doesn't really mean anything to me except for the fact that is what I've got."* Neither was able to articulate small sample versus large sample variation or talk about expectations in the long run. One student thought the die was fair but attributed the results to the way the die was rolled, after which she reasoned that more throws were needed to test for fairness. Three students initially thought the die was biased. However one of these students, through reasoning aloud, clarified her thinking by conjecturing that there was not enough data and that the die would have to be thrown many times to know whether it was biased, and therefore it was highly likely that it was not biased.

Another student was adamant the sequence was impossible. He was asked to write a sequence that would be more likely. He wrote: *"1,4,2,5,6,3,2*." He thought this was unlikely as all the numbers were obtained. He tried again: *"3,4,6,3,4,3,2*." He mused that he had got three three's and two four's. He then realised he was getting close to the example given. He began to see that each combination was just as likely as the other. He said: *"Maybe I am changing my mind"* and finally *"I've changed my mind, that's fair*." The third student simply declared the die was unfair. His attempt at a sequence was: *"2,3,6,5,1,5,4*" followed by the statement that a fair die would probably show one repetition. His idea of randomness was that it was uniform. The idea of small sample variation did not seem to be in these two students' repertoire or experience.

These observations differed in two respects from the first exploratory study. First, two students reasoned that there would need to be many throws and second, another two students declared the die was biased with one subsequently changing his mind.

Potential Issues

Were these differences in observations a result of school instruction? The two students who reasoned about long run relative frequency were the most recent school leavers of the interviewees, while the one student who declared the die was biased was the oldest interviewee. The interviewee who subsequently changed his mind would not have experienced die tossing in his schooling. Whatever the reasons it should be noted that causal reasoning was not predominant in the first or second exploratory studies of this research.

Prison Newspaper Question

(This item was chosen to explore how the students reacted to and interpreted a newspaper article.)

Private prisons will help stop suicide.
New Zealand Herald, Monday April 24, 1995

A perennial problem facing prison administrators is inmate suicide. Levels fluctuate greatly without apparent cause but a hopeful tendency over the past few years has been that rates have generally fallen.

After rocketing in the early 1980s to an all-time high of eight (288 per 100,000) in 1985, prison suicides showed a downward trend.

Between 1985 and 1989 there was an average of 5.8 suicides a year; between 1990 and 1993 it was only 4.25. In 1993 there was just one.

These reductions occurred at a time when the prison population grew by 68 per cent, from about 2800 in 1985 to 4700 in 1993.

But suddenly and perplexingly, prison suicides seem to have taken off again. There were 10 suicides in 1994 and, with another two so far this year; the trend threatens to continue.

Justice officials cannot say why.

Figure 6.3 Prison Newspaper Question

Prison Newspaper Question 1:
Read this article (Fig. 6.3) and tell me in your own words what you have read and how you react to the information.

Observations

One student focussed on the detail of the data, wondering how each figure was calculated and critiquing how the data were presented: *"the time frames that are used aren't consistent . . . if you average 93 and 94 you get an average of 5.5 per year instead of 1 one year and 10 another."* She also wanted to know about the author, possibly, to check out biases and hidden agendas behind the article. Another student suggested the data presentation was not appropriate: *"averaging this out like that for periods of four years* *isn't necessarily a very good way of doing it. It would be better to show year by year the amount of suicides not the variation."* Another queried whether the suicide rate had gone up, as the prison population had increased: *"because the population increased, I mean, just at a glance, without looking, I would be dubious as to whether they* [number of suicides] *actually have or not."* Three students did not critique the data presentation only described it: *"in the 1980s they* [number of suicides] *went quite high and then they dropped off considerably but in 1994 . . .."* Such numerical information was not easily connected to subject knowledge by the students.

Potential Issues

The 'worry' questions instigated by this article concerned reliability of the data, whether different ways of presenting the data would lead to different conclusions, whether the given data supported the central tenet of the argument, and whether using a rate measure would lead to a different conclusion from a raw data measure. All these issues, raised by three of the students, suggested that they had adopted a critical mode of thinking whereas the other students could be described as having adopted an uncritical thinking mode.

Further questions from their responses were raised. Have measurement issues and text data been features of statistics courses? Should the comparison have been based on raw data or been expressed as a rate? When human life was involved perhaps data were looked at differently. (e.g. At the current rate of aircraft accidents the number of fatalities has been acceptable. In ten years time, with the increase in aircraft traffic, a jumbo will go down every week using the same rate. Will this be acceptable?).

Prison Newspaper Question 2:
Later prompt: the justice official could not explain why, can you?

Observations

To explain the increase in suicides, all students thought of a number of causes, an expectation for statistical thinking, and two talked about independent and dependent events.

". . . see and look at problems like overcrowding, whether or not people are psychologically disturbed."

