CHAPTER 2

STATISTICAL UNDERSTANDING

The gradual evolution of the term statistics to its present broad form was discussed in the previous chapter. This included the presentation of a detailed definition of various focus areas within statistics. The chapter also included some research related to understanding in statistics. Key concepts in probability and the feedback from a variety of teaching strategies have been shown to influence a student's statistical understanding.

The main purpose of this chapter is to review the literature which reports on research in the field of students' understanding in statistics and the SOLO Taxonomy. A detailed résumé of research into cognitive development, including the identification and attempted modification of misconceptions in statistical reasoning, can be found in Garfield and Ahlgren (1988, pp.50-56). It is unnecessarily repetitive to outline all this research and so only that research pertinent to the present study is considered in this chapter. A detailed summary of research into probability and statistics by Shaughnessy (1992) is classified broadly as investigations of either describing how people think or influencing how people think.

This chapter contains four sections. Understanding in Statistics, the first section, deals in detail with various aspects of understanding in statistics. Research into misconceptions and errors is considered and then statistical heuristics are discussed as possible solutions to explain students' misconceptions. Some major statistical research projects are outlined, followed by research into specific areas of statistics education. The second section, Levels of Understanding in Statistics, considers the necessity for levels and where to direct the search for such levels. Research into attempts to describe the sequencing of tasks is followed by a consideration of some attempts at assessment instruments. Various researchers' attempts at developing hierarchies are then presented. Third, The SOLO Taxonomy, outlines research which has led to the development of the taxonomy and this is followed by a description of how it has been used. This scheme is presented as a possible framework for describing levels of understanding in statistics. The concluding section, Research Questions, suggests a number of research questions which arise out of the literature review and prompt the present study.

Understanding in Statistics

Compared to understanding in probability, there has been relatively little research into student understanding in statistics. When listing eight major criticisms of the
Statistical Understanding

statistical ability of tertiary statistics students, Oosthuizen (1991, p.62) rated “total lack of true understanding” top of the list. However, no evidence was presented to substantiate the claim, in fact very little research into understanding in statistics was undertaken until the 1970s and 1980s.

This section addresses relevant research into statistical understanding. First, consideration is given to research which has investigated misconceptions or errors in students’ statistical thinking. Researchers in this area were interested in finding out what was wrong with, or missing from, statistical understanding. Second, research which has dealt with statistical thinking, or more precisely, the use of statistical heuristics in reasoning, is outlined. An account of the major projects into, or including understanding of, statistics follows. Finally, those papers which deal with understanding in specific topic areas of statistics are reviewed.

Misconceptions and Errors

Students have many misconceptions about statistical concepts and these, along with other errors, cause many difficulties. This section is divided into four parts. First, the reasons for difficulties are outlined. Next, attempts to develop instruments to detect errors and misconceptions are presented. This is followed by some research into the correction of these misconceptions and finally, some research into correction of errors is discussed.

Reasons for Difficulties

Many errors are made by students when solving statistical problems. Often these errors are due to basic misconceptions but there can be other reasons as well. Garfield and Ahlgren (1987) isolated three main reasons for students encountering difficulties:

(i) the concept was unlike anything the student had met before;
(ii) the concept encountered interfered with intuitive ideas that the student already had; or
(iii) the student may have understood the question incorrectly.

Recent research has been studying the nature of prevalent intuitive preconceptions, as well as isolating common errors, a discussion of which may be found in Garfield and Ahlgren (1988, pp.52-55).

A major project on students’ understanding in mathematics was undertaken by the CSMS Mathematics Team in England in the 1970s and the results were presented in a book edited by Hart (1981(b)). A standard test was designed by interviewing students and then using the information to revise questions. The interviews were used to find out what
methods the students used to answer questions and the errors that they made. To ensure that students did not just repeat methods taught by the teacher, questions were asked in unfamiliar forms. The research was divided into areas dealing with specific mathematical topics. Although the book did not include a section specifically on statistics, it did include one on graphs. While there are numerous misconceptions identified (Hart, 1981(b), pp.120-129), those of most significance to statistics were:

(i) students had difficulty with the idea that there were more points on a line other than the ones plotted;
(ii) students often chose awkward scales for graphs, sometimes making the information impossible to read; and
(iii) students had incorrect perceptions of distance-time graphs, some thought the journey was up and down hills and found it difficult to deal with the abstract notion of ‘distance from the origin’.

The project also included work on developing levels in students’ understanding and this work is discussed later. In the following, discussion of research into the detection of errors is followed by mention of research which attempts to devise means of correcting misconceptions, and errors.

*Instruments for the Detection of Errors and Misconceptions*

There has been little research into the development of instruments which could be used to detect statistical errors. An early attempt at studying errors was made by Boveda (1975) who designed a diagnostic instrument to be used to identify common difficulties that behavioural science students encountered in an introductory tertiary-level statistics course. This concept test was then administered and a detailed analysis of the resulting errors was made. The items in the test were clustered by stratum that represented given content areas and Boveda (1975, pp.101-106) summarized the various errors which occurred in each stratum. Some of the errors of interest were:

(i) failure to distinguish between the average as a central value of a group and variation about an average;
(ii) misapprehension of measures of central tendency;
(iii) misconception about infinite population; and
(iv) misconception of biased and unbiased estimates.

The analysis identified errors made as a guide to a student’s mastery of the subject and led to suggestions of some misconceptions students may have harboured. However, the report did not include any assessment of students’ growth of understanding. Instruments which have attempted to measure this include hierarchies devised to identify levels of understanding, which are discussed later in this chapter.
More recently, the aid of the computer has been enlisted in the development of diagnostic tests designed to describe statistical errors in terms of the conceptual context in which they occur. During its development the intention-based artificial intelligence system, called GIDE, displayed useful features, some of which are:

(i) analysis of computational and symbolic aspects of the solution;
(ii) carrying through student errors in computations;
(iii) implicit matching of steps which students knew but left out; and
(iv) explanation within the context of the problem.

A sample printout of the solution to a problem showed detailed explanation at various lines of the solution. In developing the software, the designers attempted to explain what the error was and why it was an error by using the relationship between a student's solution and a number of possible appropriate solution strategies. As well as describing the software, Sebrechts and Schooler (1987, p.84) suggested a number of reasons why the task of understanding errors was so complex:

(i) students do not follow a single well-defined solution path;
(ii) the solution often reflects only part of the reasoning process; and
(iii) learning procedures depend upon an active integration of specific steps and although students may know procedures they may be confused conceptually and unable to correctly sequence a series of steps.

A number of the principles used in the development could be useful constructs, applicable over other fields as well. The greatest potential of the intention-based approach is its ability to diagnose an error in terms of a specific goal that was intended. As well as detecting errors and misconceptions, there have been various attempts at correction.

**Correction of Misconceptions**

Mevarech (1983) attempted to rectify misconceptions which had been identified. His study investigated a model used by non-mathematically orientated college students in solving problems in descriptive statistics. The procedure used in the experiment is of interest because after the subjects were divided into groups, which were subjected to various introductory courses in statistics, they were given a ‘test’ which played the part of an errant student whose misconceptions had to be identified by the ‘diagnosticians’, the students in the study. The results showed that, although most students recognized the computational formulae required, their knowledge was not sufficient for them to be able to acquire appropriate schema. This meant that, while they were able to recognize that some of the problems were solved incorrectly, they were not able to supply the correct procedures. Mevarech concluded that the results were consistent with other research in that college
students “translate directly from a problem to an equation without regard to the semantics of the problem” (pp.417-420).

Having identified these misconceptions, Mevarech attempted to correct them. It was essential in the corrective activities that understanding was emphasised so that, as well as solving problems, students were encouraged to explain solutions, guess answers before computations and to check that answers were sensible. Results of the program showed that exposure to this diagnostic learning process and feedback concerning correct procedures did increase students’ achievement. Most importantly, Mevarech concluded that students failed to understand concepts of the behaviour of a set of means and of the variance, and that students must engage in corrective activities to overcome these misconceptions. Some research has attempted also to correct other aspects of mistakes which are made in statistical work. It may not always be misconceptions which are causing problems.

Correction of Errors

One interesting aspect of the detection of errors is a student’s own capacity to detect errors. Allwood and Montgomery (1982) attempted to devise a taxonomy of how statistical problem solvers detected their errors. They also investigated how statistical problem solvers responded to instructions given for checking solutions. The subjects were asked to vocalize all their thoughts as they worked on a problem, and if they were quiet for more than fifteen seconds they were reminded to verbalize. When no error detection help was given, more errors were found by spontaneous discovery but when help was given, by way of an instruction manual, more errors were found by routine checking.

The results indicated that there were two main error detection methods, noticing conflicts and repeating the solution. Instructions given on debugging, that is, error detection, seemed to have some effect on making subjects detect their own problem-solving errors and also seemed to have some effect on the way in which errors were detected (Allwood & Montgomery, 1982, pp.136-138).

It would appear that while students will always make errors, it is possible to train them to detect some of their errors. Misconceptions are more difficult to detect and remediate. Some attempt has been made to develop corrective activities but more use needs to be made of activities where students are encouraged to explore, think ahead and explain their way through problems. This aids in a better development of statistical heuristics.

Statistical Heuristics

Some attempt has been made in the study of the use of statistical heuristics, that is, rules for making judgements to explain why people develop misconceptions and
Statistical Understanding

retain them despite contrary experiences. Before presenting their own research and findings Nisbett, Krantz, Jepson and Kundt (1983, pp.340-345) summarized recent work establishing failures to reason statistically and reviewed evidence, both anecdotal and experimental, indicating that people do sometimes reason statistically.

In summarizing this review, they concluded that there was good reason to believe that people do possess statistical heuristics - "intuitive, rule-of-thumb inferential procedures that resemble formal statistical procedures". People appear to be able to apply these heuristics at an early age and "the formal understanding of statistical principles - that is, of the rules governing the behaviour of randomizing devices - increases, at least until adolescence" (Nisbett, Krantz, Jepson & Kundt, 1983, p.345). Even adults, untutored in formal statistics, seem to be able to reason stochastically and to generalize from instances, a task which implies at least a rudimentary application of the law-of-large-numbers heuristic. The discussion also showed that three factors were important in making the application of statistical heuristics easier:

(i) clarity of the sample space and the sampling process which are often difficult in problems in the social domain;
(ii) recognition of the role of chance in producing events; and
(iii) cultural or sub-cultural prescription to think statistically.

