DEVELOPING CRITICAL NUMERACY AT THE TERTIARY LEVEL

by

Marian Kemp

A thesis submitted in part completion of the requirements for the degree of Doctor of Education

Murdoch University

August 2005
DECLARATION

I declare that this thesis is my own account of my own research and has not previously been submitted, in whole, or part, for assessment at any tertiary institution.

Marian Kemp
August 2005
Western Australia
CONTENTS

DECLARATION

CONTENTS

PREFACE

ABSTRACT

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

1 INTRODUCTION .................................................................1

1.1 Overview of the Dissertation
1.2 Introduction to Chapter 1
1.3 Numeracy
1.4 Numeracy Demands
   1.4.1 Everyday Life
   1.4.2 Community Life
   1.4.3 Learning
   1.4.4 Work
   1.4.5 Recreation
1.5 Summary

2 LITERATURE REVIEW ..........................................................24

2.1 Introduction
2.2 The Concept of Numeracy
2.3 Perspectives of Numeracy
   2.3.1 Numeracy as Mathematical Skills
   2.3.2 Numeracy as Mathematics in Everyday Life Taught in the
       Mathematics Learning Area
   2.3.3 Numeracy Across the Curriculum
2.4 Developing Numeracy
   2.4.1 Learning and Using Mathematics
   2.4.2 Developing Cognitive Skills
2.5 Learning Numeracy Through Other Subject Areas
2.6 Reading and Interpreting Tables and Graphs
2.7 Summary
3 THE EMPIRICAL STUDY

3.1 Introduction
3.2 A Pilot Study
3.3 The Aims of the Study
3.4 The Sample
3.5 The Table Interpretation Tasks
3.6 The Measuring Instrument
3.7 Coding of the Tasks for Analysis
3.8 Summary

4 QUANTITATIVE ANALYSIS

4.1 Introduction
4.2 The Rasch Model
  4.2.1 The Rasch Model for Polytomous Responses
4.3 Analysis using the Rasch Model
  4.3.1 Person Separation Index
  4.3.2 Item Characteristic Curves and Associated $\chi^2$ Values
  4.3.3 The Category Characteristic Curves
  4.3.4 The Relationship Between the Person Distribution and the Locations of the Thresholds
  4.3.5 The Relationships between Locations of Thresholds and the Item Distribution
4.4 Statistics for the Initial Analysis
4.5 Category and Item Characteristic Curves for Initial and Final Analyses
  4.5.1 CCC and ICC for the Initial and Final Analyses of Environmental Concerns 1
  4.5.2 CCC and ICC for the Initial and Final Analyses of Suicide 2
  4.5.3 CCC and ICC for the Initial and Final Analyses of Homeworkers 1
  4.5.4 CCC and ICC for the Initial and Final Analyses of Homeworkers 2
  4.5.5 CCC and ICC for the Initial and Final Analyses of Suicide 1
  4.5.6 CCC and ICC for the Initial and Final Analyses of Environmental Concerns 2
4.7 Summary

5 QUALITATIVE INTERPRETATION OF THE MODIFIED RESPONSE CATEGORIES OF THE TASKS

5.1 Introduction
5.2 The Measuring Instrument
  5.2.1 Final Categories for Environmental Concerns 1, Suicide 2 and Homeworkers 1
  5.2.2 Final Categories for Homeworkers 2
  5.2.3 Final Categories for Suicide 1
PREFACE

During the course of my career as a teacher at secondary and tertiary levels it has become increasingly apparent to me that many students, from all kinds of backgrounds, tend to skim over those aspects of information that are quantitative. It seems that many lack the confidence to engage with the quantitative material, be it in the form of tables, graphs charts, embedded in the text or written in prose, while others may not be aware of the importance of dealing with this kind of information themselves.

This is of concern for several reasons. In the context of education, students need to be able to research information about their chosen area of study. They need to be able to engage with available data and be able to understand what it represents and some of the implications. They also need to be aware that there is a range of interpretations that can be made based on the same data. Interpretations are always influenced by the perspective of the person interpreting the data and when these are reported in print or other forms of media, they may reflect some element of bias. Without the confidence and ability to engage with the reported information the reader may be misinformed, or mislead by the author’s own viewpoint.

An awareness of the need to encourage and enable students to engage with the quantitative material led to the study on developing students’ numeracy to be reported on in this thesis.
ABSTRACT

Students at university encounter quantitative information in tables and graphs or through prose in textbooks, journals, electronic sources and in lectures. The degree to which students are able to engage with this kind of information and draw their own conclusions, influences the extent to which they need to rely on the interpretation of others. In particular, students who are studying in non-mathematical disciplines often fail to seriously engage with such material for a number of reasons. These may include a lack of confidence in their ability to do mathematics, a lack of mathematical skills required to understand the data, or a lack of an awareness of the importance to able to read and interpret the data themselves. In this dissertation, the successful choice and use of skills to interpret quantitative information is referred to as numeracy.

The level of numeracy exhibited by a student can vary depending on the social or cultural context, his/her confidence to engage with the quantitative information, the sophistication of the mathematics required, and his/her ability to evaluate the findings. The first part of the dissertation is devoted to the conceptualisation of numeracy and its relationship to mathematics.

The empirical study that follows this is focused on an aspect of numeracy of importance to university students: the reading and interpreting of tables of data in a range of non-mathematical contexts. The students who participated in this study were enrolled in degree programs in the social sciences. The
study was designed to measure the effectiveness of a one-hour intervention workshop aimed at improving the levels of the students’ numeracy. This workshop involved reading and interpreting a table of data using strategies based on the SOLO taxonomy (Biggs & Collis, 1982).

The SOLO taxonomy was developed mainly as a means of classifying the quality of responses across both arts and science disciplines rather than classifying the students. The categorisation uses five levels: prestructural, unistructural, multistructural, relational and extended abstract. It can be used as a diagnostic tool at all levels of education as it can be seen as a spiral learning structure repeating itself with increasing levels of abstraction. It can also be used as a teaching tool because it is based on an analysis of responses, rather than of individual students.

A measuring instrument, also based on the SOLO taxonomy, was designed to gauge the levels of the students’ responses to these tasks. Each response was allocated a level that was subsequently coded as a number from zero to seven. Because the responses were in distinct ordered categories, it was possible to analyse the scores using the Rasch Model (Rasch 1960/80) for polytomous responses, placing both the difficulty of the tasks and the ability of the students on an equal interval scale. The Rasch Model was also used to evaluate the measuring instrument itself. Some adjustments were made to the instrument in the light of this analysis. It was found that it is possible to
construct an instrument to distinguish between levels of students’ written responses for each of the chosen table interpretation tasks.

The workshop was evaluated through a comparison of the levels achieved on the scale by individual students before and after the workshop. T-tests for dependent samples indicated a significant improvement (p < 0.01) in student performance for the majority of students. In addition, the positive change in overall means for the table interpretations tasks indicates that the workshop using the Five Step Framework was effective overall. The Five Step Framework and the measuring instrument are appropriate for teaching and learning at a range of educational levels, and they will be used to contribute to further research in the area.
LIST OF TABLES

Table 2.1 Perspectives of Numeracy
Table 3.1 Health Risk Factors related to socio-economic status
Table 3.2 Steps to Interpret a Table of Data
Table 3.3 Homeworkers
Table 3.4 Environmental Concerns
Table 3.5 Methods of Suicide
Table 3.6 The Subtasks for the Table Interpretation Tasks
Table 3.7 Criteria for Allocation to Initial Levels
Table 3.8 The Allocation of Tasks
Table 3.9 The Distribution of Tasks
Table 4.1 Individual Item Fit for the Initial Analysis
Table 4.2 Uncentralised Thresholds for the Initial Analysis
Table 4.3 Rescoring for Environmental Concerns
Table 4.4 Rescoring for Suicide 2
Table 4.5 Rescoring for Homeworkers 1
Table 4.6 Rescoring for Homeworkers 2
Table 4.7 Rescoring for Suicide 1
Table 4.8 Rescoring for Environmental Concerns 2
Table 4.9 The Overall Combination of Categories
Table 4.10 Individual Item Fit for the Final Analysis
Table 4.11 The Uncentralised Thresholds for the Final Analysis
Table 5.1 Criteria for Allocation to Initial Levels
Table 5.2 Final Categories for Environmental Concerns 1, Suicide 2, & Homeworkers 1
Table 5.3  Final Categories for *Homeworkers 2*
Table 5.4  Final Categories for *Suicide 1*
Table 5.5  Final Categories for *Environmental Concerns 2*
Table 5.6  The Comparison of Pairs of Tasks
Table 6.1  Categories for *Four-Category Analysis*
Table 6.2  Individual Item Fit for the *Four-Category Analysis*
Table 7.1  The Proportion of Students Who Improved After The Workshop
Table 7.2  Results for Dependent t-tests for *Final* and *Four-Category Analyses*
Table 8.1  Features of the Tables for *Environmental Concerns, Homeworkers & Suicide*
Table 8.2  Tasks Ordered by Location of Multistructural/Relational Thresholds
LIST OF FIGURES