". . . you don't know that they're independent either do you? It could have been a suicide pact that was suddenly made."

It would appear that disaster coupled with a familiar context promoted causal thinking. Current concerns in the community were drawn on for possible causes.

Two students gave explanations that involved the role of random variation. One of these students was able to articulate small versus large sample variation ideas and had an appreciation of the sampling behaviour of averages but said that he would also investigate causes because it was a human problem.

"They're getting such a wide variation because you are dealing with a small number . . . the smaller your sample . . . the bigger the variations you can expect . . . so the larger the population you have, the more there's sort of potential, things seem to even out. In the same way that if you do a survey over a week you get a lot of ups and downs - if you did it over 20 years you would find it pretty constant."

The other student suggested one cause could be statistics: *"it's got to have some influence."* He was able to articulate probabilistic ideas even though he seemed to have a deterministic notion about random variation or statistics. The four other students did not proffer the idea of random variation.

Potential Issues

All students went through the phases of reading and understanding the text, describing the data and interpreting the data deterministically. Thus context knowledge was a source for statistical reasoning. Has this context knowledge been recognised as important in the teaching process?

Two students went the extra step of interpreting the data probabilistically. One of these students had concluded that the die of the die question was unfair and for the map question had produced deterministic reasons. Why had this question triggered those ideas? Had the context of the question or the extra information on the suicide rate over a number of years produced those ideas, or was there an interview effect?

The subject knowledge that could have been triggered by this article was that, at first glance, ten suicides in one year was an unusual result. The students could have then *recognised* that the data could be roughly modelled by a Poisson distribution. Some quick mental calculations could then have been *produced* to find out whether the result could be explained by variation. Even though none of the students recognised this as a Poisson process they should have still considered variation as a possibility and hence been prepared to interpret this data from a probabilistic as well as a deterministic perspective.

Students should also be aware of the reality of such a situation in that, firstly, because it was a process that would unfold over time, there should be a watching brief on future data. Secondly, that because it was a human problem, a cause would often be 'found', even if it was not the actual one, in an effort to appease critics. In fact, a year later, an article appeared in the newspaper reporting that a group looked into the high number of suicides in prison and had found at least ten risk factors for, or causes of, suicide. On the review group's recommendation 'a successful method of reducing suicides in prisons had been developed'.

Prison Newspaper Question 3:
Further into the article the author Dr Greg Newbold mused that "it could be that the recent surge in suicides is a random blip." What did he mean by this statement?

Observations

The four students who had not suggested this as one way to perceive the data readily agreed with the statement. Random blip to one student meant *"couldn't think of a reason"* or later on *"an accumulation of circumstances that cannot really be explained."* Another student had a visual picture of a blip on a computer screen that happened at random and that random meant 'not able to predict'.

"It wasn't uniform, couldn't predict where it was going to be. . . . What I am saying is that I don't think you could predict it [recent surge in suicides], *I don't think you could say based on the last 10 years worth of observation, on how many suicides, that this year is going to be that amount. No, I don't think you could do that, I think that there is always going to be that margin of error - there's just too many physical influences outside - outside of the statistical evidence."*

He compartmentalised his reasons into the two categories of statistical evidence and outside physical influences. He did not appear to understand small and large sample variation but did have a good idea of random processes. The other students said the blip was the result of a cause, such as a 'heatwave', and with further prompting produced the idea of a trend with an occasional anomaly.

Potential Issues

Classroom instruction typically has used random devices such as coin tossing in which there would be an underlying belief of inherent constancy. However for social phenomena there would be an underlying belief that any perceived constancy would change over time and space. Therefore the connection between ideas of randomness for coin tossing and ideas of randomness for social data have been subtle. There has appeared to be no link made between how people think and the teaching process. Probabilistic thinking would seem to require a global overview and an ability to stop thinking about the detail of the numbers and causes. It would seem that this is a new way of thinking for students when confronted with social data. This finding was supported by the recent work of Konold et al. (1996) and Hancock et al. (1992) who found that students had difficulty in adopting an aggregate-based perspective, preferring to reason instead on an individual basis. Another facet for thinking about group propensities would be the awareness of random variation. Randomness has an abstract concept. Successful practical applications have required a deep understanding of random behaviour and an appreciation of how randomness could be used to model variation. Random variation has been conceptually imposed on real processes at levels which were not yielding to deterministic thinking. Even statisticians would find it difficult to define random variation (Wild, personal communication, 1996).

Prison Newspaper Question 4:
The author also stated that "Maoris were 1.5 times more likely to commit suicide than non-Maoris." What comments would you have after reading that statement? What additional information would you like to see in evaluating this statement and why?