Recent research into statistical strategies is attempting to find out when students develop these statistical heuristics and how the development can be encouraged. Some research into the use of statistical strategies in problem solving is now considered, followed by research into recently developed software which encourages the use of statistical strategies.

Strategies in Statistical Problem Solving

A recent study, which attempted to find out how students solved statistical problems when they had not had previous exposure to the ideas, and to determine when students were ready to reason statistically, was conducted by Bakum (1988). The pilot study showed that fourth graders were not ready for statistics but later research indicated that both sixth and eighth graders were ready. Even though no instructions were given on concepts, results indicated that students were ready to understand the concepts. Investigation into the mode of presentation found that greater success was shown, even by higher level students, when they were presented with concrete activities first of all. This may have been due to the fact that these students were ready to move from concrete to formal stages of thinking and benefited from the concrete experiences (Bakum, 1988, pp.84-87). Of particular interest was the fact that students were interviewed as they solved the problems. This procedure was employed as the researcher needed to assess the strategies used.

36
In their studies, Nisbett, Krantz, Jepson, and Kunda (1983) used University students as subjects to investigate the operation of the heuristic of 'generalizing from instances'. The results of the first study showed that much more caution was exercised in generalizing from a single case when the population being investigated was considered by the subjects to be heterogenous than when it was considered to be homogeneous. They also showed that subjects considered it more important to have a large sample size when the population under study was heterogeneous. The second study reinforced earlier research which showed that one is less willing to generalize for a group with which one is more familiar than less familiar. This is probably because the familiarity makes one see the better known group as a more heterogenous one (pp.348-352).

These two studies, along with a third studying the effect of highlighting the degree to which evidence should be regarded as a sample from the population, indicated that subjects were able to reason statistically "when they recognize the heterogeneity of the events in the question and the sample-like nature of their evidence about the events". The final study is of interest because subjects wrote open-ended answers to problems and then these answers were coded as to whether they reflected the use of statistical principles, namely, the law of large numbers and the regression in principle. Having decided that people's intuitive reasoning skills do include statistical heuristics, any formal training in statistics should consist of a refinement of these skills rather than a grafting on of procedures (pp.354-357).

From the above research it would appear that there are statistical strategies which people have a natural tendency to use and these should be encouraged, possibly by the use of games. However, in their research into the cognitive effects of instructional games, mentioned earlier, Bright, Harvey, and Wheeler (1983(b)) tested students at each of four cognitive levels, knowledge, comprehension, application and analysis. Results showed that, although the games were effective at improving performance at the knowledge cognitive level, they were not so at the comprehension or application cognitive levels. A problem highlighted by the research was that basic concepts needed in statistics may be contaminated by students' inaccurate perceptions of 'real-world experiences (pp.166-167).

One particular heuristic, 'representativeness', was considered by Bentz and Borovenik (1984) to be a fundamental strategy in statistics. The problem of selecting samples which are representative of a population is a crucial one. It was not until the second analysis of the results of the research that the idea of representativeness was recognized "to be in the core of statistics". This heuristic allows traits of samples to be transferred to the underlying populations and, though it causes errors in calculations in probabilistic situations when more representative events are thought probable, it is a fundamental idea in statistics.

There are various statistical heuristics which are evident in students' statistical thoughts and it is necessary that these be nurtured and encouraged at every possible
opportunity. One possibility for encouraging people to use statistical strategies is the use of computer software.

*Software to Encourage Statistical Strategies*

The use of the computer allows interactive investigations so that students are able to see quick visually displayed responses to various possibilities in statistical problems. A team of researchers at BBN Laboratories in England have developed a Reasoning Under Uncertainty (RUU) Curriculum which emphasized reasoning and learning-by-doing as methods for helping students to understand statistics. The ELASTIC (Environment for Learning Abstract Statistical Thinking) software uses interactive graphics to support the curriculum, the advantages being that it: interactivity, visualization and dynamic links created a laboratory which students used to explore the underlying meaning of basic statistical concepts and processes. For example, in one section of ELASTIC, Stretchy Histograms, students were able to manipulate the hypothetical distribution as represented in a histogram. As they stretched or shrunk the height of the bars, graphical representations of the mean, median and quartiles were dynamically updated to reflect the changes. Another section, Shifty Lines, allowed students to experiment with line fitting. They could also change the slope and y intercept, and then observe how well the resulting line fitted the data.

The course was designed to support students' reasoning and problem solving in three ways:

(i) the curriculum content provided them with background knowledge they needed to design and conduct their study, for example, how to determine sample size and how to sample randomly;

(ii) the software provided them with the tools they needed to easily explore and analyze data; and

(iii) the concepts were presented graphically and students were able to spend their time and energy on interpreting data and drawing conclusions.

The effectiveness of the program was evaluated using classroom observations and informal reports from teachers. In general, the program was felt to be effective but not always for the same reasons. Teachers of mathematically disadvantaged students felt that the program gave their students insight into qualitative statistical reasoning and a chance for real-world experiences. The effectiveness of the course for mathematically sophisticated students, though, was felt to be the way it forced students to stop and think about statistics in a new way. The results of the open-ended case study problems, which students undertook in both pre and post tests, were unavailable as the time-consuming process of coding and analyzing student responses was still progressing (Rosebery & Rubin, 1989, pp.217-218).
Statistical Understanding

Research has shown that students are ready to reason statistically at an early age but they need to be presented with a learning situation in which their relational understanding of statistics can develop. Well structured games and computer software have proved to be useful tools to this end. To investigate just how students learn various aspects of statistics and exactly what they do understand, some major projects have been undertaken.

Major Projects

Most commonly, research into the understanding of various aspects of statistics deals with the specific areas of graphing, measures of central tendency, such as, mode, mean and median, and measures of spread. Larger projects undertaken by research teams often investigate many of these areas. Four major projects are discussed here, the research being carried out in France, England and the United States.

French Research Project

Although statistics was not being taught in French Schools to 11 to 15 year olds in the 1970s, there was research into the possible development of such teaching. Dumousseau (1983) described research carried out from 1973 to 1978 in France by various institutes for research into the teaching of mathematics, in particular the National Teaching Research Institute. The observations of secondary students, made in their classrooms rather than in a laboratory situation, were used to infer how the student grasped situations when working on statistical projects. Students, working in groups, were given tables of data and time to investigate and organize ideas from the data with as little interference from the teacher as possible. Then the students had to explain their work and justify their actions. Freedom in the choice of subject and the methods employed stimulated the students’ interest (Dumousseau, 1983, pp.147-149).

General results of interest, from this research, showed that students were: not deterred by having to deal with large amounts of information; strongly motivated by real data; ready to undertake long projects with many calculations; and able to work on random phenomena. More specifically, the results of investigations into a number of statistical concepts were presented, the more interesting of these are now outlined. It was surprising the number of students who, when confronted with a problem with two variables, did not use graphical representation to help understand what was happening. However, initial failures and reports to the class for criticism gradually led to the production of progressively more efficient graphs. Although students knew the average, they often did not know what it meant. The mean to most students was a number in the sequence, the mode and the mean were often mixed up, and the median was used frequently. The ideas of ratio, frequency and
proportionality occurred spontaneously in statistical calculations wherever comparisons were made and mastery of the ideas was achieved only slowly and at different personal rates. Exercises which put too much accent on calculation did not favour mastery of the ideas behind these concepts. Students showed little interest in reporting on the spread of data and particularly in representing it by a number. After a suggestion by the teacher, a group presented a report which included a scatter diagram and the other students showed interest in the cloud of points on presentation of this report. Students mainly used graphical methods or tables of spread to investigate correlation.

An interesting result of the research was the awakening in the students of an awareness to random phenomena and prediction. They were keen to make predictions when they knew they would be able to check on them (Dumousseau, 1983, pp.153-157). This project was important for demonstrating that students could improve statistical skills by investigation.

*Concepts in Secondary Mathematics and Science (CSMS) Research Project*

At the same time as the French project, a large research project was underway in England. The major research by the CSMS Team (Hart, 1981(b)), mentioned earlier, included not just methods used by students and the errors they made, but also the development of a hierarchy of levels of understanding. This was based on the analysis of the results of widespread testing. The research covered a number of mathematical areas including ratio, algebra, decimals and fractions but the area of most interest, statistically, was graphing. In deciding on the hierarchy test items were not just graded from hard to easy but were grouped together to form a 'type' and then these groups were ordered into a chain such that success on a harder group of items automatically meant success on an easier group. Six strict criteria were formulated and these had to be satisfied before items formed a group (p.7). The tests, in their final form were designed as a tool to be used by teachers to assess the level which had been attained by students in individual topics. From the grouped questions, the stages found in the levels of learning were situations where student responses showed:

(i) the visual aspect and understanding the meaning of new conventions;
(ii) the application of new conventions - problems may require two steps;
(iii) the first appearance of abstraction - question not always tied to a diagram;
and
(iv) abstraction and a lot of knowledge needed to solve the problem.

Overall, the tests found that 50% of the secondary student population could cope with the first two stages.
Statistical Understanding

More specifically, for graphing, only the first three of the above stages were included because although there were items harder than stage (iii), which were answered correctly, they could not be included in the hierarchy because this was only done by a small percentage of students. Items in the graph test included questions in familiar and unfamiliar forms and three groups were created into which the items fell. These groups were set at the following three levels:

Level 1 - involved plotting points, interpreting block graphs, interpreting scattergrams and recognizing that a straight line represents constant rate.
Level 2 - involved simple interpolating from graphs, recognizing the connection between rate of growth and gradient, using of scales shown on graphs, interpreting simple travel graphs and being aware of the effect of changing scale on a graph.
Level 3 - included understanding the relationship between a graph and its algebraic expression.