Figure 4.1  ICC for the Initial Analysis of Environmental Concerns 1
Figure 4.2  ICC for the Initial Analysis of Homeworkers 2
Figure 4.3  CCC for the Initial Analysis of Environmental Concerns 1
Figure 4.4  CCC for the Initial Analysis of Homeworkers 2
Figure 4.5  Person-Item Location for the Coded SOLO Levels
Figure 4.6  Threshold map for the Initial Analysis
Figure 4.7  Person-Item Location for the Coded SOLO Levels
Figure 4.8  Threshold Map for SOLO Levels for the Initial Analysis
Figure 4.9  CCC for the Initial Analysis of Environmental Concerns 1
Figure 4.10  ICC for the Initial Analysis of Environmental Concerns 1
Figure 4.11  CCC for the Final Analysis of Environmental Concerns 1
Figure 4.12  ICC for the Final Analysis of Environmental Concerns
Figure 4.13  CCC for the Initial Analysis of Suicide 2
Figure 4.14  ICC for the Initial Analysis of Suicide 2
Figure 4.15  CCC for the Final Analysis of Suicide 2
Figure 4.16  ICC for the Final Analysis of Suicide 2
Figure 4.17  CCC for the Initial Analysis of Homeworkers 1
Figure 4.18  ICC for the Initial Analysis of Homeworkers 1
Figure 4.19  CCC for the Final Analysis of Homeworkers 1
Figure 4.20  ICC for the Final Analysis of Homeworkers 1
Figure 4.21  CCC for the Initial Analysis of Homeworkers 2
Figure 4.22  ICC for the Initial Analysis of Homeworkers 2
Figure 4.23  CCC for the Final Analysis of Homeworkers 2
Figure 4.24  ICC for the Final Analysis of Homeworkers 2
Figure 4.26  ICC for the *Initial Analysis* of *Suicide* 1
Figure 4.27  CCC for the *Final Analysis* of *Suicide* 1
Figure 4.28  ICC for the *Final Analysis* of *Suicide* 1
Figure 4.29  CCC for the *Initial Analysis* of *Environmental Concerns* 2
Figure 4.30  ICC for the *Initial Analysis* of *Environmental Concerns* 2
Figure 4.31  CCC for the *Final Analysis* of *Environmental Concerns* 2
Figure 4.32  ICC for the *Final Analysis* of *Environmental Concerns* 2
Figure 4.33  Threshold Map for the SOLO levels for the *Final Analysis*
Figure 4.34  Person-Item fit for the *Final Analysis*
Figure 6.1  CCC for the *Four-Category Analysis* for *Environmental Concerns* 1
Figure 6.2  ICC for the *Four-Category Analysis* for *Environmental Concerns* 1
Figure 6.3  CCC for the *Four-Category Analysis* for *Environmental Concerns* 2
Figure 6.4  ICC for the *Four-Category Analysis* for *Environmental Concerns* 2
Figure 6.5  CCC for the *Four-Category Analysis* for *Suicide* 1
Figure 6.6  ICC for the *Four-Category Analysis* for *Suicide* 1
Figure 6.7  CCC for the *Four-Category Analysis* for *Suicide* 2
Figure 6.8  ICC for the *Four-Category Analysis* for *Suicide* 2
Figure 6.9  CCC for the *Four-Category Analysis* for *Homeworkers* 1
Figure 6.10 ICC for the *Four-Category Analysis* for *Homeworkers* 1
Figure 6.11 CCC for the *Four-Category Analysis* for *Homeworkers* 2
Figure 6.12 ICC for the *Four-Category Analysis* for *Homeworkers* 2
Figure 7.1  Scatter Graph for the *Final Analysis*
Figure 7.2  Scatter Graph for the *Four-Category Analysis*
Figure 7.3  Threshold Map for the *Final Analysis*
Figure 7.4  Threshold Map for the *Four-Category Analysis*
ACKNOWLEDGEMENTS

For the ideas, encouragement, and guidance given to me throughout the long process of researching and writing this dissertation, I am very grateful to Sue Willis, David Lake and David Andrich.

I sincerely appreciate the support, encouragement and comments on the draft by Pat Forster and Barry Kissane.

I thank Marie Arandiga for her careful editing.

Thank you especially to my daughters who have been very patient with me over the time-consuming project and to my friends who have encouraged me to continue with it.
CHAPTER 1  INTRODUCTION

1.1 Overview of the Dissertation

Firstly, this thesis describes the numeracy demands of people in their everyday life, community life, learning, work and recreation. Secondly, it presents some national and international interpretations of the term numeracy and associated terms such as quantitative literacy, mathematical literacy and statistical literacy, all of which involve the application of some mathematics in a particular context. Thirdly the thesis examines ways that numeracy can be developed in relation to the learning of mathematics inside and outside of the mathematics learning area.

Fourthly, the thesis reports on an empirical study conducted with first year undergraduate students at a Western Australian university. The applied research study was developed as a result of the researcher’s experience in teaching undergraduate students that led to an awareness of the tendency of most students to skim over quantitative material that they encountered in their studies.

The aims of the study were to:

1. To develop a measuring instrument to evaluate students’ responses to table interpretation tasks, and

2. To investigate whether students’ ability to interpret tables of data can be improved through an intervention workshop using a Five Step Framework.
The thesis reports on the design of the workshop and the analysis of the data from the study including the refinement of the measuring instrument and an evaluation of the effectiveness of the workshop.

Chapter 1 gives an introduction to the concept of numeracy and describes some of the issues of concern for the development of numeracy.

Chapter 2 provides a review of the research and professional literature concerning numeracy. It examines the concept of numeracy using national and international descriptions and definitions, and considers the nature of numeracy from a range of perspectives. Ways of developing and measuring numeracy are reviewed, incorporating an examination of the relationship between numeracy and mathematics and the issue of transfer.

Chapter 3 describes the design of the study. It includes the aims of the study, the students who participated in the study, the development of the one-hour intervention workshop and measuring instrument based on the SOLO taxonomy, and the tasks completed by the students before and after the workshop.

Chapter 4 illustrates the use of the Rasch Extended Logistic Model for the analysis of qualitative responses and describes in detail the quantitative analysis of the measuring instrument, including the adjustment of the initial categories to improve the instrument.
Chapter 5 gives a qualitative analysis of the criteria of the measuring instrument developed in Chapter 4.

Chapter 6 explores the feasibility of having a measuring instrument based on only 4 categories.

Chapter 7 is concerned with the changes in the levels of students’ responses to the table interpretation tasks. Scatter graphs show the changes visually and statistical differences are reported giving the results of dependent t-tests. The effectiveness of the workshop is considered in the light of these results.

Chapter 8 includes discussion about the contributions of a number of factors to the levels of students’ responses.

Chapter 9 summarises the findings of the study and considers the implications for education and further research in the development of numeracy.

1.2 Introduction to Chapter 1

This chapter introduces numeracy as the field of research for this thesis. It includes an introduction to the concept of numeracy illustrating its importance to people in a wide range of situations and the ways in which it is being developed in educational settings. It identifies an area of concern where undergraduate students lack the necessary skills to engage with quantitative information that is typically encountered in tertiary education, including the reading and critical interpretation of tables of data.
1.3 Numeracy

The term *numeracy* is used in a variety of ways in the literature. These range from defining numeracy as a set of basic mathematical skills through to a rich description of the use of mathematics in a whole range of different contexts. Some of these definitions will be described and discussed in Chapter 2. The multiplicity of descriptions of numeracy may be accounted for by the fact that the importance of being numerate, and what that entails, has changed considerably over the last four decades and continues to change (Steen, 1990). The improvement in computers and other technologies means that the speed and ease of communication has increased, and often that information includes numbers. Numeracy now extends to dealing with quantitative information in electronic form. In this fast moving world there are pitfalls for the unwary and, as Steen (1997) points out:

> As information becomes ever more quantitative and society relies increasingly on computers and the data they produce, an innumerate citizen is as vulnerable as the illiterate peasant of Gutenberg’s time (p. xv).

In the USA, 50 percent of those interviewed as part of the large scale National Adult Literacy Survey had major difficulty with situations in life which involved numbers, quantities, mathematical concepts and algorithmic processes (Gal, 1993). Paulos (1988) found that people could not deal with big numbers or probabilities, both of which contribute to confusion about political polling, medical consent and safety regulations. He found that when
doctors estimate the risks of various operations and procedures they were
“way off the mark, often by several orders of magnitude” (p. 8).

The way that probabilities are reported can be difficult to interpret, for
professionals as well as lay people. Health benefits from surgery or
medication are sometimes reported in terms of absolute risk reduction, and
sometimes in relative terms. This is likely to confuse people about the degree
of a treatment’s benefit. For example calculation in terms of the relative risk
can look more beneficial (Gigerenzer, 2002).

To investigate expert responses to different ways of describing risk some
British researchers analysed the responses from a team of expert regional
health personnel in the UK to proposals for funding for four cardiac
rehabilitation and four cancer-screening programs. These proposals were
couched in terms of the stated benefits of the programs, which were in fact of
equal benefit, but the proposals were framed in four ways. For each
programme one proposal was framed in terms of relative risk reduction, the
second was in terms of absolute risk reduction, the third was based on the
number of people who needed to be treated to prevent one death and the 4th
gave the proportion of surviving patients. The team selected the ones
expressed as a relative risk, with only 3 out of the 140 experts realising that
the four proposals reflected the same clinical results (Gigerenzer, 2002).

In Australia, Willis (1996) also has concerns about numeracy and particular
ramifications for certain groups:
Improving numeracy outcomes is particularly important for those who are economically and educationally disadvantaged since numeracy skills are socially distributed, that is, people who are regarded as insufficiently numerate are predominantly working class, are disproportionately female and more likely to be members of certain ethnic and racial minority groups. Perceptions about the numeracy demands of particular tasks and, consequently, what sorts of people succeed with them will have a powerful effect on who gets chosen for, and who chooses, themselves into, certain of life’s opportunities at home, at work and in public life (p. 5).

This emphasises the importance of numeracy for all, and the need to develop the appropriate skills from an early age. Numeracy needs to be high on the educational agenda to ensure that certain groups are not disadvantaged.

Almost everyone encounters material in printed, oral and electronic forms inside and outside educational settings. Within these there are verbal, tabular, graphical and symbolic representations of the information to be conveyed. Verbal representation is written or spoken text and may include mathematical terms such as birth rates, decreasing inflation, exponential growth and so on. Tabular representation includes information in tables ranging from more straightforward, as with bus timetables, to more complex tables of data found in statistical reports. They include numerical data related to different categories and are often longitudinal. Graphical representations include bar graphs, line graphs or pie graphs frequently found in newspapers where they can easily mislead the readers unless they pay attention to the legends, axes and scales. Symbolic representation is typically restricted to the science and mathematics areas and is found in the form of formulae. It has been suggested that the ability to make transformations between the different representations, so as to make the
information more accessible and understandable, would contribute greatly to numerate behaviour (Chapman, Kemp, & Kissane, 1990; Kemp & Kissane, 1990).

The term **numeracy** is not the only one used in the context of understanding quantitative information. Others, described in Chapter 2, are terms such as quantitative literacy, statistical literacy, mathematical literacy, critical numeracy, document literacy and civic literacy. In this dissertation, numeracy is taken to mean - making the decision to use appropriate mathematics in situations where mathematics is useful, to apply the mathematics and then to critically evaluate the results. This definition is a synthesis of other definitions and particularly influenced by the definition of the Open University (1980). The level of sophistication of the mathematics used and the degree of critical thought depends on the mathematical experience of the person involved. The term **critical numeracy** is used to encompass all the characteristics of numeracy with an emphasis on the critical nature of it.

### 1.4 Numeracy Demands

When there is discussion about numeracy, reference is often made to its relevance to ‘everyday lives’. There is a wide range of situations in which people need to meet numeracy demands. The activities in which people meet numeracy demands varies but as Gal (2004) points out there are many contexts within these categories where people need to be “effective data
consumers” (p.50) in a variety of ways. This section provides descriptions of some of the many numeracy demands in a range of contexts. To illustrate these situations examples are given for five categories: everyday life, community life, learning, work and recreation. The empirical study reported on in this thesis is concerned with understanding and interpreting data presented in tabular form that people encounter in many of these situations.

1.4.1 Everyday Life

The concept of what constitutes everyday life varies depending on people’s situations that can change rapidly. These usually include practical aspects that are often monetary and based on economic considerations. In the financial field, children move from simple to more complex situations involving the handling of money. Budgeting becomes possible when they have some jurisdiction over their own finances through the earning of wages or pocket money. There are school excursions and fairs, school camps, and family holidays and other opportunities where planning of money, time and space involves numeracy.