Observations

Three main themes could be drawn from the students' interpretation of this statement. The first theme was the judgement of the number in context. Three of the students queried the calculation of 1.5. Statements were made such as: *"would need to know the proportion of Maoris in prison before agreeing."* They wanted to know whether the number was based on the raw data or on a population proportion yet when probed further they were unable to articulate how the 1.5 ratio could be worked out.

The second theme was that contextual knowledge was a key component of processing statistical information. The students proffered causes of why Maori were 1.5 times more likely to commit suicide and thought of factors they would look at for an explanation. These causes were drawn from their own background knowledge and reflected current community and media information.

". . . because the Maoris in this country suffer from a cultural loss which is so deep rooted within them . . . I could quite believe this statistic."

". . . doesn't surprise me . . . economic circumstances and all those sorts of reasons."

For some of these students the third theme indicated that their experience, their context knowledge, prevented them from questioning or looking at the data from another angle. Two students appeared to be in a transition phase between accepting and questioning the statement, though this could be attributed to an interviewer effect. They tentatively raised such concerns as:

" . . . would like a breakdown of crimes committed";

". . . maybe suicides in general have gone up. I mean suicides in the population in general have gone up, then it might be something to do with society. "

The adoption of a critical stance and the raising of concerns about statements in reports was tenuous. Five students did not think of stratifying on variables other than Maori before agreeing with this statement.

One student after several prompts suggested stratifying on other variables.

"Splitting into Maori and non-Maori is - it's only one way of categorising people. You could split it into gang members and non-gang members, violent offenders and non-violent offenders . . . I'm not sure [about] *splitting everybody by race. There's too many other factors involved why people commit suicide other than whether they are Maori or non-Maori."*

He was aware of the multiplicity of factors that could result in people committing suicide and that there were many variables that could be proposed as an explanation of the variation. Such a critical attitude should be promoted in statistics.

Potential Issues

The comparison base for numbers was not well understood. Ratios would be a common way of expressing statistical information, yet have they been addressed in typical statistics courses? Have students been given raw data such as the number of suicides in each category and the proportion of the population in each category, and then asked to work out the ratio? Have they been asked to work out rates for raw data such as the number of deaths on the road per 100,000 people, or per 100,000 cars, or per 100,000 Km travelled, as a method of comparison with other countries? And then asked to interpret, from these several perspectives, whether New Zealand has a poor road fatality record compared to other countries? Have they discussed the fact that raw data might be appropriate when comparing yearly fatalities in New Zealand only? Students ought to have a good grasp of comparison bases for figures and a sense of how to compare data numerically.

Students also might need to be aware that their context knowledge might limit their search for alternative ways of looking at the data. That is, their own beliefs and experiences could affect the way they perceived and received data and information. In instruction, such context, as in this newspaper article, could prove to be a rich source for prompting hidden beliefs and opinions and promoting the development of a critical attitude.

An awareness that other explanatory variables might be better predictors might require an ability to look globally at the data, to re-aggregate and then to re-stratify. Perhaps students lacked the strategies and thinking tools to propose alternative explanatory variables or perhaps they were not in the habit of critiquing the classification of data. Hancock et al. (1992) stated that a consideration of what might have been neglected in the data collection process, and the realisation that data were a model that only partially represented the real-world situation, were key ideas in critically thinking about data-based arguments.

Basketball Question

Suppose that a basketball player over a long season has scored 70% of her free throws. Three-quarters of the way through a tournament game she attempts five free throws and scores on only two. The team manager attributed her performance to normal variation, that she scored 70% in the long run and that 70% was only an average so that you had to expect some low scores now and again. Critique the manager's comments.

Observations

All students thought the manager's explanation sounded reasonable and added that the odd blip was due to a cause or that they could see why other people might attribute it to something else. One student articulated his opinion as *"the manager's comments are OK if that is the way he wants to look at the score and not on 'we want to win'"* and then he gave possible causes.

Potential Issues

Describing such data as random variation has not solved practical problems on the day, but it has given data on consistent performers. In reality a low score would precipitate an analysis of why it had occurred so that changes could be made. Furthermore the data might be part of an unstable process.

Has statistics teaching viewed such data from two perspectives? School instruction has seemed to have concentrated on probabilistic thinking from a random device perspective (Biehler, 1989), which has not been a realistic portrayal of randomness in statistics. Yet in understanding randomness, simple models have been needed on which to build. The problem has been that teaching has not continued the process to the more complex models involving measurement data. Has teaching focussed on this probabilistic perspective so that we have ended up with 'school statistics' and 'real statistics'? Has the probabilistic perspective been the sole base for teaching or has it just been another way of looking at the data? Such multiplicity of perspective and interpretation ought to be explicitly, drawn attention to, by the teacher (Mason, 1989). The causal aspect would be an important part of thinking statistically. The dual way of thinking would be 'statistical thinking'.