Students progressed well through the first two levels but found it difficult to attain the third level of understanding. Hence, although secondary students were capable, with respect to many aspects of graphing, there was a large gap between reading information and an appreciation of an algebraic relationship (Hart, 1981(b), pp.134-135).

A critical assessment of this project was presented by O'Reilly (1990) who was concerned that levels of understanding may be used as a way of "prescribing rather than describing" students' understanding (p.89). The critique also pointed out that the project attempted to match hierarchies across the various mathematical areas or topics as a guide to the relative difficulty of the topics for teachers. However, this was found to be too difficult because the topics had varying numbers of levels and the facility range of items within a level varied from topic to topic.

National Assessment of Educational Progress

As the English team grappled with the description and comparison of levels, assessment of mathematical understanding was also the focus of a large project in the United States. During the 1977-78 school year the second mathematics assessment was conducted by the National Assessment of Educational Progress (NEAP) in the United States. Items were prepared for a test to evaluate students in various content areas at four levels of cognition - knowledge, skill, understanding and application (Ward, 1979, xiii). The sections of the report of most interest here were those on probability and statistics, graphs and tables, and reasoning and making judgements. The Technical Report for the Assessment, released in 1985, included all items used and the data which were collected (Chang & Ruzicka, 1985).
Also, the Detailed Report, which followed in 1986, included the analysis of the various items relating to statistics in eighth grade and twelfth grade (Travers, 1986).

The original report included the questions asked and details of the success rates of students in answering the questions. As far as statistics was concerned the project found that familiarity with the terms mean, mode and median did not appear to increase with age between 13 and 17 years and average was by far the most familiar term. In the graphing and table problems, students were successful at reading information and comparing two sets of information but were not very good at projecting to the future using past trends (Ward, 1979, pp.19-22). This result is consistent with that of the CSMS Project team which also found a large gap between reading information from a graph and an appreciation of the information.

Although reasoning and making judgements are considered to be essential to problem solving, they are also essential to statistics. In test items on these aspects, many students demonstrated the ability to realize that if an object only had one of two required characteristics then it was not the same as the original. However, few students were able to recognize that if matching on both characteristics occurred, then it may be identical rather than certain to be identical. Of interest, is the fact that some of the questions in this section had a ‘not enough information’ option in the answer, and although only 25% of the 9 year-olds correctly selected this option, 53% of the 17 year-olds did (pp.22-23). This would appear to indicate that the ability to analyze the question and recognize what information is required improved with age.

All three projects discussed so far covered a wide range of mathematical topics, of which statistical aspects were only a small part. At a similar time, there was a large project in England which was investigating issues directed more specifically towards statistical education.

Schools Council Statistical Education Project

This project was undertaken in England and Wales in the 1970s with students from 11-to-16 years of age. The Project’s aims including assessing the present situation in statistical education and producing teaching materials (Holmes, 1980, p.13). A further report by the Schools Council Project was produced in 1981 and included: a review of project material; the results of the survey of teaching of statistics in schools; a detailed study of statistics at examination level; and, an overview of the teaching of statistics in various subject areas (Holmes, Kapadia & Rubra, 1981). The project initially surveyed ten percent of schools and concluded that the problems with statistics teaching were, if anything, more widespread than had been suspected.
Units of teaching material were written by the Project team and then students were tested on them. Of most interest to the present thesis, is the hierarchy of statistical concepts and techniques which were developed by the team. While investigating the levels of understanding involved, the team concluded that there was in fact not one hierarchy but several, and elements of the separate hierarchies could be used in parallel. Each concept had different levels of understanding with formal definitions and proofs not being needed until well into the later stages. An approach to statistics suggested was to stress the importance of intuition, and then worry about calculations and definitions later (Holmes, 1980, p.40). After a detailed analysis of the concepts and techniques, Holmes presented flow diagrams which indicated the order of difficulty of the various items. Each of the five main areas, data collection, data tabulation and representation, data reduction, probability, and interpretation and inference, from the definition of statistics considered in Chapter 1, were further subdivided for the purposes of the flow diagrams. These flow diagrams are considered in a later section of this chapter.

Although these projects have supplied important information about statistics teaching and about students' understanding there is still more which needs to be investigated. Apart from the large projects, research has been undertaken which concentrates on students' understanding in particular topic areas of statistics.

Specific Areas of Research

The most commonly studied area in statistics is that of measures of central tendency. Other areas considered are measures of dispersion, sampling and graphing.

Measures of Central Tendency

Given the amount of research that has been undertaken on the mean, mode and median it is surprising that more has not been done to improve the way these concepts are approached in everyday teaching. Lappan (1988) presented a concise review of the research into these three concepts, and included the following important points:

(i) Misconceptions and difficulties in understanding the mean still exist at college level.
(ii) Many students only understood an average as a totally abstract formula, devoid of any meaning.
(iii) Many students chose a central value as the median, whether the list was ranked or not, suggesting an impoverished definition of the median.
(iv) The mode was often identified as the largest number, rather than finding the most frequently occurring data item.
Lappan encountered problems where students appeared to have unsuccessfully applied rote-learnt formulae for calculating the mode and median. This was consistent with the results of Barr’s (1980, p.41) research where responses of students, who were in the 17-to-21 age group, showed confusion when it came to applying basic rules.

Russell and Mokros (1991) tried a different approach for investigating students’ ideas about statistics by focusing on the students’ own constructs of the concept of ‘average’. They found that when students were given algorithms to use, they often lost their intuitive ideas. As students have problems relating formal statistical calculations with their own intuitive ideas, perhaps more should be done to help develop ways of mapping the newer algorithms onto the intuitive methods the students already possess.

Research by Pollatsek, Lima and Well (1981), concerning the appropriate weighting and combining of means in an overall mean, indicated that many tertiary-level students did not understand this concept. Often students used a simple mean when it was not appropriate, which indicated a lack of understanding of the concept. One question, in fact, did not require calculation of a weighted mean and was included to test students’ concept of estimating unknown scores in a sample. Rather than using the given average to estimate, many students calculated the average of the values given so far in the sample, or the value which would make the average of the sample the same as the given average. All these results only reinforce the notion that students have a poor understanding of the mean and tend to apply a formula, no matter what the circumstances.

It is only recently that there have been attempts to investigate students’ understanding of central tendency. One early suggestion, to assist with developing understanding of the mean, was the use of random samples to help understand the distribution of the mean (Leake, 1971). More recently, the balance model has been suggested as useful in helping to understand the mean. Results of research by Hardiman, Well and Pollatsek (1984), with 48 university students, showed that subjects trained using the balance beam learned more than just the mindless application of an algorithm for a weighted mean. They were able to decide between situations when a weighted mean was necessary and those when a simple mean would suffice. Also, when they justified their decisions, they invoked proportional logic. Both the success and the justification indicated that training with the balance beam had led to an increased understanding of the mean.

The use of the balance beam as an apparatus for teaching the concept of the mean was also supported by Cassell (1989). He suggested that the true strength of the mean lay in looking at its mathematical definition in terms of expected value. An earlier suggestion, of finding the centre of gravity of cut-out templates of distributions, he felt had two drawbacks in practice, namely, the difficulty in relating the centre of gravity to a specific numerical value and the need for a student’s understanding of a continuous distribution. As
younger students only have experience with discrete distributions, the simple balance beam was considered far more useful to demonstrate the mean.

As well as the general concept of central tendency, there has been some research into how students link the various measures of central tendency. Lappan (1988) suggested linking concrete experiences to procedures and giving students the chance to develop their own language along with the more conventional strategies of putting variety into problems and letting students collect and summarize their own data as methods of improving understanding. This strategy could apply equally well to other aspects of statistics besides measures of central tendency.

The main problem with measures of central tendency is that students appear to blindly apply rote-learnt formulae, which is more than likely due to a lack of understanding of the concepts. More research is needed to investigate the best way to present these measures to students so as to improve their understanding. This applies even more so to measures of dispersion.

**Measures of Dispersion**

Research into the understanding of measures of dispersion has not been as prolific, nor as comprehensive, as that into the measures of central tendency. As most students meet the concept of range early in their schooling and appear to understand it, discussion here centres on variance and standard deviation. Some researchers have suggested that when more ‘high powered’ statistical topics are introduced they should be handled in an intuitive manner (Love & Lovie, 1976, p.29, Matsumoto, 1981, p.126). Investigative activities would help students to develop a ‘feel’ for the ideas without having to be involved with difficult calculations. This would be particularly useful when learning about variance and standard deviation.

Research into students’ ability to estimate the variance or standard deviation suggested that there is a lack of understanding of these concepts. Lovie and Lovie (1976) showed that students used feedback information from previous estimates in an attempt to establish an interval in which the variance may lie. This research was followed up by an experiment to see if subjects were able to make better direct estimates of standard deviations rather than of variances, and an investigation into the systematic use of the range rule in helping to estimate standard deviations. Lovie (1978) randomly allocated 20 undergraduate psychology students to two different instruction sets. One group had to directly estimate the standard deviation while the other was instructed in the ‘quick and dirty’ method using the range transformation rule. Results indicated that subjects were able, in most cases, to produce better estimates using the ‘quick and dirty’ method.
Statistical Understanding

One aspect of this research by Lovie and Lovie (1976, p.33) which may give a lead as to students' understanding of variance was that the accuracy of the estimates made was influenced by the size of the variance and by the size of the mean as well. This would appear to suggest a relationship between the understanding of the mean and the variance. However, Lovie and Lovie concluded from their work that there was not a detectable relationship between estimates of means and variances. Although, this was contrary to earlier results that researchers had obtained, the findings may be explained by the fact that the instructions, during Lovie and Lovie's exercise, specifically asked students not to take the value of the mean into account when estimating the variance.