Young people are influenced by the advertising they encounter on television, on the radio, and in magazines. They become aware of what is fashionable in clothes, games, toys, food, music and technology, as they become part of the consumer society. As they get older the numeracy demands increase to include budgeting for rent, food and phone calls as well as other kinds of activities such as playing musical instruments, sailing and dressmaking. Many teenagers work part-time in a variety of jobs.
However it is not only young people who are influenced by advertising. Older people are also likely to be influenced by advertising on a variety of media including the Internet. Every day there are claims that certain products will improve one’s health and fitness, make one more desirable, transform one’s home and so on. Increasing use and availability of credit cards makes spending easier and the need to be realistic about the repayments even more important.

There are other areas where numeracy is very important when making financial decisions. There are different kinds of mortgages available to suit different lifestyles and incomes. For all of them payments can be adjusted, or bulk payments can be made which can make a large difference to the long-term cost. Many people now have home computers with spreadsheets that are “more powerful than programs available to professional accountants twenty years ago” (Steen, 1997, p, xvi). This gives people the opportunity to use these spreadsheets to work out how the decisions they make, in situations such as car loans or salary sacrificing, will impact on their economic lives.

In household budgeting, understanding bank statements and fees, paying bills for power, water and rates, running a car or motorbike, filling in application forms, understanding prescription labels, buying appliances, credit card use and so on all place numeracy demands on the individual. As they move towards retirement people need to calculate how paid work and asset values affect allowances and pensions. Making renovations and
building new houses involve considerations of relative merits of building materials and appropriate levels of energy use. Interior decoration, furniture making, dressmaking, shopping and cooking, all involve both financial and spatial concepts (AAMT/DETYA, 1997).

A serious issue in the area of health is concerned with tests for such problems as high blood pressure or cholesterol and screening tests for cancer, AIDS and so on. There is a difference in purpose behind these tests but in essence they are both not ‘error-free’. It is of grave concern that people are led to believe that the results of these tests are conclusive. Gigerenzer (2002) describes the situation, by no means unique, of a single mother who was told that she had tested positive for HIV. This led to her losing her job, her friends and having difficulties in caring for her son. She moved into a refuge for people who had tested for the HIV virus. She contracted a severe case of bronchitis and her blood was tested again. This time it was negative for HIV. In fact her initial blood sample was found to have been inadvertently switched with another person’s, but even the test on her own blood would not have been conclusive due to the errors associated with the tests.

The message of this scenario is that there is uncertainty in test results and that the doctor should have realised the need to tell her to take another test. The same is true for mammograms, where only about 1 in 10 women who test positive actually have breast cancer (Gigerenzer, 2002). The doctors and patients need to be quite clear that there is uncertainty in the results and to seek further advice where appropriate.
1.4.2 Community Life

Community life includes political, economic and social aspects of national and local life, frequently influenced by the media. People not only operate in their own everyday life but also with others as part of their membership of communities. Community life operates at many levels from the immediate neighbourhood community through to national and international communities. Mass communication enables most people to be aware of what is happening in the rest of the world within a very short time (Steen, 1997). Attitudes and behaviour are influenced by the media in a large range of issues from local ones to those involved in major national policies. The way in which quantitative information is presented can make a big difference to views and opinions.

At the local level there is work in the community, often unpaid, where decisions are made, policies developed and various activities undertaken. These include participating in school parent committees where decisions are made about school policies, canteen arrangements and fund raising; running a scout or guide group; petitioning for a skateboard park or running a neighbourhood watch program, all of which involve some financial calculations and decisions. In the wider community people are asked to endorse and support decisions made at state or federal levels concerning immigration, welfare payments, health benefits, logging, military involvement and so on (Hogan & Kemp, 2000).
The media is a crucial part of modern community life. Interpreting media reports makes many numeracy demands and raises many issues. There are cases where there is inadequate information containing incomplete data or logical deficiencies where the effects are linked to causes without justification. There is misleading information that is included to deliberately persuade people to adopt a particular point of view or information that is presented in complex ways that could encourage erroneous conclusions. There are cases where the original collection of data was neither valid nor reliable and the conclusions drawn are very questionable. People need to be numerate to avoid consequences of being misled or misinformed by such practices (Gelman et al, 1998; Gal, 2002; Watson, 1995).

Kolata (1997) who writes for the New York Times and who has been a science writer in biology and mathematics for many years, emphasises how important it is for people to realise how misleading stories in the newspapers can be. She stresses that readers need to be particularly aware of the difference between anecdotal evidence and evidence from randomised controlled clinical trials which has included numerical analysis of the data. Anecdotal evidence about silicon breast implant symptoms, and relief after removal of the implants, only suggest a causal relationship which is not subsequently backed up by quantitative data. Similarly, apparent reduction in the overall sperm counts of donors might be related to the criteria for selection of donors rather than an absolute decline. In addition she notes the importance of not being swayed by stories: we need to be aware that a one
hundred percent increase in AIDS in women does not necessarily mean an epidemic; one needs to look at the raw data. A one hundred percent increase could be from five to ten, or one million to two million. She reinforces the need to not only look at the numbers, but also to be able to think about them in the context of the situation.

Every high school graduate should be able to read and understand charts and graphs but that is only the beginning, like saying they should know how to write a sentence. They also should understand logical arguments and logical fallacies. They should understand the nature of evidence. They should understand the difference between absolute and relative risk (p. 28).

Information on crime is another issue that impinges on community life and influences the ways that people act. Everyday there are reports in the newspapers and on the television about crimes that have been committed. Reading and understanding statistical information which is produced and disseminated is not an easy task for most people (Gal, 1998, 2003). Crime statistics are often produced in tables, which include numbers and percentages, and in surveys people can respond to more than one category, so that the way the data are reported can be confusing.

For people responsible for making decisions and recommendations connected to crime investigations an understanding of statistics and probability is of particular importance. For example, Gigerenzer (2002) points out a potential flaw in a chain of inference in deciding whether an accused is guilty. It is commonly believed that if a defendant’s DNA matches the DNA found at the crime scene then this would prove that the
defendant is the source of the trace. Gigerenzer gives three reasons why this is not necessarily the case. Firstly, there may be just a mistake at the laboratory level. Secondly, even rare DNA patterns can occur in more than one person. Thirdly, the blood sample could have been transferred to the scene, intentionally or otherwise. Gigerenzer (2002) gives the example of the O. J. Simpson case, in which Simpson was accused of murdering his wife and her friend, and the “defense forcefully alleged that some of the blood evidence at the scene was introduced by the police” (p. 165).

The Chamberlain case was opened in Australia in 1981, where Mrs Chamberlain was accused of killing her baby daughter Azaria. Mrs Chamberlain was tried and was sent to prison. The case used results from screening tests to ascertain the presence of blood in the family car. The misunderstanding about the nature of screening tests lead to what has now been recognised as a miscarriage of justice. Contrary to earlier popular belief, the errors related to the probabilities concerned with the kind of blood on the child’s clothing were made by judges, lawyers and scientists and not by the jury.

The conditional probabilities underpinning the analysis of many scientific tests are not intuitive and there could be a strong case for more inclusion of learning about chance and data in the school curriculum. As the trend is for the courts of law to present their statistical arguments in terms of probabilities rather than frequencies it is harder for people to find the flaws in the arguments.
There are examples reported of published errors in the press made by the writers of the reports or advertisements. In a newspaper article from the Netherlands 1 in 25 was mistaken for 25% rather than 4% resulting in the mathematically incorrect conclusion that over three million people in the Netherlands were illiterate (Treffers, 1991). A corporate advertisement in a newspaper in the USA advertised a 200% saving on electricity when using a certain metal halide fixture, and when management were questioned about it they claimed they had used a formula so it must be correct (Dewdney, 1993). In the UK an article blamed frequent fish eating incorrectly for aggression in men. Watson (1995) investigated the source of the article and found that the author had incorrectly interpreted the probabilities and had subsequently published an erroneous conclusion. These examples indicate a need for readers to question apparently unrealistic statements.

The general public in most places in the world receives information on a regular basis related to issues such as road toll statistics, unemployment, crime, drug use, health risks and possible cures, hospitals, air pollution, drinking water pollution, ocean and river pollution, dangerous waste disposal, forest fires and the national budget. The authors of articles on these issues typically have a view that they want to present to the readers. It may be related to potential unemployment versus ecological damage, the effects on society due to the legalising of drugs such as marijuana, or the relative merits of a new speed limit.
To be able to assess the merits of these kinds of articles requires that readers understand at least the rudiments of sampling and data collection and have some notions of validity and reliability. But possibly even more important is that they need to understand that data collection is not neutral. Television and newspaper polls can be conducted in ways likely to bias outcomes since people who respond are likely to over-represent those who feel strongly about the issue.

1.4.3 Learning

Learning starts early in life and continues for many years. Learning can be thought about in a variety of contexts: primary, secondary and tertiary educational institutions, learning at home, on the sports ground, in the library, travelling the world and on the Internet. Learning involves a whole range of skills and dispositions, of cultural customs and histories.

In primary school students encounter many situations in their classrooms and school events where numeracy demands are made. They might be reading a story and encounter phrases such as ‘twice as many’ and need to interpret what that means (Hogan & Kemp, 2000). In learning about a country and constructing a flag they would have to use a scale and measure appropriately. An article about the destruction of old growth forests could require students to understand percentage and area. Students might have to work out a fair way to set up a roster for sharing sports equipment or staffing their class stall at the Christmas fair.
When students attend secondary school the numeracy demands typically become more sophisticated. Many demands are school based, but many others are encountered outside the classroom. More understanding of data collection, sampling and graphing is needed to critically interpret information presented in the media. There are many numeracy demands across the curriculum including time zones and climate variation shown with graphs, economic issues about overseas and domestic expenditure, implications of genetic engineering in science, issues to do with population growth and decline and many more.

Once students have left secondary school in Australia, and if they choose to continue with formal learning, there are opportunities for tertiary study in Technical and Further Education (TAFE) institutions, business colleges and universities. The students encounter material in a variety of ways such as lectures, tutorials, seminars, oral presentations, videos, computers, scientific and graphics calculators, Power Point presentations and Internet interfaces. The different subject areas include a variety of numeracy demands, in broadly similar ways to those studied at school. In addition students studying at university have to be aware that a higher level of critical analysis is required than at school. “Students are expected to use the necessary skills, including numeracy skills, at a high level, to think critically about the material presented and to be able to present convincing and logical arguments of their own” (Kemp, 1995, p. 377).
Many students at the tertiary level, particularly in non-mathematical disciplines, lack the confidence and the necessary skills to engage with the quantitative material in texts in the form of tables, graphs, charts, or numbers. Often these students have not followed a mathematics course when the taking of subjects was dependent on the students’ choice. In addition, they can be unaware of the importance of dealing with this kind of information themselves (Chapman, 1988; Kemp, 1995).