Landwehr et al. (1995) stated that students should be made aware of their misconceptions. The judgement of a misconception assumed that there was only one solution for a situation although statistical thinking has required dual modes of thinking which were partly incompatible (Biehler, 1994b). A teacher may have required a probabilistic perspective but the student may have given a deterministic perspective. Should dual perspectives be required in the instruction process? Later on, with experience, students might become experts and know when to listen to their intuitions (Fischbein, 1987).

> "[Experts] *know the different types of thinking and the interface between the two, and in what circumstances to use the one or the other, and they use metaphors and experience from one domain for the benefit of the other"* (Biehler, 1994b, p. 8).

Fitness Newspaper Question (New Zealand Herald, May 1995)
Tell me in your own words what the article is about (Fig. 6.4). When you read data like this what questions are running through your mind about the study?

No limit on fitness

AP Washington

It is never too late to get fit, according to a study comparing the survival of men who started and maintained exercise programmes with those who never exercised.

Mr Steven Blair, head of the Cooper Institute for Aerobics Research in Washington, says a 10-year study showed that men who were unfit but who exercised their way to good physical shape had about half the number of deaths of those who stayed unfit.

When only heart attack deaths were considered, the exercise benefit was even greater, the research found.

The research, published this week in the *Journal of the American Medical Association*, studied 9777 men who each took two measured treadmill tests five years apart between 1970 and 1989.

For those judged unfit in both, the death rate in the following five years was 122 per 10,000. For those judged fit at both tests, it was 40 per 10,000. And for those judged unfit at the first test, but fit in the second, the death rate was 68 per 10,000.

Men in the test ranged in age from 20 to 82 years. There were benefits from exercise, however, no matter what the age of the men.

Fit men aged 60, for instance, had a 50 per cent lower death rate than persistently unfit men of the same age.

"It is never too late to take up exercise to improve your risk of preventing heart attack," Mr Blair said. "A sedentary lifestyle is a very important health risk."

He said that 20 to 25 per cent of all Americans led totally sedentary lifestyles.

Figure 6.4 Fitness Newspaper Question

Observations

All students went through a prolonged reading and understanding phase. The main message was not easily comprehended. Words were read and reread, defined and clarified. With prompts, all students knew other variables that should be taken into account such as medical history, diet, weight, stress, smoking. All of the factors were current concerns in the community and in the media.

Another student, who gave no indication on the other questions that he knew about random variation, now gave a very lucid account of it. Ideas of 'in the long run', of predictable within a range, of things that 'iron out', and of a trend, were articulated.

Potential Issues

Why had this article triggered ideas about variation for a student and not any of the other previous questions? Had the basketball question suggested this idea? There seemed to be a lack of interconnection of ideas for processing information. How could instruction involve students seeing the underlying commonalities, seeing and experiencing the underlying thinking? Has teaching needed to make this thinking explicit? When context was used, have widely publicised concerns affected the way students think?

6.4 Second Interview

The second interview (see Appendix Two) included the following items. Only five students were available for the second interview. At the time of this interview four of the students had completed formal course work on significance testing and LSD (least significant difference) plots.

<u>Genetic Damage Question</u> (adapted from Margolin, 1988)
Human monitoring for exposure to and damage from genotoxic agents is commanding increasing attention. The interplay of environment and genetics on human health means that scientists need to develop methods of measuring damage to human genetic structures. A method that has been suggested is the counting of the number of SCEs observed per cell. An SCE results from a reciprocal exchange of DNA between two sister chromatids (the two spiral filaments that constitute a chromosome). Some scientists believe that an SCE is a possible measure of genetic damage in an individual's DNA. For a particular individual the average of the SCE measurements is recorded.

The graph below (Fig. 6.5) is a plot of genetic damage. **Initially** the scientist divided the data up into these racial groups. If you were the scientist what would you think and what would you do next?