Although some researchers have investigated the understanding of both variance and standard deviation, there has been little research into how this understanding may be improved. Having found the balance beam to be so successful as an aid in teaching about the mean, Cassell (1989, p.39) tried to extend this to teaching about the standard deviation by placing a scale at the end of the beam and measuring the deflection of the ends as weights were added. So far, demonstrations and discussions with physicists have only proved this useful in showing that the deflection is other than linear, but the actual standard deviation cannot be measured easily using the deflection. However, Cassell felt that even if an absolute analogy for teaching the standard deviation cannot be found, the use of the standard deviation was preferable to that of other measures of dispersion if a link to the mean could be found. He resorted to earlier suggestions that a demonstration of deviations from the mean, as opposed to other measures of central tendency, gave a good indication of the standard deviation of the data.

Cassell (1989) also discussed situations associated with certain arrangements of data which make the standard deviation more difficult to understand. The standard deviation measures deviation from central tendency whereas many students may simply consider the spread and not assess it relative to any fixed point. When data cluster around non-central points, for example, outliers, the standard deviation does not reflect the idea that data are not widely dispersed. In this case, where data are loosely spread around the mean and tightly clustered around outliers, a balance beam could be useful to demonstrate the dispersion. In fact, although there does 'appear' to be a wide spread, there is a large standard deviation.

Other researchers have also considered whether there may be a link between the standard deviation and the mean. Kelly and Beamer (1986) were of the opinion that the union between central tendency and dispersion is essential. Illustrative examples of questions given to students showed that when they were given additional information about the dispersion of data they were able to make more meaningful conclusions related to averages.

Although a discussion of 'standard deviation' usually implies 'from the mean' Gordon (1986) suggested that a measure of dispersion need not necessarily represent such a
displacement but be thought of as a 'free agent'. Numerically, this could be calculated as the average deviation of each measurement from every measurement in the data. Gordon had been investigating a suitable way of representing the standard deviation on a frequency histogram using a double-ended arrow with the length representative of the amount of deviation. If a measure of dispersion were devised, which did not depend on the central tendency, then a double-ended arrow could be used anywhere on a graph to represent the standard deviation rather than having to decide on the correct placing of the arrow to represent deviation from the mean. The idea of a measure of spread, which is independent of the central tendency, may also help students to understand spread, especially in the situations when the data are clustered around outlying values.

Despite any suggestions of new measures of dispersion which may be useful in improving understanding, it is still necessary for students to develop an appreciation of variance and standard deviation because they are so widely used. More research needs to be undertaken, not just into the understanding of variance and standard deviation, but into how to improve that understanding.

*Sampling*

Students' understanding of sampling at secondary level has not been researched to any great extent as little is taught about sampling. However, if statistics is to be approached from a more practical perspective, and students are to collect data for analysis, then knowledge about sampling and an appreciation of the processes involved are necessary.

Having hypothesized that sampling was in fact common sense, Hunt (1989) gave a questionnaire to 51 first-year tertiary-level students in England. Responses indicated some understanding of sampling. The exercise also served to make the students more aware of the gaps in their knowledge. They appeared to have a good idea of what a sample was, how to select it randomly, and some of the different types of sampling. However, not all countries teach as much statistics at secondary level as England and this may have contributed to the students' knowledge of sampling. For a comparison of statistics curriculums in the United Kingdom, Canada and the United States see Pereira-Mendoza (1989). Also, the answers students gave may simply have been rote learnt, and a more useful exercise to test their understanding might have been to have them actually participate in sampling activities.

Essential to the understanding of sampling is the link between the sample and the population. Research by Rubin, Bruce and Tenney (1991) suggested that secondary students have incorrect models of the relationship between samples and populations. Over-reliance on the representativeness of the sample leads students to conclude that the sample tells everything about the population. In the other extreme, students, who over-rely on
sample variability, may conclude that the sample tells nothing about the population. As the approach to statistics teaching is changing, sampling is becoming more important and further research needs to be undertaken into such ideas develop.

*Graphing*

This is perhaps one of the more important aspects of statistics, as a good graphical representation can assist greatly in helping students to understand various aspects relating to data. A key factor in the understanding of graphs is the development of a left-brain right-brain link and the encouragement of its development is discussed.

Graphs can be used as a tool to help treat what might be considered as ‘high powered’ concepts at an intuitive level. Matsumoto (1981) outlined a teaching unit which used graphs to investigate correlation at secondary level. Thomas (1984) presented a computer program which used graphs to illustrate the Central Limit Theorem. Use of this laboratory session with a group of secondary students made the theorem more ‘real’ to them.

Great care should be taken with the construction of graphs as many representations can be misleading. Strange effects that influence the efficiency of graphic representation may be explained in terms of the functioning of the human brain. The eyes, a major source of information from graphs, are not connected completely to both halves of the brain. The left field of vision (not necessarily the left eye) is connected to the right side of the brain and vice versa. Also, the brain halves have a lot of specialized functions (Vinberg, 1980, p.13). The left half of the brain analyzes logical, analytical, numerical, temporal, verbal and symbolic input, while the right half analyzes intuitive, holistic, spatial, non-temporal, non-verbal and concrete. The right brain can be responsible for a ‘flash of intuition’ but the left brain must methodically work through the detail when faced with a difficult problem. As the right brain does the processing of graphical information, graphs must be designed so that they are suitable for right brain analysis, excluding, as far as possible, talk, text, explanations or numbers.

It is possible that one side of the brain ‘sees’ the graph and the other does the ‘interpretation’ and this may help to explain why there is often a large gap in understanding between reading information from a graph and being able to interpret that information and use it intelligently to describe trends and make predictions. As mentioned earlier, Hart (1981(b), p.135) found such a gap in students’ understanding. Students were able to read information from graphs but they were not able to understand the connection between an equation and a graph. For these students the link between the two halves of the brain may not be as well developed. Details of the levels or stages found, by the CSMS team, in students’ understanding of graphs is given later in this chapter.
To assist in the nurturing of this connection there is a necessity for developing a smooth progression of students' training in graphwork. This was recognized by Gallimore (1989) who stressed the need for greater co-ordination across the curriculum. Students appear to meet a jumble of ideas and are expected to make a great leap in understanding from bar graphs to continuous line graphs with no intermediate steps to help them. More recently, Gallimore (1991) suggested a possible plan for the progression of primary students through graph work.

Students would also be assisted in developing the connection between the two halves of the brain if graphs could be kept simple. As the right brain picks up an impression, without understanding meanings of the numbers, it is easy to distort graphs with optical illusions. For example, diagonal shading of bars at different angles may make them appear to be taller or shorter than they really are. Also, if numbers are included on bar graphs in the body of the graph they may actually distort the visual impression of the lengths of the bars. Unnecessary grids on graphs can also distract from the visual perception that the right brain is attempting to make (Vinberg, 1980, pp.15-24). Graphs also need to be kept simple because psychological experiments have shown that the human mind can deal with, up to approximately, seven 'things' successfully at one time. This appears to be true irrespective of the complexity of the things which are involved. Vinberg concluded that although it may not be possible to specify which graph types are best, the best test is to rely on first visual impressions. We can usually 'recognize' a good graph, one that has been kept simple and on which not too much information has been included. The design of a 'good' graph should greatly enhance the possibility of a student's understanding of that graph. Students find it difficult to use graphs to make predictions because the information is not immediately visibly available on the graph. This was reflected in a study of students in Years 4 and 6 by Pereira-Mendoza and Mellor (1991) who found that the students had few difficulties in reading graphs but often had problems with tasks requiring skills at a higher cognitive level.

A suggestion for helping students to make meaningful statements about data from graphs constructed, is to have a more structured model on which to base their investigations of the information. McCann and McCann (1987) suggested the development of learning-objective models based on sequential abilities. These abilities are: organize and present data; make descriptive statements about data; and analyze the data including generalizations, inferences and predictions. Two learning-objective models, including typical tasks asked of students, were utilized for use with line graphs and stem-and-leaf plots. These models were trialled with secondary students in America and proved successful with both advanced and less able students in helping to produce: quick and easy useful visual forms of data; and a written statement as an end product of investigations.

One way to attempt to gain a better perspective of just how students do perceive graphical or visual presentations it is to allow them to invent their own
presentations. This was attempted by Ałe (1983), who asked junior secondary students to represent random processes. They were given the task of representing the outcomes of a number of coin-tossing trials, prior to having been taught how to draw a tree diagram. They came up with some very unique representations, but one interesting feature was that all the representations involved at least one circular component, representing the coin.

Keeping graphs simple helps students to understand the information that is being presented, but more careful sequencing of ideas in teaching the use of graphs to make inferences may be needed to help overcome what appears to be a rather large gap in understanding. Students need to understand graphs as this may influence their ability to process data. Brown and Silver (1989) when testing students’ ability to process data found that third graders were more competent when data were presented in tabular rather than graphical form. However, older students were equally competent with both forms of data presentation and so this factor may become less important as student become more experienced.

Summary

It is only in the last few years that there has been significant growth in the study of misconceptions and errors in statistics, but already attempts are being made to assist students. It would appear that it is easier to correct errors than misconceptions. Recent research has attempted to develop teaching strategies which assist students to develop healthy statistical heuristics. It is only in the last twenty years that large scale projects into the study of statistical education have been instigated to ensure that such research is ongoing. When it comes to research in understanding in specific areas, measures of central tendency have received the most attention, with work also in the areas of variability, sampling and graphing. The link between measures of central tendency and dispersion is important and needs to be investigated further.

The design and use of instruments to measure statistical understanding is a complex task and little research has been undertaken in this field. The lack of research perhaps suggests the enormity and/or complexity of the task. Diagnostic instruments, developed by Boveda (1975) and Sel rechts and Schooler (1987), have attempted to detect and partially explain the errors and misconceptions which occur in students’ statistical work.

Current trends in assessment in education necessitate the measuring of students’ understanding, or more specifically, the levels in this understanding. Some of the research which has aimed at identifying these levels is considered in the next section.
Levels of Understanding in Statistics

"Statistical thinking is a way of recognizing that observations of the world are not always totally correct" (Rowntree, 1981, p.18). Just as there are varying degrees of 'correctness' in observations, or data, there are also varying degrees or levels in the ability to understand data. As a person's statistical reasoning develops so does the ability to appreciate and understand data and related statistics.