This lack of engagement is of concern for two main reasons. Firstly, in the context of their study, students need to understand and interpret quantitative material in journals, textbooks, presented in lectures, in documents on the World Wide Web and in many other places, to become informed about their area of interest. Secondly, since authors’ own perspectives influence how they interpret and present information, leading to potential bias, students need to be able to draw their own conclusions from any available data.

1.4.4 Work

During the first half of the twentieth century it was relatively easy to predict the kinds of mathematical skills that would be needed in the various categories of employment. People generally entered the workforce and stayed in the same kind of work for forty years. This practice has changed and people change careers several times in their lifetime (Galbraith et al., 1992; Orrill, 1997). The careers that will be available for the next generation
have often not been imagined by the previous one. Indeed, as Orrill (1997), p. xii) points out “now change is more the norm than the exception” (p. xii).

The changed nature of the workplace has implications for both the literacy and numeracy needs of the twenty first century. Galbraith et al (1992) investigated what experts in the field perceived as the present and future numeracy demands of adults and reported their findings in terms of three scenarios described respectively as “careful, futuristic and process” (p. 582). All three scenarios referred to the importance of an understanding of statistics and diagrams, graphs and tables. They all acknowledge the move towards more calculator and computer use, and with this, the need for selection of appropriate problem solving techniques and applications. It appears that the numeracy that will be needed in the workplace will be a combination of both specific and higher order generic skills. For instance, specific skills include making up correct dosages for nurses or measuring distances and angles for surveyors, while higher order generic skills include knowing how to approach problem solving, critical thinking and analysis using numerical information.

In the Australian workplace today there are many aspects that require both employers and employees to be numerate. There are decisions to be made using numerical data in enterprise bargaining agreements, workplace agreements, superannuation, taxation, health insurance, union membership, safety regulations and more. These involve choices based on a real understanding of these issues.
The extent to which people at work have to make decisions about where and when to use mathematics and the kinds of mathematics to use varies for different situations. There is evidence to show that people who need to use specific job-related mathematics employ strategies and techniques given to them by their fellow workers (Cockcroft, 1982). These methods bear almost no connection to those learned in school. In general, routine methods work well but there are also non-routine situations that may arise that are not coped with well, and where mathematics learned in school could have informed appropriate responses. A good example is the Space Shuttle Challenger disaster in January 1986, where three seconds into the mission the Challenger exploded, killing the entire crew. There could have been a different outcome had the managers had a better understanding of statistics. The numeracy skills that students develop at school should include the ability to be critical about data, which would enable them to tackle non-routine situations with confidence, and anticipate potential errors.

When people are out of work or not in full time work, the demands can still be complex. For example, the Social Services Department in Australia demands information of a financial nature in order to assess the appropriate benefits. When undertaking further study, students claiming allowances must be careful not to exceed allowed earnings. Student loans need to be managed carefully as budgeting is particularly important when the payment is in the form of a lump sum. When people are not in work they need to be
able to cope with the demands of claiming benefits, pensions or other allowances.

1.4.5 Recreation

In today’s society, although there are people who work many hours per week, there are still many opportunities for people to follow their leisure interests. In Australia many people support and participate in a wide range of sporting activities. Many people are employed in the sports and leisure industry, while many others are amateur participants and interested spectators.

In the 2001 census it was found that 83 008 people (1.0% of persons employed in Australia) had their main job in a sport and physical recreation occupation. In the 12 months to April 2001, an estimated 4.1 million persons (27.1% of all persons 15 years and over) were involved in sport and physical activity organised by a club, association or other organisation. This involvement includes players or participants and those in non-playing roles that support, arrange and/or run organised sport and physical activity. The 2002 survey of the Australian Bureau of Statistics reported that over 9 million persons participated in some sport in the previous 12 months. The participation varied according to age and available facilities, with the most popular activities were walking, swimming, golf, aerobic fitness, tennis, soccer and cricket (Australian Bureau of Statistics, 2004).
For interested people in much of the world there are commentaries on
sporting activities on television and radio and reports in newspapers with a
high proportion including sporting statistics. There are many people who
attend sporting events. For example there were 2.5 million attendances at
Australian Rules football events in 2002 (Australian Bureau of Statistics,
2004). Followers of AFL football find it useful to have some understanding of
the meaning of a variety of performance measures including goals, marks
and points and how these affect the likelihood of winning, related to the
success in the first quarter. In England and former English colonies cricket
fans will interpret graphs about run rates, strike rates and batting and
bowling averages. There were about 800 thousand attendances at Australian
cricket games in 2002. In the American game of baseball there are earned run
averages, runs batted in and batting averages. Scores on a golf course include
negative and positive numbers; swimming and athletics involve times and
averages; netball and basketball involve goals and baskets, and so on.

Many people, especially in Australia, like to be on or at least near water
during their leisure time. There are motorboats that require knowledge
concerning engines and fuel, skis, mooring fees, and sailing boats with their
own set of requirements, including sails of different sizes appropriate for
different wind speeds. For all boats there are safety requirements that must
be met and people need to be able to interpret and adhere to the
mathematical constraints. People like to fish and need to consult tide tables
to know when it is the best time to fish. They also need to consult guidelines
to make sure the contents of their catch are of legal size. Others like to scuba
dive and must understand and follow the safety guidelines concerning
diving depths and the oxygen-carbon dioxide ratios in their blood.

There are other interests and hobbies including doing number and word
puzzles and playing games such as chess, bridge, monopoly and board and
computer games, all of which require an awareness of strategies and some
mathematical skills. Some people take up pottery or picture framing classes,
which include measuring and spatial concepts.

There are people who like to gamble in casinos, at the races, and on the poker
machines in public houses and cafés, and to buy lottery tickets. In Australia,
in June 1998, there were 1 776 businesses in the gambling services industries
employing 37 035 people and receiving $7 935m in income (Australian
Bureau of Statistics, 2000). In America the gambling in Las Vegas is well
known and the lottery is very popular. In the New York State Lotto 6 out of
54 numbers are needed to win the top prize. The probability of winning this
prize is about 1 in 26 million. The increase in the popularity of the lottery
came about in 1971 when New Jersey introduced a computer-controlled
system that allowed people to pick their own numbers. Because the method
of drawing winners was completely random, using a barrel with balls
selected at random as in Australia, the ability to choose numbers had no
effect on the odds of winning but it made the game more popular. This was
because of a bias known as the illusion of control where people believe they
have greater influence on the outcome. This also applies to the games of chance in the casinos where people can roll their own dice (Vyse, 1997).

Even though the selection of the winners is purely random, there is a substantial industry in marketing books and magazines promoting lottery and betting systems. There are books sold on numerology with an emphasis on lucky numbers. Some newspapers publish the number of times a particular number has been drawn and some agencies show lucky numbers perpetuating the myth that one number can be more likely than another number to be picked at random. More awareness of the odds might prevent excessive gambling for those who cannot afford it.

In addition, numbers appear to have a fascination for some with many people interested in numerology and astrology. Steen (1990) points out that “many adults trust astrology more than astronomy, numerology more than mathematics, and creationism more than molecular biology” (p. 20). Both Paulos (1988) and Gardner (2000) have noted the links between innumeracy and numerology and scientific illiteracy and pseudo-science. Others (Schick & Vaughn, 1999; Shermer, 1997; Vyse, 1997) are concerned with the prevalence of superstition and belief in paranormal activities and warn of the potential dangers of people being persuaded by information with components of numeracy that is presented without scientific evidence. There are countless examples connected with chance and gambling, where people do not understand the mathematics. One of Vyse’s concerns is the promotion
of alternative therapies based on anecdotes and testimonials rather than through articles published in respectable journals.

Overseas travel has become more prevalent for a variety of reasons. Some are travelling on business but many are travelling on overseas holidays, and to visit family and friends. People need knowledge of currency exchange rates, transport timetables, international time and date lines, and time management skills.

1.5 Summary

The purpose of the preceding sections was to make clear how pervasive and important numeracy is in all people’s lives. They included descriptions of numeracy in everyday life, community life, learning, work and recreation.

The development of numeracy starts when children are young as they develop an understanding of the relationship between mathematics and their world. In many ways, numeracy is similar to literacy in that it transcends subject boundaries and is fundamental to many aspects of people’s lives. It follows that the development of both literacy and numeracy be integral to the curriculum in educational settings at all levels from primary to tertiary, and that all teachers should see themselves as ‘teachers of numeracy’.

There are a variety of viewpoints and practical issues associated with the development of numeracy. These involve people’s views of what numeracy is and how it might be developed. Some people refer to numeracy as basic
mathematical skills and focus on their development at both the school and adult levels. Others look at the connections and relationships between mathematics, other learning areas and everyday lives. The numeracy demands that people encounter in their lives occur in a wide range of situations and it is important that they can make the connections and use the appropriate mathematics to cope with them well. There is a discussion of these issues in Chapter 2.
CHAPTER 2    LITERATURE REVIEW

2.1 Introduction

Chapter 1 gave some indication of the extensive quantitative demands made on people in their everyday life, community life, learning, work and recreation activities. In the literature descriptions of abilities, dispositions or literacies that people need to have to cope with these demands are given in different terms depending on the specific nature of the tasks. These terms include numeracy, critical numeracy, mathematical literacy, quantitative literacy, document literacy, statistical literacy, critical mathematics, civic literacy, scientific literacy and so on. The literature also shows that definitions and interpretations of these terms vary quite considerably nationally and internationally.

The main purpose of this chapter is to review the literature related to different viewpoints about the nature of numeracy and the ways in which it can be developed and evaluated. For completeness the other terms listed above are defined and some reference made to similarities between them. However, the inclusion of these terms is intended to illustrate the complexity of the area and to emphasise that the common denominator is the use of mathematics in a variety of contexts and to varying extents. It is not intended, in this chapter, to provide a detailed analysis of the differences between the various terms. However this is not to deny the importance of the use of the terms and the concepts underpinning them. The emphasis on the
term numeracy in this dissertation is because in Australia this term is becoming accepted by academics in the mathematics education field to encompass many of the qualities and characteristics of the other terms.

2.2 The Concept of Numeracy

There is a range of interpretations of the meaning of the term *numeracy*. At one end of the scale is a rich interpretation of the term *numeracy* as the ability to recognise when mathematics can be useful, in familiar and unfamiliar contexts, to apply the appropriate mathematics and to interpret the results. At the other extreme *numeracy* is regarded as only the basic mathematical or indeed arithmetical skills usually learned in primary school. There are variations in between these extremes. Some viewpoints include reference to the social and cultural nature of numeracy in a range of contexts.

To give some idea of the range of views that surrounds the term *numeracy*, a selection of views from prominent educators and policy makers from the UK, USA and Australia is included in this section. The selection has been made to give examples of the kinds of thinking about the concept of numeracy and is by no means exhaustive. The order is chosen to present similar ones together starting with views of numeracy encompassing higher order thinking and moving on to those where it is regarded mainly in terms of basic mathematical skills.