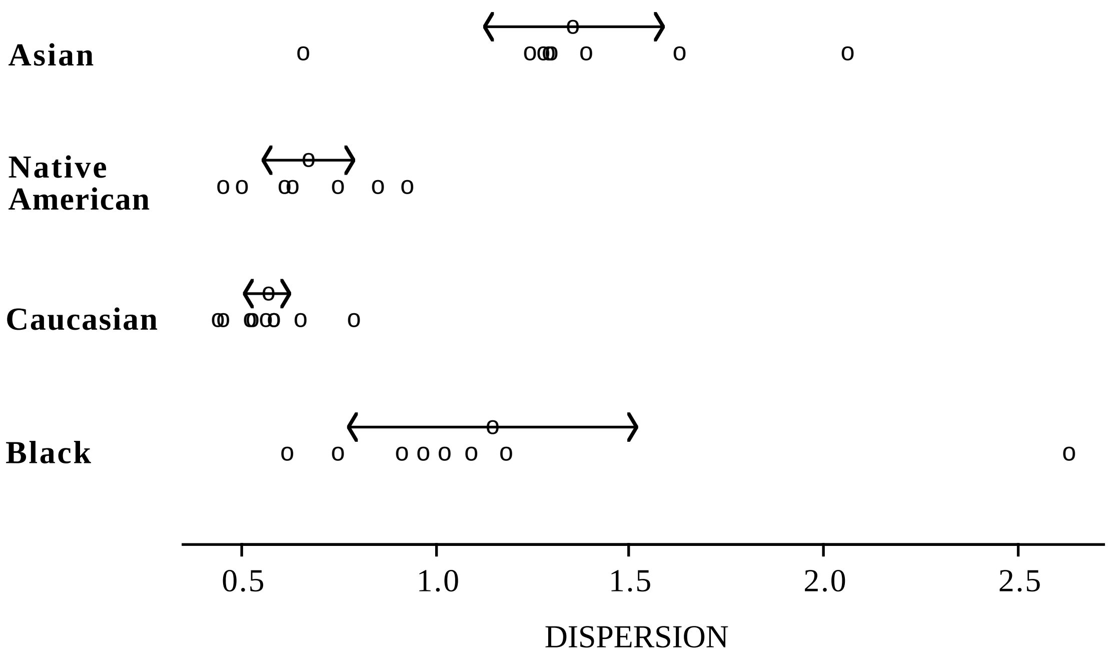

Figure 6.5 Genetic Damage Question

Observations

The first phase was the reading and interpretation of the graphical features. Data were looked at in detail: scales, means, spread, sample size, outliers, and finally the recall of how to interpret the LSD plots. They did not think of visually removing the outlier for the Black group in order to get another comparison of LSD plots. Some mentioned they would like to see larger samples, so have they misunderstood the role of the LSD plots? Many causes for the difference between the Asians and other groups were mentioned such as heredity problems, nutrition, toxins in the environment. All ideas were focussed on investigating possible causes for Asians having a higher genetic damage rate. After several prompts some suggested splitting the racial groups into male and female.

Potential Issues

Even though emphasis was put on the word 'initial' students did not think of stratifying the data on other variables (e.g. gender, occupation). When challenged about this the students said they were reading the data that were presented. They assumed that the scientist believed there was a difference between racial groups even though this was not stated. How have students been encouraged to look at the data from a completely fresh perspective? The students in class had investigated possible relationships in data tables using a computer but somehow this knowledge of stratifying on other variables was not transferred to the situation where a graph was presented. A possibility could be that the graph contained a lot of detailed information which could contribute to obscuring the bigger picture in the analysis. However it appeared that when there was systematic variation students did not seem to question the underlying assumptions. Was this because students have not done long extended tasks, not carried out a statistical investigation that went through several cycles of the enquiry process, not been challenged to explore every possibility before reaching a tentative conclusion? Have students experienced classifying and reclassifying data into different groups and noted the differences in possible conclusions that could be drawn?

Investigating Relationships Question

Students were given birth data information, a small part of which is shown below (Table 6.1). They were familiar with the computer software *DataScope* (Konold & Miller, 1995). They were first given the meaning of each variable and the reason why the data were collected. They were then asked to investigate any relationships in the data set. (Note: Since time was running out in the interview it was suggested they investigate whether there was a link between smoking (SMOKE) and birth weight (BWGT) first.)

Table 6.1 Investigating Relationships Question

ID	LOWGT	AGE	MOMWGT	RACE	SMOKE	PREM	HYPERT	UTIRR	FTVSTS	BWGT
1	0	20	46.7	white	yes	0	no	no	1	2557
2	0	23	57.8	black	no	0	no	no	1	3062
3	1	34	46.7	white	yes	0	no	no	0	1818
4	1	25	46.7	other	no	1	yes	no	0	1330
5	0	27	57.8	other	no	0	no	no	0	3969
185	1	20	66.7	white	yes	0	no	no	2	1928
186	0	22	42.2	other	no	0	yes	no	0	2750
187	1	29	53.3	other	yes	1	no	yes	0	709
188	0	26	74.7	black	yes	0	no	no	0	2920
189	0	15	60	white	no	0	no	no	0	3941

Observations

For the relationship between mother smoking and birth weight of the baby they endeavoured to plot a graph. With one exception, the bar graph was the first choice. On seeing that it was inappropriate they then chose a boxplot or scatter plot. Once the appropriate display was chosen they were asked what they would do next. One concluded that smoking was a factor in low birth weights as it was obvious from the graph. Three said they were not able to reach a conclusion as the samples were too small. What they would do next would be to take a larger sample. Even though four of the five students had covered significance testing in a formal course only one student suggested a test which she successfully carried out in Datascope. On further questioning about the ideas behind significance testing not one student was able to fully articulate the idea that: ‘Although I know that my particular samples for smokers and non-smokers have these particular medians and spreads I know that if I repeat this study with the same sample size I will get different values. So, is this difference in medians I see between the two groups due to random variation or is there a real difference? Suppose what I see is due to random variation how often will I see this difference or larger? This difference or larger occurs 4 times in a 100 through random variation only. Therefore . . .’