The question then arises as to how this understanding develops. Although a number of researchers have studied errors made by students, misconceptions harboured and attempts to correct these problems, few have delved into the understanding, or lack of understanding, of statistics possessed by students. The work of some of these researchers is presented in this section which is divided into five parts. First, consideration is given to the question of the necessity to investigate levels of understanding. Second, research into the sequencing of tasks is discussed. This is followed by a brief description of some of the models of statistical understanding which have been developed. Fourth, some hierarchies which have been developed are presented. Finally, more specific research into levels in the understanding of the mean and probability concepts are outlined.

Existence of Levels

There is a need to describe statistical understanding. Levels of cognitive growth are suggested as a means for discussing the developmental changes in this understanding. An exploratory study by Bakum (1988), with students from Years 6 and 8, was designed to gain insight into how students move from their first encounter with statistical subject matter which is new to them. In attempting to predict the success of students in solving a statistical problem, various parameters were considered and cognitive skill was found to have the most influence on a student's ability to solve a statistical problem. The Cognitive Skills Index (CSI) was used as a measure of student performance on the Test of Cognitive Skills, as published by CTB/McGraw-Hill. The test had four subtests: sequences; analogies; memory; and verbal reasoning, which measure students' academic aptitude (pp.84-86). Bakum scored an attempt by a student at solving the problem as 0 (unsuccessful) or 1 (successful), with a successful attempt being one where the strategy used would have eventually resulted in the problem being solved, even if the student had not done so. The CSI score was one of the independent variables considered in the logistics model. In most cases this score made a more significant contribution towards accounting for the success of the student than the other variables, that is, grade level, standardized mathematics achievement test score and order in which the problems were presented (p.72). It appears that, generally, the CSI would be an appropriate measure for predicting a student's chance of solving a statistical problem. This implies that those students with more developed
cognitive skills are more likely to formulate a successful strategy for solving a statistical problem, even if no prior instruction had been given.

This dependence of performance in solving statistical problems on cognitive development of the student makes it important to consider the sort of developmental stages that take place in a student’s understanding of statistics. This could then be used as a basis for deciding on a teaching sequence. The study of this development could be achieved by investigating whether levels exist in the student’s understanding. When investigating levels of understanding the question arises as to where to look for levels. There are two types of hierarchies, those which consider development of ‘stages’ in behaviour and those which consider ‘stages’ in instructional content. Hart (1981(a)) considered research into both styles of hierarchies. If levels do exist, are they found in the body of knowledge or do they exist in the student’s mind?

Bakum (1988, pp.2-4) discussed Feldman’s model of cognitive development, which he used as a basis for his study investigating the acquisition of statistical skills by students, and compared it to Piaget’s model. However, the underlying philosophy was different.

In the continuum of developmental achievements from the universal to the unique, considered by researchers, Piaget concentrated on the universal. He was interested in the advances of thought that all students achieved, irrespective of environment, and not in those aspects which make an individual ‘unique’. Feldman, on the other hand, believed that, as well as domains of knowledge that everyone is expected to achieve, there are domains of knowledge which some master and some do not. This unique, or individual cognitive development, has two basic underlying assumptions. First, the stages in the process are invariant. Every individual begins at step one and no-one is allowed to miss a step. Second, there are rules governing the movement from one stage to the next with earlier stages being incorporated into later ones. These two stages of Feldman’s model are consistent with the Piagetian theory. However, his model contrasted with that of Piaget in that Feldman felt that the development was not necessarily spontaneous, nor relevant to all individuals. Both models of cognitive development were based on ‘stages’ but Feldman held the view that the stages existed in the body of the knowledge while Piaget felt that they existed in the student’s mind.

Feldman saw the student as a ‘craftsman’ and recognized the need for disciplines to be considered from a developmental viewpoint. He felt that curricula needed to be organized so that the logic of the discipline did not confuse the developmental levels. Bakum, in his research, considered questions and issues which were consistent with Feldman’s model, having the structure existing in the body of knowledge. The next section considers research into sequencing of tasks, which assumes that there is structure in the
body of knowledge. Later discussions present research which assumes that the structure exists in the student’s mind.

**Sequencing of Tasks**

Early attempts at showing consideration for students’ understanding included the careful sequencing of tasks within the structure of a statistics course. Often this sequencing indicated attempts at allowing for levels within the body of knowledge. Harvey (1975) presented details of a proposal for a statistics and probability course developed at the University of Wisconsin-Madison. The proposal consisted of a complex network of concepts and the sequencing of their presentation.

![Mathematical Probability

Empirical Probability

Inferential Statistics

Descriptive Statistics](image)

**Figure 2.1 - Task Analysis Guideline**
*(Harvey, 1975, p.130)*

The structure shown in Figure 2.1 was used as a guideline to perform a task analysis which resulted in eleven major sections as set out in Figure 2.2. Components of each section were related to those subordinate behaviours immediately preceding them which were a necessary prerequisite. A very detailed diagram was then presented which showed the connection of all tasks as implied by the above scheme. The various tasks and their interrelatedness are too numerous and detailed to mention here (Harvey, 1975, pp. 130-141). The task analysis identified possible sequencing of objectives without specifying a single instructional sequence which should be followed.

The very detailed analysis attempted to order various tasks students would need to be able to accomplish, and sequence them so that a particular task followed on from one which was considered to be a prerequisite for its accomplishment. A problem with this model is that students may start to develop an appreciation for, rather than an understanding
of, a task higher up the order before necessarily fully mastering the task at hand. The model, on the other hand, gives the impression that one needs to have mastered one task before proceeding to its successor in the chain of tasks.

![Diagram of task analysis major sections](image)

**Figure 2.2 - Task Analysis - Major Sections**  
*(Harvey, 1975, p.131)*

One attempt to assist students to appreciate the interdependence of concepts in statistics is the development of computer software, MicroCAM (Nitko & Lane, 1991), which allows students to construct their own interpretations of the structure and interrelatedness of various concepts in statistics. The authors suggest students complete this task before and after a teaching sequence so that changes in their understanding may be assessed. This instrument appears to have its drawbacks as students may be able to connect and use these
concepts in practical situations without realizing what they are doing. It appears that students are being expected to perform a task that even researchers have found very difficult.

Another example of the sequencing of tasks in a course description is Schupp's (1987) outline of a stochastics curriculum for the middle grades (5 to 10). Overall, in the stochastics discipline, a control circuit, as set out in Figure 2.3, was suggested.

![Stochastic Discipline Control Circuit Diagram](image)

**Figure 2.3 - Stochastics Discipline Control Circuit**  
(Schupp, 1987, p.265)

Within this control circuit, a six step learning sequence was outlined (pp.267-269), with each phase carrying on its predecessor, not cutting it off. The six phases are listed below, along with the central concepts to be met in each:

1. Phase: **Experimenting**  
   Chance, label, experiment, trial, outcome, independence, probability.

2. Phase: **Quantifying**  
   Absolute and relative frequency, probability, distribution, uniform distribution.

3. Phase: **Calculating**  
   Secondary probability, expenditure, simulation.

4. Phase: **Characterizing**  
   Mean, median, mode, average deviation, expectation, expected deviation.

5. Phase: **Systemizing**  
   Resuming, precising and structuring the mathematical part of the stochastic knowledge worked out up to now.

6. Phase: **Evaluating**  
   Conditional probability, independence, binomial distribution, sample, test, statistical safety (significance).

A more detailed discussion of the content and central concepts of each of these phases, as well as typical tasks, can be found in Schupp (1989).
The sequencings suggested by Harvey and Schupp were both designed as models on which to base courses. They differ in that a task in Harvey’s model needed to be accomplished before progression, while Schupp allowed a phase to continue on into the next phase. This suggests that an alternative approach to building up a curriculum could be to isolate fundamental ideas which are essential to students, introduce them at an early age and then repeatedly return to these ideas creating a spiralling effect in the learning process. Thus understanding could be developing in a number of areas at once.

Basic Models of Understanding

Some consideration is now given to research which investigates those fundamental ideas that would need to be revisited by students continually. Gawronski and McLeod (1980, pp.86-88) suggested six fundamental ideas upon which a curriculum could be based. They are for students to:

(i) be introduced to a variety of ways of describing and representing data;
(ii) have a variety of experiences with the notion of a sample;
(iii) understand the notion of randomness or of a random variable;
(iv) understand the notion of sample space, or the collection of possible outcomes, in probability;
(v) understand the process of assigning a number (from zero to one) to an event, where the number is the probability; and
(vi) understand the notion of independence.

They believed that emphasis on these six fundamental ideas, although only one aspect of curriculum development, was an important one. It may be possible to find various notions which form the basis of students’ understanding, and use these as a structure for investigating the levels that may exist in the understanding. While investigating levels or cognitive development, some attempts have been made to develop measurement instruments.

During the process of developing a measurement instrument for content learning and problem-solving skills in an introductory statistics course Chervany, Collier, Fienberg, Johnson and Neter (1977) built up a model for the process of statistical reasoning on an elementary level. They proposed that it should consist of three stages as follows:

(i) Comprehension - the recognition of a problem or task as an instance of a more general category or prototype of methods for solving instances of the prototype;
(ii) Planning and Execution - the application of these methods to a specific instance at hand; and
(iii) Evaluation and Interpretation - the evaluation and validity of the outcome from this application against the initial problem.
These stages were then elaborated on with six steps in stage (i) and two steps in each of stages (ii) and (iii). For example, the two steps in the planning and execution stage are shown in Figure 2.4. The explanation also included recognition of the state of the students' understanding at six different points in the process. The points recognized are those at which the students:

(i) understand the problem;
(ii) know the problem type;
(iii) have knowledge of concepts given in the problem statement;
(iv) understand the problem to be solved;
(v) know how to solve the problem; and
(vi) know the problem solution.

All these points are identified within the first two stages of the process.