The writers of the Crowther Report first coined the term in 1959 in the UK:
Numeracy is defined as a word to represent the mirror image of literacy ... On the one hand ... an understanding of the scientific approach to the study of phenomena—observation, hypothesis, experimentation, verification. On the other hand ... the need in the modern world to think quantitatively, to realise how far our problems are problems of degree even when they appear to be problems of kind. Statistical ignorance and statistical fallacies are quite as widespread and quite as dangerous as the logical fallacies that come under the heading of illiteracy. (Quoted in Cockcroft, 1982, p. 11).

In referring to the danger of statistical ignorance and fallacies it is evident that the Crowther Report regarded numeracy as encompassing higher level thinking skills analogous to those required for literacy.

The first widely used definition was from the team at the Open University (UK) in 1990, which defined numeracy as:

When to use mathematics, what mathematics to use, how to do mathematics, and how to use the results provided by the mathematics (pp. 23-24),

which includes thinking about more than just the operations of mathematics.

To emphasise the importance of looking at the results and reflecting on the context and meaning critical numeracy is used by Lake (1999):

Meaningful analysis begins when the student recognises the relevant framework for the graph or table. The important pointers that establish the context of the data must be understood before the student can produce critically numerate responses. This is the first point of departure between a mathematical and a critical numeracy. In mathematics the substance of the data are held within an equation. In the biological sciences, the essence lies within the factors surrounding the measurement (p. 194).

A concern that numeracy was being largely thought of as basic mathematical skills in Australia brought about a joint initiative in 1997 of the Australian Association of Mathematics Teachers (AAMT) and the Education Department of Western Australia (EDWA), where issues and questions
concerned with numeracy including *what does it mean to be a numerate person?* and consequently *what would this mean for schools?* were raised. The conference report, *Numeracy = everyone’s business*, defined numeracy as:

> [u]sing mathematics to achieve some purpose in a particular context. To be numerate is to use mathematics effectively to meet the general demands of life at home, in paid work and for participation in community and civic life. In school education, numeracy is a fundamental component of learning, performance discourse and critique across all areas of the curriculum. It involves the disposition to use, in context, a combination of: underpinning mathematical concepts and skills from across the discipline (numerical, spatial, graphical, statistical and algebraic); mathematical thinking and strategies; general thinking skills and a grounded appreciation of context (AAMT/DEETYA 1997, p. 15).

Although this definition of numeracy was widely disseminated in Australia this view is not held by everyone as will be seen in Section 2.3.1.

From a similar perspective of looking at numeracy as more than basic mathematical skills are definitions from:

Steen (1990):

> Numeracy is to mathematics as literacy is to language. Each represents a distinctive means of communication that is indispensable to civilised life (p. 211),

and Gal (1993):

> numeracy as encompassing a broad set of skills, strategies, beliefs, and dispositions that people need to autonomously engage in, and effectively manage situations involving numbers, quantitative or quantifiable data or information based on quantitative data (p. 21).

Some maths educators emphasise the contextual nature of numeracy, as for example, Bishop (1992) who states that:
numeracy and mathematics are not the same thing. We also need to recognise that numeracy sits within a certain societal context and has been established through certain social situations. There is therefore a certain particularity about numeracy and my definition of numeracy is the mathematical knowledge needed by every citizen to empower them for life in that society (p. 148).

From the UK, Brown (2001) indicates a perspective that focuses more on the mathematical skills:

I should first make it clear that the meaning of numeracy in England is somewhat different from the meaning in Australia. Although most people involved in mathematics education would accept the Australian/New Zealand interpretation of numeracy as being about the ability to apply number skills and concepts to solve problems in the real world, our UK politicians are more concerned with the traditional ‘basic skills’ characterised by what they believe the public wants, i.e. knowing addition and subtraction number-bonds and multiplication tables, together with facility in both mental arithmetic and traditional written procedures (p. 1).

Similarly in Australia there are those who focus on numeracy as the ability to use basic mathematical skills. The Early Numeracy Research Project in Victoria is a professional development project concerned with developing the mathematical skills of young children in their first three years of school. The project developed a framework for key aspects of early numeracy learning and ways to assess the mathematical profiles of the students. The project focuses on developing mathematical skills in the strands of Number, Measurement and Space using growth points as guides. The growth points are all described in terms of mathematical achievement. The impetus for the project was a desire to improve mathematics learning (Clarke & Cheeseman, 2000)
There are other terms that are used and the most frequently used are described here. For example, *mathematical literacy* as used in PISA 2000 study which

...looked at mathematics in relation to its use in people’s lives and assessed the capacity of students to recognise and interpret mathematical problems; translate these problems into a mathematical context; use mathematical knowledge and procedures to solve problems; interpret the results in terms of the original problem; reflect on the methods applied; and formulate and communicate the outcomes. (OECD, 2003, p.8).

The term *quantitative literacy* appears to be used often, with slightly varying meanings, particularly in the USA. The definition from the report of the National Assessment of Adult Literacy (NAAL) Survey of the USA in 2003 is based on the work of Kirsch and Mosenthal:

The knowledge and skills required to perform quantitative tasks—that is to identify and perform computations, either alone or sequentially, using numbers embedded in printed materials. Examples include balancing a checkbook, computing a tip, completing an order form, or determining the amount of interest on a loan from an advertisement (White & Dillow 1995, p.4).

According to Dossey (1997):

[q]uantitative literacy may be defined as the ability to interpret and apply these aspects of mathematics’ ... (data representation, number and operational sense, measurement, variables and relations’ geometric shapes and spatial visualisation)... to fruitfully understand predict and control relevant factors in a variety of contexts (p. 174).

Forman (1997) emphasises the cross-curricula nature of quantitative literacy:

To mathematicians, quantitative literacy is not the same as mathematical literacy. It is part of all subjects, the responsibility of teachers in all disciplines - many of whom, however, may not be prepared for this new responsibility (p161). ... The general consensus is that quantitative literacy is a subset of mathematical literacy, the latter
Devlin (2000) relates quantitative literacy and numeracy in this way:

Roughly speaking, quantitative literacy - sometimes called numeracy - comprises a reasonable sense of number, including the ability to estimate orders of magnitude within a certain range, the ability to understand numerical data, the ability to read a chart or graph, and the ability to follow an argument based on numerical or statistical evidence (p. 24).

Thus it can be seen that there is a range of views concerning the extent of the mathematical nature of quantitative literacy from the computations at a kind of everyday level to those required for scientific reasoning. Some include aspects of spatial awareness and an understanding of graphs and tables. For Forman quantitative literacy is subset of mathematical literacy.

As well as evaluating levels of quantitative literacy the NAAL evaluated levels of document literacy which refers to:

The knowledge and skills needed to perform document tasks - that is, to search, comprehend, and use non-continuous texts in various formats. Document examples include job applications, payroll forms, transportation schedules, maps, tables and drug or food labels (p.4).

This has tasks that seem to overlap the quantitative literacy somewhat and certainly require the use of some mathematical skills.

So far the definitions have focussed mainly on mathematical and spatial demands. The importance of citizens being able to make informed decisions which is a part of civic literacy defined by Millner (2001) to be ‘the knowledge and capacity of citizens to make sense of their political world’ (p.1) involves
an ability to evaluate and interpret statistical information which is often referred to as statistical literacy.

Wallman (1993) defined this statistical literacy in her presidential address to the American Statistical Association:

‘Statistical Literacy’ is the ability to understand and critically evaluate statistical results that permeate our daily lives – coupled with the ability to appreciate the contributions that statistical thinking can make in public and private, professional and personal decisions. (p. 1).

Gal (2004) proposes a more encompassing understanding of statistical literacy that incorporate communication as well as interpretation of statistical data and appreciation of the contribution of statistics. He proposes that in this context the term statistical literacy:

...refers to broadly to two interrelated components, primarily (a) people’s ability to interpret and critically evaluate statistical information, data-related arguments, or stochastic phenomena, which they may encounter in diverse contexts, and when relevant (b) their ability to discuss or communicate their reactions to such statistical information, such as the understanding of the meaning of the information, their opinions about the implications of the information, or their concerns regarding the acceptability of their conclusions (p. 49).

Directly related to this concept of interpreting and understanding statistical information is critical mathematics described by Frankenstein (2001) as involving the understanding of mathematics, the understanding of the mathematics of political knowledge, the understanding of the politics of mathematical knowledge and the understanding of the politics of knowledge.
For Yasukawa, Johnston & Yates, (1995) in Australia their interpretation of critical mathematics emphasises the need to be aware of the ways in which mathematics is used to interpret the social world. They link critical mathematics to numeracy and the social responsibilities inherent in it.

“Critical mathematics” is sometimes used to describe an approach to teaching which recognises that mathematical knowledge needs to be constructed and owned by the learner. This use of the word “critical” is limited to subtasking mathematics within the domains of academic mathematics. “Critical” to us means not only this but also the awareness of how mathematics can be used either to reflect or to violate realities in the wider social world....Being numerate involves not only having this critical awareness, but also involves the responsibility of reflecting that critical awareness in one’s social practice....Unlike mathematics, numeracy cannot pretend to be objective and value free (p. 815).

These terms are all related to the interpretation of quantitative information which is a particularly important component of in a scientific literacy. The PISA surveys have measured scientific performance in the following context:

The emphasis on the PISA 2003 assessment of science is on the application of science knowledge and skills in real-life situations, as opposed to testing particular curricular components. Scientific literacy is defined as the capacity to use scientific knowledge, to identify questions and to draw evidence-based conclusions in order to understand and help make decisions about the natural world and the changes made to it through human activity (PISA, 2004b, p.282)

To emphasise that scientific literacy is not restricted to students but is an essential part of life for all citizens Murcia (2005) looks at the characteristics of the person:

A scientifically literate person will have a broad and practical understanding of science that contributes to their competence, interest and disposition to use science to meet the personal and social demands of their life at home, at work and in the community (p.40)
This section has listed some of terms that are used in relation to the application of mathematics in a range of contexts. It has not defined the relationships between the terms, like between numeracy and mathematical literacy or scientific literacy and numeracy for example because this is not the intent of this chapter. However there are evidently similarities and commonalities in the descriptions which all seem to emphasis the need to people to be able to chose and use mathematics, or statistics, in interpreting their world.

2.3 Perspectives of Numeracy

The definitions and descriptions of numeracy and associated terms highlighted in Section 2.2 had similarities in their reference to the use of mathematics in a variety of contexts. In this section the focus will be on the ways that numeracy is interpreted in practice and to illustrate these this section examines some of the kinds of numeracy projects and learning activities being undertaken in schools.

Perusal of some of these projects aimed at developing numeracy can give some insights about what is being taught and learned. Most of these projects are located in the primary sector, since that is where most of the government and other grants have been provided to support projects designed to develop numeracy. Other articles written by mathematics educators about their own work also give us insight into what is actually happening in the classroom. Three main perspectives with some internal variations in emphasis are
described below with some examples to indicate the sources of such a categorisation.