Potential issues

The lack of awareness by the students that boxplots allowed comparison of data from different groups raised the issue of teaching method. Was this lack of awareness due to instruction being based on learning to do each display separately, with little opportunity being given to students to choose the appropriate display?

Four students could procedurally carry out a significance test from a standard coursework problem. However when confronted with the raw data and then the graph three students did not suggest using such a confirmatory method. The ideas behind significance testing were not well understood. Were these ideas not well understood because they had not experienced or conceptualised random variation in a host of contexts and situations? Or had instruction revealed only the procedure not the method of thinking? How could the instruction process encourage students to adopt a critical attitude? How should teaching encourage such thinking as: ‘None of the ways I've looked at the data reveal any structure. Perhaps I have not asked the right question and if I did it might reveal structure so in the meantime I will model it as random’. To understand significance testing the student needed an understanding of random and systematic variation, and an awareness of the dual cultures of thinking probabilistically and deterministically.

6.5 Some Conjectures

Conjectures about Developing Students’ Interpretation Skills

Gal et al.'s (1995) cognitive and dispositional components for interpretation were the ability: to comprehend; to invoke 'worry' questions about; to evaluate; and to challenge media reports. Using this framework for the interview items, it appeared that the students could comprehend statistical graphs and had partial comprehension of numerical information in texts. The students had a partial list of 'worry' questions suggesting an unstructured approach towards evaluating items. They were able to express opinions on the items but were not good at raising concerns. A critical attitude was partially developed.

Therefore there should be a consideration that students would not be able to use these interpretive skills unless they: had first and foremost *context knowledge of the situation*; were able to *judge numbers in context*; were able to *understand variation*; were able to *reason within uncertainty*; and were aware that data could be interpreted differently dependent upon the *personal viewing lens* of the author. Reasoning within uncertainty from a context perspective would mean reasoning with insufficient information, and from a subject perspective would mean reasoning with margins of error. Another component, a ***metacognitive*** component, for interpretation should be considered. This component has two aspects: knowledge of cognition and regulation of cognition (Shaughnessy, 1992). Knowledge of cognition included knowledge of strategies and self knowledge of beliefs and attitudes. The interpretation of statistical information in reports might need a systematic approach so that students have knowledge of the strategies to use rather than relying on their intuitions. Regulation of cognition included monitoring how decisions were made under uncertainty and mentally stepping aside to reflect on the process of decision making. If students were aware that their interpretation of statistical information was based on their experiences and perspective (Bartholomew, 1995; Hancock et al., 1992; Barabba, 1991), then they might deliberately search for alternative explanations or judgements.

The comprehension of information in a text should also be addressed from a reading perspective. Resnick (1987) stated that four kinds of knowledge were needed for readers to comprehend a text: linguistic knowledge; topical knowledge; and knowledge about rules of inference and conventional rhetorical structures. Successful readers possessed more of these kinds of knowledge. Cognitive research in reading suggested that for higher order reading, multiple interpretations of texts were part of comprehension. It would seem that this would also apply to the interpretation of media reports from a statistical perspective. The important role of knowledge about the text's subject matter in higher order reading concurred with the findings in this investigation.

Another matter that bore looking at was the role of structuring a statistically based story. According to Resnick (1987) and Curcio (1987), standard rhetorical forms were used by the reader as a scaffold for interpretation. That is, the order in which information was presented played a key role in the reader's understanding and interpretation of the story. Students might not have had enough experience in reading statistically based stories to be able to intuitively understand the framework of the story and hence this would make understanding and interpretation more difficult.

Conjectures about Critical Thinking

From this investigation some characteristics of statistical thinking have been tentatively conjectured. Two modes of thinking could be identified, critical and uncritical thinking, in an item analysis of the interview tasks. The uncritical thinking appeared to be characterised by the first stage of reading and understanding the problem, understanding how the data were represented, and thence giving a description or summary of the situation. This would appear to be an important first stage in interpretive analysis.

The second mode of critical thinking appeared to operationalise higher order thinking skills whereby the student had to produce or create something out of the given situation. This mode seemed to have several identifiable parts. One part reflected critical thinking, whereby 'worry' questions about the data representation and reduction were conveyed. Through the use of subject and context knowledge the student wondered whether the data could be measured, classified, stratified or interpreted in some other way. This part could be classified as critical thinking from a deterministic perspective.