![Figure 2.4 - Planning and Execution Stage](image)

(Adapted from Chervany, Collier, Fienberg, Johnson & Neter, 1977, p.20)

At a time when various instructional methods had been suggested for introductory statistical courses, and no one had been shown to be better than another, Garfield (1981) attempted to isolate factors which influenced the attainment of statistical competence. The framework used for evaluating this attainment was an adaptation of that used by Chervany, Collier, Fienberg, Johnson and Neter (1977) with the same three general
stages but some alterations to the steps within each stage. The model used for the process of problem solving (Garfield, 1981, p.4) was:

A. Comprehension
   1. Identifying statistical terms.
   2. Identifying known information.
   3. Identifying unknown information.
   4. Identifying the problem type.

B. Planning and Execution
   5. Selecting a statistical procedure to be used.
   6. Applying a statistical procedure to data.
   7. Obtaining a solution.

C. Evaluation and Interpretation
   8. Verifying the solution, checking to see that the solution makes sense.
   9. Interpreting the answer statistically.
  10. Interpreting the answer in terms of the problem.

Then tasks were developed to determine which of the steps, or levels, of the model students were able to complete in solving statistical problems after having been instructed using a problem-solving approach. Garfield (1981, p.193) concluded that the model was extremely successful both in helping to organize the instruction that was given and to evaluate student learning.

The models discussed so far have been attempts to sequence the fundamental ideas needed in the general problem-solving process for a statistical task. It may be that similar ideas could be useful for developing a hierarchy of levels of understanding in statistics and some early attempts are now considered.

**Developing a Hierarchy**

One problem with teaching statistics is that within any one given application several different concepts and techniques may be needed, and different levels of understanding of concepts and competence in techniques may be appropriate. The Schools Council Project on Statistical Education in England, as mentioned earlier, developed several detailed hierarchies of statistical concepts and techniques, the elements of which could be addressed in parallel. This scheme used the idea of introducing fundamental concepts and then returning to them at a later time. The concepts and techniques were introduced into teaching material but were not meant to be completely developed when they were first introduced to the student. It was felt that too great a stress on techniques hindered understanding, whereas they should be an aid in developing understanding (Holmes, 1980, pp.40-41).
A detailed list of concepts and techniques to be met was presented (pp.42-50) under the five main areas Data Collection, Data Tabulation and Representation, Data Reduction, Probability, and Interpretation and Inference. Then flowcharts were used to present a detailed ordering of the concepts and techniques within each of these areas. These are far too numerous and detailed to present in full but an example is the Tabulation flowchart shown in Figure 2.5.

![Flowchart](image)

**Figure 2.5 - Data Tabulation and Representation Flowchart**

(Holmes, 1980, p.52)

Some flow charts had a simple one direction of flow but most had some sort of cross linking as in the above flowchart. The concepts were ordered so that easier ones were met before more difficult ones but it was not necessary for an item to be completely understood before proceeding to a later item. In fact, it may be sufficient to have only met some examples of the earlier item. Thus an hierarchy of understanding was developed based on an interdependent schema of concepts and techniques.

Another approach to developing an hierarchy of statistical understanding was to adapt definitions from an earlier taxonomy. In studying the cognitive effects of the use of games in teaching, Bright, Harvey and Wheeler (1983(b)) used cognitive levels which were defined in terms of Bloom’s 1956 Taxonomy. Of the original six levels, knowledge, comprehension, application, analysis, synthesis and evaluation, only the first four were used because in the USA school curriculum the last two did not appear often enough to warrant their inclusion. Students selected for the study had not previously been exposed to the statistical concepts or skills required for the games. As mentioned earlier in this chapter, all the games tested seemed effective at the knowledge cognitive level but not at the comprehension level or application level. Researchers suggested that the explanation for this probably lay in the students’ need for growth in understanding. They began with essentially no knowledge and had to build on this within a relatively short time span.

This suggests that understanding only develops when sufficient time is given, thus supporting the concept of a basis of fundamental ideas being presented and then
returned to continually in the learning process. Then, within each of these ideas, students develop levels of understanding.

There has been some research recently into levels of understanding (see for example Küchemann (1981) and Boott (1984) in algebra, Van Hiele (1986) and Pegg and Davy (1988) in geometry, and Hart (1981(b)) in various areas of mathematics but little of this work has been in the area of statistics. Two areas, however, where such research has been undertaken, concern the mean and probability notions.

Levels of Understanding of the Mean

There have been at least two studies which have attempted to discover levels of understanding of the mean. The first outlines three types of knowledge about the mean and the second details fundamental ideas about the mean. Pollatsek, Lima and Well (1981) interviewed undergraduate college students as they attempted to solve problems involving the appropriate weighting and combining of means into an overall mean. Many of the students calculated the simple mean when it was not appropriate. For many, it was the only method they had available, but still the fact that they used it demonstrated a lack of understanding of the mean. The authors suggested that there was three kinds of knowledge about the mean, functional, computational and analog.

Functional knowledge was understanding of the mean as a meaningful real-world concept. Results of interviews indicated that many of the students did not have such real-world knowledge, although there was a suggestion that problems involving more concrete quantities, such as, weight, might be easier to deal with than those involving more abstract quantities, such as, marks.

Computational knowledge involved knowing a formula as well as how to obtain the appropriate pieces of information to use in the formula. Although all students knew the formula for calculating the mean, they demonstrated a lack of understanding of the formula as they were unable to find a missing score to make a specified mean. Not many of the students knew the formula for calculating a weighted mean, thus demonstrating little computational knowledge.

Analog knowledge involved visual or kinesthetic images of the mean as a ‘middle’ or balance point. The representation of the mean as a point about which the deviations must cancel out might help students to overcome the gross errors made in calculating weighted means. Almost no evidence was found in the study to suggest that subjects used analog knowledge when dealing with the weighted mean problem. The three types of knowledge discussed, functional, computational and analog, may in fact apply to other statistical concepts as well as the mean.
Strauss and Bichler (1588) investigated how students understood the properties of the mean and how this understanding developed. To determine this development, seven fundamental properties of the mean were isolated as follows:

(i) the average is located between the extreme values;
(ii) the sum of the deviations from the mean is zero;
(iii) the average is influenced by values other than the average;
(iv) the average does not necessarily equal one of the values that was summed;
(v) the average can be a fraction that has no counterpart in physical reality;
(vi) when one calculates the average, a value of zero, if it appears, must be taken into account; and
(vii) the average value is representative of the values that were averaged.

These properties tap three aspects of the concept: statistical (i), (ii) & (iii), abstract (iv), (v) & (vi) and representative of a group of individuals (vii). In all, eighty students were tested over the four age groups 8, 10, 12 and 14 years. The tasks used to test the properties were represented in three different forms, hypothetical (story), concrete and numerical, and the physical matter used in the problem was presented in both continuous and discrete form. The word 'average' was not used in the test and the tasks were designed so that calculation of the average was not required. The student had to make a judgement and then justify it.

There was found to be little effect on understanding by changing the medium used or the type of physical matter. However, results indicated that students' understanding of the mean changes with age and that properties (i), (iii) and (iv) were easier than (ii), (vi) and (vii). From the analysis of students' responses, Strauss and Bichler concluded that there were different developmental paths for some of the properties of the average. As they found two levels of difficulty they suggested that the curriculum developers construct different units for the average for the two levels. It was also interesting that the distinction between continuous and discrete materials, a factor in understanding, was not found to be as important as Piaget suggested.

Despite the detailed analysis of levels of understanding of the mean, little research has been undertaken into levels of understanding in other aspects of statistics. There has been some research, however, into levels of understanding in the related field of probability. Although probability is not of direct interest to this study, some of the research into levels of understanding in probability is worth noting.

**Levels of Understanding in Probability**

Two suggested hierarchies of levels of understanding in probability are discussed. The first, is a series of levels based on experimental information and the second is
a totally theoretical framework. From the results of the Chance and Probability Concepts Project, Green (1983(a)) concluded that there were levels in the understanding of probability concepts ranging from 0 to 3. Such a level was assigned to each student and the results showed that there was a steady increase in level with age and that average students entered secondary school (age 11) at Level 1 and left (age 16) at Level 2. A typical item from each of the three Levels 1, 2 and 3 was discussed briefly and a Guttman Scalogram Analysis of 58 test items allowed conceptual levels to be assigned to items. Most students did not attain the level of ‘formal operational’ (Piagetian Stage III) by Year 5 (17 years old) and presumably left school at the ‘concrete operational’ level (Stage II).

Olecka (1983) applied Vienne’s six-stage process, for going from concrete to formal operational in the development of mathematical concepts, to probability teaching. The six stages were:

(i) initial reaction;
(ii) discovery of regularities in situations;
(iii) searching for isomorphisms - comparison of several games possessing the structure;
(iv) representation of isomorphic situations in one form;
(v) study of the representation by description of its properties - usually symbols are needed at this stage; and
(vi) formalization of the system.

As the task was too difficult to perform globally, the stages were applied separately to various probability concepts and then the models combined. A table was presented that linked the stages in order for seven different probability concepts. However, the table was only a suggestion and has not been subjected to experimental verification.

Both research projects have aspects which are of particular interest. Green’s Project was able to suggest levels and to pinpoint the stages achieved by students during secondary schooling. Olecka attempted to describe levels in different areas of probability and then link them together.

Summary

The research considered in this section highlights some important points. A student’s performance in solving statistical problems depends on his or her cognitive development. Levels do appear to exist in this development but as yet, measurement instruments have proved difficult to construct. There have been attempts to develop general hierarchies which may be used to explain these levels.

The need for assessment instruments for understanding has been recognized by many researchers. As Jolliffe (1991(b)) pointed out, when dealing with large groups of
students it is difficult to ascertain formally whether concepts are understood. Questions need to be developed which can be used to assess understanding but allow marking to be kept within realistic bounds. It is possible that open-ended questions may be useful as an instrument for measuring students' understanding.

More recently, the SOLO Taxonomy has been used to attempt to explain student understanding in many areas of study and may be useful to explain statistical understanding. The next section investigates the development of the taxonomy and compares it to other taxonomies.