2.3.1 Numeracy as Mathematical Skills

*Numeracy is defined as basic mathematical skills and the focus is the development of basic mathematical skills in the mathematics learning area.*

Projects with this view focus on the learning of mathematics in the mathematics classroom. Generally people holding views congruent with this category regard improving students’ learning of mathematics in the mathematics classroom as of prime importance. There is generally an emphasis on basic mathematical skills for young students that have often been restricted to *number*, but more recently there has been a trend to include *measurement* and sometimes *space* as well. Usually for older students the mathematics becomes more complex but there are many who still need to learn, and reinforce, basic skills. Some projects are set up to deal directly with students whilst others are directed at professional development for teachers.

For example, the *National Numeracy Project* in the UK was set up in 1995 in response to the TIMMS (1999) study and was concerned with the improvement of mathematics, including *number*, *measurement* and *data handling*, in primary schools. Later *space* and *shape* were added. The Numeracy Framework was produced to give guidelines to teachers on the expected student outcomes. It also included written materials to help
teachers improve their teaching. The Numeracy Task Force was set up as a research project for five years and produced the National Numeracy Strategy and training programs for teachers (Askew, 1999; Price, 1998). The Leverhulme Numeracy Research Project (Brown & Askew 2000) looking at underachievement in numeracy, incorporates the National Numeracy Strategy and focuses on the mathematics in the classroom (Brown, 2000; Brown & Askew, 2000).

In Australia and New Zealand, the projects Count Me In and Count Me In Too are school based professional development projects that focus on students’ mathematical strategies and have an emphasis on number (Gould, 2000; Bobis & Gould, 1999; Mulligan, Bobis & Francis, 1999). Another program, Count Me Into Measurement, focuses on measurement (Outhred, 2001). The Early Years Numeracy Research Project in Victoria also focuses on teaching and learning mathematics in the classroom (Clarke & Cheeseman, 2000; Clarke, Sullivan, Cheeseman & Clarke, 2000), as well as First Steps in Mathematics from Western Australia (DEET 2004) and The Year Two Diagnostic Net in Queensland (McMahon, 2001).

2.3.2 Numeracy as Mathematics in Everyday Life Taught in the Mathematics Learning Area

Numeracy is described in terms of the effective use of mathematics in everyday life but the focus in practice is of learning mathematics in the mathematics classroom, with varying levels of inclusion of real life problems.
For projects within this perspective the main focus in classrooms is on developing mathematical understanding. Students’ exposure to ‘everyday life’ examples usually consists in using ‘real world’ problems in textbooks or on worksheets. In a small number of cases the students learn mathematics by applying mathematics to situations outside the classroom that may have relevance to them. An example of this is planning a hypothetical trip, or looking at how to buy a stereo system.

For example, the *Junior Secondary Numeracy Project* in South Australia (Kuss, 2000) developed a series of assessment items related to the interests of the students outside of school. The *SAUCER* Project in Western Australia (Northcote & McIntosh, 1999) looks at the mathematics used in adult life so as to incorporate appropriate mathematics into the classroom. The *Victorian Early Years Program* (Hammond & Beesey, 1999) uses mathematics that is seen as relevant to the children and related to their experience outside school. *The Middle Years Numeracy Research Project* (Siemon, 2000; Siemon, 2001; Siemon & Stephens, 2001; Siemon & Griffin, 2000) incorporates real life contexts into teaching and assessment.

### 2.3.3 Numeracy Across the Curriculum

Numeracy is described in terms of the effective use of mathematics in everyday life and the teaching is located in other learning areas as well as the mathematics learning area.
For the projects within this perspective primary school teachers encourage students to use mathematics in all areas of their education. In secondary and tertiary education this perspective necessarily involves teachers from other learning areas and disciplines other than mathematics, and requires them to identify and make explicit the use of mathematics in their areas to aid effective learning.

There are relatively few reports and articles that describe work on developing numeracy across the curriculum where teachers are concerned about developing students’ awareness of the effective use of mathematics for learning in other subject areas and for use in a wide range of situations. For example, in the state of Queensland in Australia, the Integrated Curriculum Project (Goos, 2001) is aimed at giving pre-service teachers a cross-curricula approach to the teaching of numeracy. Thornton and Hogan (2004) report on a project in the ACT aimed at a school-wide focus on numeracy across the curriculum, including professional development activities and classroom observation and reflection. Hogan also reports on an across the curriculum approach in a numeracy project in Western Australia (DEET, 2004).

The three perspectives described in this Section are summarised in Table 2.1

Table 2.1 Perspectives of Numeracy

<table>
<thead>
<tr>
<th>Perspective</th>
<th>Teaching and Learning</th>
<th>Example of Project</th>
</tr>
</thead>
<tbody>
<tr>
<td>Numeracy as mathematical skills with development of basic skills in the maths learning area</td>
<td>Described as ‘numeracy’ basic maths skills are taught in the context of the maths learning area without any real-life contexts (except perhaps money or length etc exercises)</td>
<td>Leverhulme Numeracy Research Project in the UK</td>
</tr>
<tr>
<td>Numeracy as mathematics in</td>
<td>The development of maths skills is</td>
<td>The Middle Years</td>
</tr>
<tr>
<td>everyday life taught in the mathematics learning area</td>
<td>based in the maths learning area but there are real-world contexts such as school based fairs or other class or school activities.</td>
<td>Numeracy Research Project in Australia</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Numeracy across the curriculum—teaching located in learning areas of mathematics as well as the mathematics learning area.</td>
<td>Students learn about maths across the curriculum by learning in the other areas. Tables and graphs interpreted in social studies and science. Fractions and part measures in cooking etc.</td>
<td>The Integrated Curriculum Project in Australia</td>
</tr>
</tbody>
</table>

### 2.4 Developing Numeracy

The research literature and reports on the projects above are mainly concerned with the development of mathematical concepts and skills, and does not offer much to guide the development of numeracy interpreted as the using of mathematics in *everyday life* as described in Chapter 1. From the literature there appear to be two main components that contribute to numerate behaviour, each of which will be examined in this section. The first component involves the ways in which students learn mathematics, and how they develop their ability and skills in applying mathematics to real world situations and problems. The second component concerns the use of cognitive skills, which range from lower order thinking, such as using information and stating facts, to higher order skills including analysing, comparing, generalising and hypothesising. Some possible ways of developing both of these vital aspects of numeracy will be highlighted and explored in Sections 2.4.1 and 2.4.2.
2.4.1 Learning and Using Mathematics

In this section the first component, the learning of mathematics, is discussed and the relationships between the ways of learning mathematics and the use of mathematics in familiar and unfamiliar situations are explored. In the literature the learning of mathematics is mainly explored in two contexts, (a) out-of-school, such as in apprenticeship or work situations and (b) learning within the mathematics classroom and using mathematics across the curriculum.

Researchers in the area of situated cognition use examples from the out-of-school context to illustrate the ways that people are able to learn and develop mathematical methods in authentic situations. There are varying views among the field of researchers regarding the use of an apprenticeship model in school learning. They generally agree on the importance of authentic types of situations for the learning of mathematics. Out-of-school methods, categorised in this view as real maths, are often different from those learned in a mathematics classroom, usually described as school maths. People may develop their own strategies or use those learned in an apprenticeship or transferred from experiences in similar environments. This literature helps us to understand the complex nature of learning mathematics and the importance of flexibility in allowing students to develop their own ways of thinking, and to appreciate the wide range of approaches to problems that are possible and valid. Section 2.4.1.1 includes some of the findings of these
researchers who have analysed learning mathematics in out-of-school contexts so these may be taken into account for developing numeracy.

Just as with the development of numeracy in the out-of-school setting there is a huge body of mathematics education literature concerned with improving ways that students can learn mathematics more effectively in the mathematics classroom. It concerns students from early childhood up to those undertaking university studies. The literature includes all kinds of aspects of mathematics education including developing basic mathematical skills, mental computation, using technology and developing all kinds of mathematical concepts. The relevant aspects from this literature for developing numeracy are discussed in Section 2.4.1.2.

In addition to the learning of mathematics outside of school and in the mathematics classroom students learn and use mathematics across the curriculum. There is little reported research in this area. Here students still learn mathematical concepts in the mathematics classroom but they also develop an awareness of the wide applicability of their mathematical knowledge in other areas inside and outside school. They may also develop their mathematical knowledge in other subject areas or in activities undertaken outside the classrooms. Here students are encouraged to make sensible decisions about where the use of mathematics would be relevant and to think critically about their findings. This third context combines aspects of the first two contexts and thus has the potential to produce numerate citizens. This is considered in Section 2.4.1.3.
2.4.1.1 Learning and Using Mathematics in Out-of-School Situations

This section looks at how mathematics is learned and used in out-of-school contexts. These include everyday, community, social, working and recreational aspects of our lives that have a range of cultural and social factors that influence the way that people behave. Researchers have found that many children and adults develop mathematical strategies and processes in their out-of-school lives, which enable them to function adequately in those contexts. These methods are often different from the algorithms that students learned as school mathematics.

Examples of people who have constructed their own strategies include Liberian tailors (Reed & Lave, 1981); merchants on the Ivory Coast (Ginsberg, 1982); traders in Oksapmin, Papua New Guinea (Saxe, 1982); Brazilian cane farmers (de Abreu & Carraher, 1989); carpet layers (Masingila, 1993); street children in Brazil (Nunes, Schlieman & Carraher, 1993); and paediatric nurses in hospitals (Pozzi, Noss, & Hoyle, 1998; Noss, Pozzi & Hoyles, 1999). Lave (1988) compared the mathematics used by adults in mathematics tests with the mathematics they used in shopping situations and found that the methods were highly dependent on the setting within which they were used.

Lave, Murtaugh and de la Rocha’s (1988) investigation of weight watchers and the work of Nunes, Schlieman and Carraher (1993) found that those who were prone to error on mathematics tests were in the main able to operate without error in familiar, practical contexts, and that the mathematical methods that they used in the two kinds of contexts were different.
In the workplace it is common for the procedures used to differ from the standard algorithms used in the classroom. Masingila (1993) describes how carpet layers estimate the quantities of floor coverings needed. They need to understand concepts of length and area, but in addition they need to be able to cope with constraints related to naps, seams and hard use areas. Complex decisions to be made concerning the most effective way to measure and position the floor-covering depending on the shape of the area, the presence of obstacles, and whether the cost was to be minimised. The carpet layers take all of these into consideration adjusting the amount of carpet to the shape of the floor whether they were using carpet tiles or lengths of carpet twelve feet wide. These specific skills are learned in their training as apprentices and developed through their careers.