Another part of critical thinking appeared to be conveyed through production of subject and context knowledge with the recognition, for example, that variation should be a consideration (and hence modelling the situation, such as the suicide situation, with a Poisson Distribution might be appropriate). Specifically, questions should be raised on whether the variation was real or random, or what was the impact of the existence of, and the extent of, the unexplained variability on the situation. This part could be classified as critical thinking from a probabilistic perspective. The more statistical tools and context knowledge students had, the more advantaged they should be in these parts of critical thinking. However the first step would be to recognise the situations in which that knowledge should be used.

A further part to critical thinking would, perhaps, be the production of a considered judgement and an insightful interpretation of the statistically based report. It was interesting to note that parallels to these modes of thinking might be drawn from the work of Ben-Zvi and Friedlander (1996) who identified those two thinking modes in students involved in statistical investigation activities.

Conjectures about Causal Thinking

From a statistical perspective it appeared that the students propensity to consider only causal aspects without consideration towards probabilistic aspects was non-normative. However on reflection from a practical realistic perspective it was conjectured that the students ‘were right’. Any problem presented should involve the search for causes and a causal analysis (see Section 9.3).

Conjectures about Measurement

A theme arising from the student interviews was the issue of what quantitative measures and classification measures should be used to interpret the data. A lack of querying such issues by the students would suggest a lack of experience in this area. Part of interpreting or setting up a data-based argument would be to consider whether the measurements reflected the real situation, and whether another measurement would provide another plausible explanation or a new perspective on the situation.

Conjectures for Teaching

It would appear that the interpretation of statistical reports might need to be explicitly taught as the students in this study did not seem to have strategies for interpretation. If interpretation was taught then several implications might need to be considered.

The first conjecture for teaching was that students might not have the context knowledge to interpret reports. Therefore, before interpreting a report, they would need to be given readings to build up their background knowledge, or articles should be chosen that have a high media profile or where the student-context-knowledge base was good. Students would be better to share their context knowledge and subject knowledge through working co-operatively on interpreting an article. Such an approach would in fact be simulating a multi-disciplinary effort that would be prevalent in solving real statistically based problems.

The second conjecture was that students might need a framework as a strategy for critiquing statistical information. An example would be the empirical problem solving framework, based on PPDAC (MacKay & Oldford, 1994) the cycle of scientific investigation, which would require each step in the cycle to be first identified and then critiqued.

> “• ***P****roblem: The statement of the research questions.*
> • ***P****lan: The procedures used to carry out the study.*
> • ***D****ata: The data collection process.*
> • ***A****nalysis: The summaries and analyses of the data to answer the questions posed.*
> • ***C****onclusion: The conclusions about what has been learned"* (adapted from MacKay & Oldford, 1994, p. 1.8).

Such issues as measurement and classification would then be addressed at several points in the cycle. Another tool, such as the fishbone diagram, could be useful for positing possible explanatory variables during the critiquing phase.

The third conjecture was that variation should be experienced using data. An ability to take a more global perspective of data rather than a local perspective (looking at individual data) was an important ingredient of statistical thinking. An understanding of how dice, or how ideas of randomness, could be used as a model for exploring data might be necessary for learning how to process such information. Konold et al. (1991) suggested that a ‘component’ view of phenomenon, that is an understanding that measures could be decomposed into two sources of variation, random (unexplained) and systematic (explained) variation, was needed. However in quality control another view has been taken: *"All variation is caused. Unexplained variation in a process is a measure of the level of ignorance about the process"* (Pyzdek, 1990, p. 102).

To overcome this conceptual barrier random variation might need to be defined as a multiplicity of causes that were chosen to be ignored in the data (Falk & Konold, 1991), and that when a pattern could not be discerned, the phenomenon was modelled as random. From these various definitions it would appear that an understanding of variation in social data might be based on these suppositions: (1) some variation can be explained; (2) other variation cannot be explained on current knowledge; (3) random variation is a model which can be superimposed as a means of coping with unexplained variation; and (4) randomness is a convenient human construct which is used to deal with variation in which patterns cannot be detected (see Section 9.3).

Another problematic area of variation has been the propensity of media articles to focus on a clustering of random events such as a spate of aircraft accidents or an outbreak of a flesh-eating disease. *"Despite our constant exposure to the effects of randomness, few of us have a good grip on what it is, or how it behaves, still less how to generate it"* (Matthews, 1995, p. 38). To understand how such events could occur through random behaviour, specific news or news-type items could be used. These news situations could be modelled and the simulation should demonstrate how such clustering or variation above or below a trend could happen through chance. Students should always keep in mind that there could be a causal explanation. Such an activity should connect ideas between random behaviour in random devices and in social data.