The SOLO Taxonomy

Over the years, taxonomies have been created in attempts to classify, sort or explain many different phenomena. One theoretical framework that has been used to explore students' understanding in a variety of topics is the SOLO (Structure of the Observed Learning Outcome) Taxonomy. This taxonomy attempts to provide a language which can be used to categorize the levels of students' responses at the various stages in the development of understanding. Examples of the application of this taxonomy include algebra and geometry (Pegg, 1992), area (Watson, Chick & Collis, 1988), fractions (Watson, Campbell & Collis, 1993), volume (Campbell, Watson & Collis, 1992) and science concepts (Levins, 1992; Levins & Pegg, 1993).

Prior to 1992 there was no study investigating the use of the SOLO Taxonomy to help explain statistical reasoning, despite its application in many other areas. Before considering the implications of the SOLO Taxonomy to students' understanding of statistics it is appropriate for exploring the theory in detail. The first part of this section is a brief overview of the development of the SOLO Taxonomy. The second part investigates possible styles of measurement which may be used in conjunction with the taxonomy. The next part considers applications of the taxonomy. Finally, the SOLO Taxonomy is compared with other categorization systems.

Development of the Taxonomy

Since the concept was first introduced in 1979, the SOLO Taxonomy has steadily evolved into its present form. The original taxonomy resulted from a study by Collis and Biggs (1979), which attempted to devise a taxonomy educators could use to analyze student responses to tasks, and hence evaluate student outcomes (Biggs & Collis, 1980). The taxonomy classified responses on a series of levels described by the words prestructural, through unistructural, multistructural and relational to extended abstract. These
levels range from the pre-structural responses, where irrelevant cues are used to produce a response, through to extended abstract responses, where many related and given, and, also, related and hypothetical, cues are used to produce a number of interrelated responses. The scale was hierarchical in that each level built on the previous one. Detailed descriptions of each of the levels in the hierarchy given in Biggs and Collis (1982) included diagrammatic representation of the structure of the responses at each level. This diagram is reproduced in Figure 2.6.

Further research brought to light situations where the SOLO Taxonomy was found to be lacking. Two distinct aspects of the sequence and structure of learning were observed and these were labelled the 'mode' and the 'structure'. The 'mode' was defined as the level of abstraction of the contents learned and the 'structure' as the SOLO level of those contents within a 'mode'. This structure formed a 'learning cycle' of five general stages within each mode, the previously mentioned prestructural, unstructural, multistructural, relational and extended abstract. Once an individual integrated all facets into one mode of functioning then extension was made into the next mode (Collis & Biggs, 1983, p.153).

The more advanced structure of the SOLO taxonomy was consolidated by Biggs and Collis (1991, pp.62-64) who presented five developmental modes with the SOLO levels existing within each mode. These modes or stages (with approximate ages) were:

- **Sensori-motor** (from birth) - interaction with the world is in the most concrete way, a motor response to a sensory stimulus.

- **Iconic** (from around 18 months) - actions are made more abstract by internal representation in some form, thought draws heavily on imagery and is frequently affect laden.

- **Concrete-Symbolic** (from around 6 years) - significant shift in abstraction from direct symbolization of the world through oral language to symbol systems that apply to the experienced world.

- **Formal** - (from around 16 years) - a superordinate abstract system used to generate hypotheses about alternative ways of ordering the world.

- **Post Formal** - (from around 20 years) - questioning the conventional bounds of theory and practice, and establishing new ones.

Within this cycle of learning, four major transitions, from one mode to the next, occur. Of particular interest to educators are not just the features of the shifts but where exit levels from the various modes relate to schooling (Biggs & Collis, 1989, pp.158-161).

The basic features of the five SOLO levels which are considered to exist within the above modes (Biggs & Collis, 1991, p.65) are:
Figure 2.6 - SOLO Response Structures
(Adapted from Biggs & Collis, 1982, pp.24-25)
**Prestructural** - The task is er gaged, but the learner is distracted or misled by an irrelevant aspect belonging to a previous stage or mode.

**Unistructural** - The learner focuses on the relevant domain, and picks up one aspect to deal with.

**Multistructural** - The learner picks up more and more relevant features, but does not integrate them.

**Relational** - The learner now integrates the parts with each other, so that the whole has a coherent structure and meaning.

**Extended Abstract** - The learner generalizes the structure to take in new and more abstract features, representing a higher mode of operation.

As later modes evolve, earlier developed modes can also operate in parallel to them, making the model multimodal. Within the mode structure, four distinct approaches to learning that students may traverse were distinguished (Biggs & Collis, 1991, pp.66-67). These are represented diagrammatically in Figure 2.7. It should be noted that the extended abstract level in one mode is equivalent to the unistructural level in the next mode.

![Figure 2.7 - Modes and Learning Cycles](image)

(Adapted from Biggs & Collis, 1991, p.66)

Apart from the course of optimal development (depicted by line A), where the transition from relational in one mode to unistructural in the next is a key factor, and the

66
unimodal course of learning within one mode (line B), there are two multimodal paths. These are the ‘top-down’ facilitation of lower order learning (line C) where one uses higher levels of understanding to enhance operation in a lower mode and the ‘bottom-up’ facilitation of higher order learning (line D) where lower levels are invoked to enhance performance at a higher level. The authors presented practical examples of each of these four paths.

Any particular skill has a target mode. This is the base to which supplementary modes, when relevant, may be added as aids. It is important for a student to be able to focus on the target mode and then use the other modes as supplements. To facilitate this, teachers need to be able to recognize characteristics of each mode and encourage appropriate use. The ‘plus-one’ strategy may be useful. If a student indicated performance at a particular level then the teacher should try to pitch instruction at the next level up. The move from the iconic mode into the concrete-symbolic mode is the first shift where formal instruction is necessary and this is one of the reasons why schools are maintained (Collis & Biggs, 1991, pp. 99-205).

**Figure 2.8 - Possible Course of Action for Cognition Task**
(Collis & Romberg, 1991, p.103)
Further aspects of the relationship between the iconic and concrete-symbolic modes have been revealed by research into a problem-solving path by Collis and Romberg (1991). This path is reproduced in Figure 2.8, on the previous page. In the problem-solving path a respondent begins by choosing either an iconic or concrete-symbolic path. However, at stages B or C, it is possible that interaction between the two modes can take place.

Later research has shown that the model of cycles within modes may in fact be more complex than that shown in Figure 2.7. When analyzing responses in an attempt to sequence responses in the measurement of volume Campbell, Watson and Collis (1990) found that the unistructural-multistructural-relational (U-M-R) SOLO development sequence occurred, not once, but a number of times within the concrete-symbolic mode. Pegg (1991) identified a similar structure.

As research has progressed the SOLO model has expanded to incorporate new ideas, while at the same time maintaining the validity of initial perceptions. This is one of the desirable features, or strengths, of the taxonomy, that the basic framework has been able to change in response to attempts by researchers to describe response structures.

**Taxonomy Measurement**

As the taxonomy developed so did the need for instruments to measure the levels identified. Both open and closed format questions can be used for this purpose. Both style of question are discussed in detail by Collis (1986). Open-format questions give the student a question or task which requires a response drawing on long-term memory to complete. The responses to such questions or tasks are then scored to a particular SOLO level depending on the understanding exhibited. Examples of the grading of responses to a history task (pp.14-16) and a mathematics item (pp.17-19) were presented in detail. Closed format questions are designed so that there are a number of parts, each of which requires a response at a higher level than the previous part. These questions combine the super-item concept with the cycle of learning of the SOLO Taxonomy and require that the hierarchy of levels is known before the question is designed. Collis (1986) also discussed gradings of responses to a mathematics item (pp.20-21) and a science item (pp.22-23).

Both styles of question have shown positive and reliable results when used to assess the SOLO level of responses (p.24) and so the choice of assessment style may depend on other factors. If the hierarchy of levels is known, a closed-format question may be used as it makes grading of responses less time-consuming, even though the actual questions require careful design. However, if the hierarchy of levels is not known, and needs to be investigated, then an open-format question style appears most appropriate.
Taxonomy Applications

Apart from continuing research to improve on the model, the SOLO Taxonomy has other uses, such as curriculum and assessment tool design. The taxonomy is proving invaluable in assisting in the development of school curriculums where previous attempts to sequence concepts have had little theoretical basis. It is being used also to develop assessment tools which are consistent with the philosophy of newly designed curriculums.

The value of the early research became more practical when Biggs and Collis (1989) showed the possible applications of the SOLO Taxonomy in the development of school-based curriculum. The taxonomy was presented as "a language for describing the level and quality of learning both within and across the curriculum" (p.151). Features of the exit levels of the four transitions in the concrete-symbolic mode were described. These are the key stages in the educational development of a student in secondary schooling. A detailed example of an application in History is given. Importantly too, it was suggested that even though the target mode for secondary schooling is concrete-symbolic, educators need to keep in mind that developmentally, prior modes also need to be nurtured. The beauty of this approach to curriculum design is that it is applicable across all Key Learning Areas.

However, if these levels are to be used as a guide to curriculum design, then there also needs to be changes in assessment procedures. Recently, there has been a push towards process-orientated tests, and hence a need for a cognitive framework for examining responses to open-format test items. Collis and Romberg (1991, pp.93-98) considered features of mathematical thinking in the various modes and presented possible courses of action as outlined earlier in Figure 2.8. They perceived the biggest problem to be that, although the curriculum emphasized multimodal functioning, the assessment recommendations showed a bias towards concrete-symbolic functioning. Examples of tasks used to test abilities in the iconic or concrete-symbolic modes were presented, along with a third type of item, which set out to test both modes. To analyze responses to these items a sample of responses would be supplied which could be used to decide whether a particular response did, or did not, satisfy the category criteria determined by the original task. Such an assessment is multimodal, and so can not be assigned a number in one dimension. An 'aims by level of mode' matrix was suggested for organizing the classification of sample responses to draft items (pp.113-114). A large bank of problems with sample responses in different modes was supplied (pp.116-125).