It is common in school for students to be given exercises that are designed to target specific understandings. For example, a textbook question “Find the cost of carpeting a floor that is 15 feet by 12 feet at $24.95 per square yard” (Masingila 1993 p. 85), does not give realistic answers if the students are not provided with constraints, such as the size of the carpet tiles or how the carpet is cut from rolls of certain widths. This type of question concentrates on the calculation of area, followed by the estimation of cost based on a unit price. To make this task more authentic, students could be given some constraints to consider, perhaps one at a time. Eventually students could be given a complex situation with a number of realistic constraints that they could approach in a problem solving way.
In another area where apprentices learn special ways of using mathematical procedures Zevenbergen (1997) describes a case study about the way that swimming pool builders in Australia go about their task. There are the excavators, the box-and-framers, and the concreters. There is a considerable amount of measuring and estimating involved in the process of making the swimming pool but these people do not use formal measuring instruments. She found that when the experts used their informal methods they were more accurate than when the novices used formal measuring instruments. The apprenticeship model works well for the novice pool builders and for the carpet layers, who are engaged in specific contexts with real constraints and are actively learning the trade. When building a pool, the box-and-framers construct a frame made from reinforced iron bars, which is self-supporting. There are building regulations that need to be adhered to if the pool structure is to pass inspection. These regulations specify a 22 cm lip for the concrete surrounds of the pool with the parallel and perpendicular bars, of various lengths, forming the base and sides meeting this lip. Each of the bars must be no more than 30 cm apart. When the pool is an irregular shape there are more angles to be worked out. All of the construction is done without formal measuring. The workers bend the bars to the appropriate angles and lengths without measuring. They use the length of their feet to estimate the distance between the bars. The distances between the bars could be less than 30 cm but to put in more bars than the minimum causes unnecessary expense. Checking revealed that fewer than 10% of the distances
had an error that was more than 0.5 cm. The apprentices learn these methods, which make the procedure much quicker than formal measuring on site, and are then able to operate effectively (Zevenbergen, 1997).

However, Acioly and Schlieman (1986) found that, whilst the bookmakers in Brazil could work successfully using only the local methods, it was only the ones who had learned mathematics at school who could adapt to novel situations. Masingila, Davindenko and Prus-Wisniowska (1996) emphasise the need for students to build links between the mathematics learned inside and outside of school so that they can develop “generizable schemas” (p.177). If people can only operate within a particular context using specific procedures they are unlikely to be able to transfer their knowledge to other contexts.

There are other situations that are not standard. Masingila, Davidenko and Prus-Wisniowska (1996) point out that percentage change is a common concept both in retailing and school mathematics. In school, percentage profit is worked out as the change from the original amount divided by the original amount. So a change from $29 to $99, an increase of $70, would be a change of $(70/29) \times 100\% = 241\%$. In the sales trade the retailer usually calculates the profit as a percentage of the selling price, so calculates $(70/99) \times 100\%$ giving a profit of about 71\% (p. 177). These differences are potentially misleading.
Other misunderstandings can arise when people think they are ‘doing the same thing’. For example, they may think that adding on a Goods and Services tax (GST) and then increasing the price by a particular percentage, is the same as increasing the price by that percentage and adding on GST. An example that illustrates the difference between the results of the two processes starts with an item costing $10 and increasing the cost by 25% with a GST of 10%. If the cost is increased by 25% and then GST is added the resulting cost is $12.50 + $1.25 = $13.75. The other way yields a cost of $11 +$3.75 = $14.75. Similarly, increasing an amount by 20% and reducing the new amount by 20%, does not bring you back to the original amount.

Willis (1998b) illustrates the difference between work and school practice with the example of a young employee who was working in a fishmonger’s shop. The employer was absent and the employee had to deal with a request for a quote for 40 prawn cocktails each containing 6 prawns. She subsequently weighed 10 prawns and found they weighed 120 grams and given the cost of $14.95 per kilogram found the cost of 240 prawns using school methods for ratio. However there are standard procedures in the trade for this kind of situation. “A caterer describing what she would do, said ‘we just assume so many prawns to the kilo for each size’ and ‘this is the way you would do it in the trade’” (Willis, 1998b, p. 36).

Lave (1985) looked at the differences between problem solving methods that adults use inside and outside of school. She found that in school students used the learned algorithms even when they could not use them successfully
but out-of-school they used a variety of techniques to help them work out the answers. In the latter cases, the problems were embedded in the contexts that were meaningful to the problem solver, and this motivated and sustained problem-solving activity (Lester, 1989). In addition “the mathematics used outside school is a tool in the service of some broader goal, and not an aim in itself as it is in school” (Nunes, Schliemann & William, 1993, p. 30). This motivating factor is fundamental but it does not explain directly why the people needed to work out their own techniques rather than to use those methods that they had learned at school.

When considering why students tend to use their own idiosyncratic methods, Boaler (1998) suggests that there is some agreement between mathematics educators that students do not use school-learned methods outside of school because they do not understand them. In accord with this view, Sierpinska (1995) proposes that one of the reasons that students do not use what they have learned at school to solve everyday problems is because they do not have the knowledge; they only have the “information or mechanical skill” (p. 6). Bransford and Schwartz (1999, p. 63) cite Lee (1998) to point out that “claims about ‘transfer failure’ have been traced to inadequate opportunities to learn in the first place”. This may also be because, as Resnick (1987) pointed out that:

...the process of schooling seems to encourage the idea that ... there is not supposed to be much continuity between what one knows outside of school and what one learns in school (p. 15).
Nunes *et al* (1993) point out that ‘street maths’, the kind of mathematics developed outside school, used by farmers, apprentices, traders and so on is not the learning of particular procedures repeated in an automatic, unthinking way but involves the development of mathematical concepts and processes. Street mathematics represents one example of cognition in practice as Lave (1988) calls it. It is a type of thinking that is carried out under the constraints of social, empirical and logical rules that aims at accomplishing a result (p. 153).

The acceptance of constraints and then working towards a solution is part of the process that motivates people to persevere and find solutions. Students enrolled in apprenticeship schemes acquire the mathematical skills required for the trade through a combination of further study and on-site training. They can learn mathematics in the context even though they were often underachievers in mathematics in the school environment.

Other researchers in this field of situated cognition, Brown, Collins and Duguid (1989) emphasise the need to think of knowledge as located in situated forms of activity rather than just a content mental exercise:

Activity and situations are integral to cognition and learning, and how different ideas of what is appropriate learning activity produce very different results. We suggest that by ignoring the situated nature of cognition, education defeats its own goal of providing usable, robust knowledge. And conversely, we argue that approaches such as *cognitive apprenticeship* (Collins, Brown & Newman, 1989) that embed learning in activity and make deliberate use of the social and physical context are more in line with the understanding of learning and cognition that is emerging from the research (p. 32).

These cognitive apprenticeship methods refer to the kinds of enculturation found in craft or trade apprenticeships.
Agre (1997) analysed the work of two situated cognitivists, Lave (1988), concerning the workplace, and Walkerdine (1988), about school contexts, and noted that despite their differences they “share a compelling view of mathematics not as an abstract cognitive task but as something deeply bound up in socially organised activities and systems of meaning” (p. 71). He points out that this is in accordance with other workers in the field who want to see mathematics viewed as a sociological phenomenon that is part of all aspects of peoples’ lives.

These views of learning have implications for the development of numeracy in educational settings. In mathematics classrooms, students are often presented with rules and formulas that they are then expected to use in sets of exercises, which engenders the misconception that mathematics is simply procedural. Thus students are likely to assume that there is only one correct method of completing a calculation or solving a problem. They are not usually encouraged to take into account the kinds of constraints that would occur in the real world. Based on their observations of learning mathematics outside the classroom Brown, Collins and Duguid (1989) emphasise the need to provide students with authentic settings, by which they mean situations that do include the realistic constraints. They say that students “[n]eed to be exposed to the use of a domain’s conceptual tools in authentic activity – to teachers acting as practitioners and using these tools in wrestling with problems of the world” (p. 34). In this respect they advocate that the teachers work with the students in realistic problem solving activities where the
teachers do not already know the answer, so that the students can observe the ways that the teacher as mathematician approaches the problem.

2.4.1.2 Learning and Using Mathematics in the Mathematics Learning Area

In contrast to the previous section the purpose of this section is to examine changes and trends in mathematics curricula and teaching within mathematics learning areas. It highlights the kinds of changes that have been brought about by an increase in the sophistication of technology available to many students in classrooms in the developed world. Internationally there are comparisons made to investigate commonalities and differences and evaluate successful approaches. It can reasonably proposed that there is some relationship between the degree of success in learning mathematics and the development of numeracy if the links and connections are made between mathematics and numeracy.

There have been many changes in the field of mathematics education over the last few decades. The changes to mathematics curricula including content and teaching methods are bringing about changes in expected learning outcomes. From the mathematics education literature it appears that in many Western countries there is more emphasis on the development of students’ understanding of mathematical concepts than on memorisation and rote-learning. The change in emphasis towards deeper understanding varies between schools, states and nations.
One of these changes is the incorporation of technology into mathematics education that has brought about different ways of thinking in countries that can afford calculators and computers. Some discussion about the use of technology in education is included in this Chapter because it is almost impossible to consider any mathematics education in the absence of technology in developed countries. In the empirical study reported in this thesis the students consider tables of data and a thorough understanding of data analysis can be developed using technology (Ben-Zvi & Garfield, 2004b).

The introduction of technology into mathematics education has influenced primary, secondary and tertiary levels of education. The technology includes calculators (four-function, scientific and graphics) and computers. The impact can be identified in the changes in the curriculum as well as the changes in activities and experiences in the classroom.

There is ongoing international discussion about the appropriateness of the use of calculators and computers in the mathematics classroom. There are variations in the extent to which the technology is integrated into the mathematics courses and this is changing all the time. (Kemp, Kissane & Bradley, 1995, Bradley, Kissane & Kemp 1996; Kissane, 2000, 1995; Waits & Demana, 2000; Forster & Taylor, 2000; Forster 2004,) The literature suggests that the impact of technology has been greatest in those countries for which essentially all students have access to computers and graphics calculators.
This access and impact is reflected in the state or national guidelines. For example in Western Australia the *Curriculum Framework* (1998) and the *Student Outcome Statements* (Education Department of Western Australia, 1994), which guide teachers in the content and process of the curriculum, assume that all students have access to computers or calculators where appropriate.

The introduction of technology into schools has influenced both the content of curricula and the teaching methodology. Although worldwide a very small minority of students have access to their own laptop computers in schools, it is generally more the case that any access to computers in school or higher education institutions, has been in a laboratory setting. Graphics calculators, which can be regarded as small hand held computers, can make a large difference to the curriculum because they are more accessible due to of their size and cost (Bradley, Kemp & Kissane, 1994, Moore, 1997a).