Some other conjectures for teaching involved understanding text information and differing interpretations of data. In order to understand numerical information in text, students could be given raw data and asked to write their own newspaper article using rates and ratios. Beins (1993) found that students who wrote press releases after their analysis of data sets acquired better computation and interpretive skills than students who did not. To understand how different interpretations of data could be made, students should be encouraged to argue and interpret data from different perspectives. De Lange (1987) commented that the critical judgement of statistical data was not easy to teach. One successful method used in his Hewet Project was to take two opposing statements and request students to find out which one was correct by providing them with the statistical material on which the statements were based. The Mathematics in New Zealand Curriculum (1992) also recognised this as a possible method of learning to interpret statistical information. *"Working in groups of three, students collect data on a topic, and use their data to support different sides in a debate"* (p. 194). Begg (1995) has stated that reasoning with uncertainty in statistics and reasoning with certainty in pure mathematics were different types of reasoning, and that perhaps teachers should make students aware of the difference.

6.6 Conclusion

From this analysis of some students' thinking it would seem that teaching has enabled students to learn statistical techniques, language and graphical representation yet it has not adequately fostered the ability to think statistically. Students used multiple frameworks for reasoning statistically. Perhaps this was because of the context of the problem (Watson & Collis, 1994), or that statistics teaching has been compartmentalised into EDA, probability and inferential methods (Biehler, 1994b), and hence the underlying thinking of how to process statistical information has never been interconnected.

Teaching has not ascribed to listening to, or challenging, students' statistical beliefs and therefore students have held on to their own beliefs while learning statistical theory (Borovcnik & Bentz, 1991). If teaching was to be effective then activities, that encouraged students to adopt a critical attitude (de Lange, 1987), and to evaluate their intuitions (Konold, 1991) and their context-knowledge beliefs, ought to be developed. Students should be made aware that cultural conditioning might influence the way data were perceived.

To think statistically a student must have understood and experienced systematic and random variation and perhaps deliberately used the dual modes of thinking probabilistically and deterministically. Ideas of randomness have to be built up from random devices, to all kinds of measurement data, for an understanding of how randomness was used to model such phenomena. This might encourage a fundamentally different way of thinking about the world (Porter, 1986; Falk & Konold, 1992; Moore, 1992; Landwehr et al., 1995). Also, for effective teaching, it would be crucial that context knowledge and subject knowledge should be recognised as underpinning the art of statistical thinking.

Statistics cannot be taught like mathematics. Statistics is moving away from mathematics back towards its roots as a scientific enquiry process (Cobb, 1991; Biehler, 1994a). Statisticians are challenging the teaching of statistics as procedures and emphasise that statistical courses also need to develop students' statistical thinking (Bailar, 1988; Barabba, 1991; Snee, 1993; Wild, 1994). The interpretation of statistical graphs and everyday numerical information in media reports requires the development of a critical attitude and the development of statistical thinking. According to Resnick (1987), some features of higher order thinking are: that it requires the ability to form judgements and interpretations; that it requires effort; that it involves self-regulation of the thinking process; and that it may involve uncertainty. Watson and Moritz (1997) suggest that the questioning of claims reported in the media represent the highest level of statistical thinking. Using these definitions of higher order thinking, this investigation supports the view that the interpretation of statistical information in reports is in the category of higher order thinking. These general definitions may need to be clarified to include some peculiarities of statistical thinking such as the ability to think both probabilistically and deterministically, or both aggregate-based and individual-based. Interpretive skills, which involve a synthesis of ideas from diverse places, will not be as easy to teach as the methods, and procedures, of statistics.

6.7 Summary

Further exploratory studies are needed to investigate the characteristics of good statistical thinking, the role of variation, what skills are required to interpret statistically based reports, and what approaches should be used in the teaching and learning process. In order to ascertain the characteristics of statistical thinking the following aspects have been conjectured as possibilities that should be considered:

- the interdependence of context knowledge and subject knowledge;
- the operation of different modes of thinking:
    - critical and uncritical;
    - probabilistic and deterministic;
    - aggregate-based and individual-based;
- the role of cognition, disposition and metacognition;
- the role of variation;
- the seeking of causes;
- the role of measurement;
- extracting from many different contexts the essential features of the problems;
- ways of developing students’ interpretation skills;
- and the role of frameworks and thinking tools in teaching.

At this stage we made a design decision to interview six professional statisticians in the third exploratory study and six undergraduate project students in the fourth exploratory study. We felt that uncovering some of their thinking processes, during empirical enquiry, would help in the construction of a framework for statistical thinking. Time constraints prevented me from observing and interviewing statisticians and project students, through the course of year-long investigations. Therefore the subjects would be required to reflect on past investigations in which they had been involved.