The SOLO Taxonomy has been used also as an infrastructure for creating a mapping procedure for analyzing the structure of mathematics responses (Collis & Watson, 1991). This technique allows an answer to be classified by the way in which it is structured. A detailed description of sample-response structures was given, showing various features of
the five SOLO levels. Questions were designed in each section attempting to elicit responses at each level. Maps of routes of responses were given at each level from unistructural through to extended abstract. One advantage of these maps is that when a student makes an incorrect response the reasoning path can be pinpointed using the map.

In both cases, open-form at test items and mapped responses, attempts are being made to elicit answers which allow educators to assess the level at which a student is able to operate. At this stage more research is needed to produce a clear picture of what features ‘typical responses’ exhibit at the various levels. The changing philosophy will, however, see major changes to both curriculum and assessment procedures in the near future. It is now reasonable to ask why the SOLO Taxonomy should be used as a basis for these changes when there are already other taxonomies available.

Comparison with other Taxonomies

Many attempts have been made over the years to create hierarchies to describe understanding. These hierarchies can be used to describe a teaching sequence, a logic sequence within a topic, or a sequence of understanding within a student. For comparison of a number of different hierarchies in these areas see Hart (1981(a)). The present section only attempts to demonstrate why the SOLO Taxonomy is more useful than the two well known taxonomies, namely those developed by Bloom and Piaget.

The Bloom Taxonomy is widespread in its use. It was specifically designed to order test items in terms of hierarchical levels of response. Although useful, this taxonomy has its limitations. First, the taxonomy was intended to assist with item selection in test construction. This means that it was not suitable to evaluate the quality of a student’s responses and hence difficult to apply meaningfully to responses to open-format questions. Second, the design of questions to draw out responses at any of the higher levels is difficult. Often the judgements about the quality were arbitrary and made a priori by teachers (Biggs & Collis, 1982, p.13). Hence, the SOLO Taxonomy is more useful than Bloom’s because it can be applied to open-ended responses and the levels used have arisen ‘naturally’ out of student attempts at understanding.

Piaget described cognitive development in discrete stages but the model was very restrictive as the levels were tied too closely to age. Once a student was at a particular level for one task, then he or she was supposed to be at that level in any task. Research has shown that this form of categorization has limitations also. First, the theory suggested an evenness within a domain of performance and also across domains. Later research, however, has shown that there this was not so across domains. For example, a student can be ‘early formal’ in mathematics while still being ‘early concrete’ in history. Second, each stage of development was considered to replace its predecessor. More recently, this idea of
replacement of preceding stages has not been confirmed by empirical evidence. For example, Biggs & Collis (1991, pp.60-53) have shown that their modes actually enhance performance within other modes. The basic problem with the Piagetian model is that it is a classification of the student and not of the response. SOLO is more useful because it allows students' responses to be at different stages in different domains and allows multimodal functioning to occur.

The taxonomies developed by Bloom and Piaget have proved useful in the past. However, research has shown that both have downfalls. The SOLO Taxonomy goes a long way towards compensating for these downfalls. The main advantage of SOLO is that it has evolved through analysis of student responses, gathered by interview and test procedures. This means that the model changes consistently with results of analyses of the variable to be measured. Hence, there is a self-correcting feature associated with SOLO's development.

Summary

The evolution of the SOLO Taxonomy over the past fifteen years has been an ongoing process. As with other taxonomies, SOLO was devised to improve on observed short comings in other taxonomies. Then, as research has brought new facts to light, the model has grown to absorb these findings. The taxonomy is proving very useful for describing the levels of understanding in student responses. In the educational sphere, the taxonomy is a useful guide for the design of both curriculum and assessment tools. Research continues to show that it is more versatile and realistic than most other taxonomies suggested to date. At this stage all evidence appears to support the continued use of the taxonomy in a wide variety of areas.

Conclusion

This chapter has identified extensive research into certain aspects of understanding in statistics. Common misconceptions and errors have been analyzed. Instruments have been designed for the detection of misconceptions and errors. Also, it has been found that students have a natural tendency to possess certain statistical heuristics, and the use of these should be encouraged. More in-depth research has been undertaken in a number of specific areas, such as, statistical measures and graphing. These studies included attempts to improve students' understanding. However, the major problem with all this research is the actual measurement of students' understanding. Being able to assess and describe this understanding will make it easier to determine the effects of change in the
teaching environment on this understanding. There has been some research into the measurement of students’ understanding of statistics but not as much as in other areas of the mathematical sciences, such as algebra or geometry.

It is only recently that research has endeavoured to investigate levels of understanding and find a taxonomy which might describe these levels. Early attempts at describing levels mostly involved sequencing tasks, which meant that the search for ‘levels’ was concentrated on the instructional content. More recently, research has concentrated on the ‘stages’ which exist within a student’s behaviour. In the last decade, the search for a suitable taxonomy to describe the level of understanding exhibited by student responses was greatly boosted by the contribution of Biggs and Collis and their development of the SOLO Taxonomy. This taxonomy has many advantages over other previously proposed taxonomies and its use as an aid for developing instruments to measure students’ level of understanding has become widespread in many subject areas.

Research into devising a scheme to measure understanding has proven difficult, but it appears that levels do exist. Recently, the SOLO Taxonomy has been used to assess student responses, with repeated success, in a number of subject areas. In consideration of the framework from Chapter 1, it was considered necessary to find a method of measuring students’ understanding of statistics. Having now considered research on understanding and levels of understanding in this chapter the following research questions become relevant:

Q 2.1 Is there a hierarchy of growth of students’ statistical understanding in statistics?

Q 2.2 If this hierarchy can be identified, does SOLO offer a broad framework for its explanation?

Hierarchies have been found to exist in students’ understanding in many areas of mathematics so it is reasonable to expect to be able to also find them in statistics. Consideration of trends in the literature review in Chapter 2 suggests a general template which may be applied to student responses to indicate increasing sophistication of statistical thinking. This template has three broad categories. First, are those who are only able to reproduce information given in the question and not refer specifically to the data. Second, are those who ‘look’ at the data but are only able to ‘see’ features or properties. Third, are those who are able to explain how the features of the data can be used for some statistical task, such as, prediction. These three categories are appropriate to be used as a general framework on which to attempt to build a hierarchy.
If it can be established that an hierarchy exists, then an attempt will be made to use SOLO to explain the range of responses. The present study is mainly concerned with secondary students' understanding and so the ikonic and concrete-symbolic modes are of most interest. Also of interest, are the cycles of levels within modes which have been observed more recently (Pegg, 1991). Thus further questions which need to be addressed, if the SOLO framework is applicable, are:

Q 2.3 What are the descriptions for the SOLO modes and levels?

Q 2.4 Do cycles of levels exist within the SOLO modes?

Researchers have found that one of the most important moves in secondary teaching is from the ikonic to the concrete-symbolic mode (Campbell, Watson & Collis, 1992). Therefore it is of the upmost importance that something be found which is able to mark the transition from the ikonic mode to the concrete-symbolic mode. From the SOLO discussion in this chapter, it is evident that responses in the ikonic mode are personal and intuitive in nature and often draw on imagery and everyday language in their expression. However, responses in the concrete symbolic mode are more process related, expressed in a more symbolic language and are no longer based on appearances. Of particular interest are the features of responses to statistical questions which could be used to distinguish between the two modes, hence the question:

Q 2.5 What are the major differences between responses in the ikonic and concrete-symbolic modes?

If SOLO proves to be a suitable tool for classifying students' understanding, then it will be interesting to know whether certain factors influence the level of a student's response. Two factors suggested in the literature are academic year and mathematical ability. While gender is not identified, specifically, it was felt appropriate that this should be included also. This issues result in three research questions:

Q 2.6 Does a student's level of understanding of statistics depend on the student's academic year?

Q 2.7 Does a student's level of understanding of statistics depend on the student's mathematical ability?
Q 2.8 Does a student’s level of understanding of statistics depend on the student’s gender?

The assessment of the models is a difficult task and it is going to be necessary to be able to open up students’ thinking. Open-format items are to be used in this study and the responses coded. If the hierarchy is found to exist, then the robustness of the levels needs to be considered.

Q 2.9 Will the performance of a range of students in successive chronological years produce a similar set of conclusions concerning levels of understanding?

Q 2.10 Has the form of the assessment, that is, written test, been able to accurately measure the students’ level of understanding, as compared to an oral test?

Q 2.11 In an interview situation what are the effects of prompting?

Statistics is mainly covered as an integrated part of the mathematics course in Australia. The statistical components of the mathematics courses for Year 7 to 12 appear in Appendix B. Some work is covered in junior secondary years and lower level senior courses but statistics is greatly lacking from the higher-level courses. As there is no major focus on statistics in the New South Wales curriculum, a diverse range of everyday experiences influence a student’s statistical understanding. Of interest is whether there is a growth in a student’s statistical understanding over time, hence the final research question:

Q 2.12 What are the influences of one year’s schooling on students’ level of understanding?

The twelve research questions implied a need to split the research into two separate studies, if all issues were to be addressed thoroughly. The purpose of the first study was mainly to establish whether an hierarchy existed, whether SOLO was suitable to describe the levels, and how these levels related to academic year, mathematical ability and gender. The second was a longitudinal study to address the issues of robustness and growth.

Aspects of earlier research also highlighted some considerations which need to be made with respect to the design of the study. These include the sample size and the style and range of questions. Thus far, research into levels of understanding have involved relatively small sample sizes. Larger samples need to be taken so as to be able to investigate more deeply the levels of responses and to be able to look in depth at each academic year.
This also allows the influence of other factors, such as, mathematical ability and sex of the student to be taken into account. Questions given to students in previous research into SOLO levels have often been designed so that they are testing at a certain level. These short answer or multiple choice style questions restrict the amount of information that students are able to convey. Research has shown that open-format questions are best for soliciting such a response because of the tentative nature of current research into statistical understanding. The range of topics covered by questions investigated so far has shown a strong bias towards probabilistic concepts, as earlier statistical research has done. What is needed is research covering a fuller range of statistical concepts, such as those identified by Holmes in his definition.

The above concerns were taken into consideration when designing the present study to address the proposed research questions. The following chapters outline the research and attempt to answer the questions detailed above. The design and results of the first study are presented in the following three chapters.