Whether students are using computers or graphics calculators, the main curriculum areas where they are used are substantially the same. The introduction of scientific calculators enabled students to perform calculations with real world numbers rather than ‘cleaned up’ numbers for easy manipulation, when calculations were done by hand or with logarithm tables. We have now moved on to a world where computers and graphics calculators can graph functions and statistical data, construct tables of values, solve equations, work with matrices, run programs, and much more.
In those countries where technology is available, curriculum writers have thought carefully about what should be in the content of the curriculum realising that now some tasks that were previously quite time consuming can be done easily and quickly on technology. Students can now solve equations that were inaccessible to them through graphing or constructing tables, they can work with data looking at different methods of data analysis and considering the merits of each. They can find confidence intervals and perform hypothesis tests. In the last few years even computer algebra systems have become available on some of the graphics calculators. This development has lead to debate about whether students should be allowed to use them in schools and public examinations, and what modifications to curricula are desirable.

It is very important that students learn to choose the most appropriate technology to the task in hand (Bradley, Kemp & Kissane, 1996). This may be a scientific calculator or a graphics calculator or a computer but in may be much more efficient to use mental methods. While the technology is there available for use it is still necessary for students themselves to make decisions about the ways to solve problems or perform computations. Indeed, the introduction of technology has not removed the need for higher order thinking skills. If anything, these are more important than ever before with the need for problem solving and mathematical modelling in the workplace.
There are concerns about this introduction of technology into the classroom. Kaput and Thompson (1992) highlighted some of the potential mismatches between the environment for learning mathematics and the introduction of technology. In this context Galbraith and Haines (1998) investigated the attitudes and beliefs of students learning in a computer based learning environment. The course followed by first year undergraduate related concepts to practical applications and used technological and analytical methods. The study sought to identify specific attitudes related to learning mathematics and towards technology so as to examine any relationships between them. They found that “computer influence is dominant in determining computer-mathematics interactions which will be expected to have significant impact when integrating the use of computers and graphics calculators into the undergraduate curriculum” (p. 288). This indicates that when the attitudes to computers and to learning mathematics are divergent there are potential issues to be concerned about. This seems to indicate that the use of technology for learning mathematics needs to be carefully planned to incorporate positive attitudes towards the use of the technology.

Although there is generally less emphasis on the use of technology in primary than secondary or tertiary classrooms there is some range in the usage. The large number of projects, referred to earlier in this chapter, which are concentrating on improving children’s mathematics, under the umbrella of numeracy, show that there is an enthusiasm to improve primary school children’s mathematical knowledge. It would be reasonable to expect that
there will be a variety of success stories within these projects. New curricula, that emphasise the importance of process as well as content, contribute to classroom practice in such aspects as mental methods, practical work and discussion. For example, projects such as *First Steps* (DEET, 2004) in Western Australia encourage teachers to think in ways that will enable them to provide suitable learning environments for learning mathematics.

Current literature is showing how vital it is to make explicit the links between mathematics and numeracy with the increasing importance of statistical literacy (Gal, 2002, 2003; Watson, 2003; Watson & Callingham, 2003; Ben-Zvi & Garfield 2004). It is essential that the development of statistical literacy from an early age is included in the mathematics curriculum. Teachers of mathematics for all ages of children can help students to develop this literacy in the mathematics classroom through a range of strategies and development of critical thinking.

### 2.4.1.3 Using Mathematics Across the Curriculum

In the previous Section it was highlighted that approaches to learning mathematics in the mathematics classroom (or mathematics learning area) have changed over the last decade. In itself this is a move in the right direction for mathematical concept development but this does not necessarily enable or encourage students to use their mathematical knowledge in other situations outside the mathematics classroom. If teachers do not help students to understand the connections between the different
contexts in which mathematics is useful, it is more difficult for students to make those links.

There is considerable variation in the ways that this is encouraged in Western Australia, for example, curriculum documents for primary and secondary schools show that there is a section *Numeracy Across all the Learning Areas* in the Curriculum Framework (Curriculum Council, 1998). It states that “At all levels teachers of mathematics should help students to use their mathematics to solve practical problems and as a tool for learning beyond the mathematics classroom” (p. 215).

The main purpose of this section is to highlight characteristics of educational experiences that enable students to use their mathematical knowledge in familiar and unfamiliar situations, inside and outside the classroom. In this thesis the term *transfer* is used to denote the use of mathematical or procedural skills in these different situations. The nature of transfer from different perspectives will be illustrated below.

Students need the ability to transfer knowledge from one context to another as well as a sound knowledge of mathematics. From the mathematics education literature it can be seen that the ability to use mathematics in other contexts is related to the ways in which the mathematics is learned in the mathematics classroom. It appears that some teaching methods, and student experiences, are more likely to facilitate transfer than others. The type and quality of the learning environment influences the extent to which students
learn. The students need to be able to feel comfortable with each other and to be able to discuss their work and ideas together. This leads to negotiation and creates similarities with real world experiences (Boaler, 1999). In addition open-ended, problem-solving approaches can be more effective than traditional textbook approaches (Cognition and Technology Group at Vanderbilt, 1990).

There is considerable debate as to what is meant by transfer. Much of the discussion concerns the ways in which different ways of learning mathematics might enhance the transfer of learning. The literature includes examination of aspects of the learning of mathematics in the mathematics classroom including the relative merits of memorization, discussion and problem solving as well as the influence of the social environment in which the mathematics is learned.

Previous theories on transfer, based on the work of Thorndike (1929), hypothesised that near transfer (from one school task to a highly similar school task) and far transfer (from school subjects to non-school subjects) could be facilitated by teaching knowledge and skills in school subjects that have elements *identical* to activities encountered in the transfer context (Klausmeier, 1985). However these theories did not take into consideration the characteristics of the activities or the learners. In modern theories there is more focus on the kinds of learning activities that are appropriate, as well as the nature of the learners’ backgrounds (Singley and Anderson, 1989). Some
discussion concerning the types and components of appropriate learning activities is included in this Chapter.

All learning involves some transfer as construction of knowledge is based on previous learning (Resnick, 1989). The literature suggests that transfer of ideas and concepts is an active process that is not the same as being able to complete a different subtask in a set of exercises. It has traditionally been considered to be the ability to use previously gained knowledge in novel situations without explicit prompting. However, more recently Bransford and Shwartz (1999) proposed that this view of transfer should be extended to include people’s “preparation for future learning and their ability to learn in knowledge rich environments” (p. 68).

Many studies have shown that the manner in which the subject matter is learned is an important indicator of the extent of transfer. Proponents of a mathematics curriculum which includes open-ended, practical, investigative work and which includes the requirement for students to make their own plans and decisions and to apply their mathematical knowledge claim that these are beneficial in a variety of ways, including the promotion of transfer (Cognition and Technology Group at Vanderbilt, 1990; Roth, 1996).

Researchers taking a situated learning perspective are concerned with the ways in which the teaching approach influences the ways in which people learn, and transfer this learning to be used in other situations. Boaler (1997a, 1998, 1999, 2000, 2002) analyses a project undertaken to compare the learning
of mathematics in two different learning environments. Two UK schools with students from similar socio-economics groups were chosen. One had a content-based approach and the other a process-based approach. Students in the content-based school had a “view that mathematics was all about memorising a vast number of rules, formulas and equations, and this view appeared to influence their mathematical behaviour” (Boaler, 1998, p. 45). They were unable to tackle subtasks that were slightly different from what they expected or for which there was no cue for them to follow. At the process-based school the students worked on open-ended projects where they had to think for themselves, interpret situations and choose the appropriate mathematical procedures. The approach was based on a “philosophy that students should encounter a need to use mathematics in situations that were realistic and meaningful to them” (Boaler, 1998, p. 47). These students were able to see the connection between the mathematics they did in school and the mathematics they needed outside school whereas the students in the other school did not. Evaluations of the students showed that the students using the process-based approach performed better on applied activities, as would be expected, but performed no worse on the traditional closed subtasks.

Further analysis led Boaler (2000a) to consider why the students with the content-based curriculum did not use their school mathematics outside the classroom. The students reported that:
“they did not even attempt to make use of school-learned methods in the real world, not because of the form or structure of the problems they encountered … but because the environments of the classroom and their everyday lives were too disparate. The students believed that adopting classroom practices in the real world was inappropriate, so they did not attempt to draw on school mathematics” (p. 114).

Masilinga, Davidenko & Prus-Wisnioska (1996) found three key differences in comparing in-school and out-of-school practices: the goals of the activity, the conceptual understanding of persons in each context and the flexibility of constraints. They found that the goals of people in out-of-school contexts were quite different from the students’ goals as long the students saw the problems as school problems. When in that everyday situation the students viewed the problems differently. The work of Masingila et al, based on Saxe’s (1991) framework, led them to suggest that many of the differences between mathematics learning and practice can be narrowed down by providing experiences that engage students in doing mathematics in school that are similar to the ways that mathematics is done outside of school.

Similarly, Boaler (1997, 1998, 1999, 2000a, 2000b) analysed the differences between the different classroom communities at the schools. At both of the schools, students demonstrated the situated nature of their learning. Boaler suggests that it is not enough to learn mathematical concepts, even when they have a chance to construct their mathematical understanding. They also need opportunities to use their mathematics, adapt and change methods, and to discuss and negotiate with other students. The environments in the classrooms at process-based school were closer to the real world and so students generally were more able and willing to use school-learned
methods than students from environments that were more procedural and school bound.

Furthermore, Boaler’s (2000c) analysis of in-depth interviews with students from six schools in Northern Californian schools, and references to the literature make it clear that the environment in mathematics classrooms has a large effect on how students view mathematics, along with their willingness and confidence to use it in out of school situations. When looked at through a lens of situated learning the traditional maths classroom, where mathematics is seen as a set of procedures incorporating artificial *real-world* problems, is unlikely to enable students to link their mathematics with the real world. On the other hand, classrooms in which realistic constraints, negotiation and learning happen within a culture of social interaction and problem solving appear to give students the opportunity to see the relevance of the mathematics they are learning to the real world. Thus ‘transfer’ becomes, if not automatic, then at least more likely.

### 2.4.2 Developing Cognitive Skills

The learning and use of mathematics as part of becoming numerate has been considered in Section 2.4.1. In this Section the second component of developing numerate behaviour, the development of cognitive skills, is discussed. Clanchy and Ballard (1995) emphasise the importance of developing cognitive skills at all levels of study. The development of higher order thinking skills enables people to move from a position where their typical responses to stimuli are trivial, in the form of isolated pieces of
information, to a new position where responses are more meaningful, and include explanations and predictions. The importance of this development is explained in Chapter 3.

Two ways of looking at the development of cognitive skills are of relevance in the context of developing numeracy, and each of these has influenced the development of this study. Both perspectives draw on the work of Jean Piaget but each brings different perspectives that contribute to the discussion about developing thinking, and they will be described in this section.

Firstly, this study was influenced by the findings of a cognitive acceleration research program that developed a method of teaching for cognitive acceleration. In 2000, the Minister for Education in the UK, the Right Honourable Blunkett, acknowledged the success of the cognitive acceleration program developed by Shayer and Adey from the Centre of Advanced Thinking at King’s College. Their method of teaching for cognitive acceleration became part of a DfEE initiative and was introduced into 13 local authorities in September 2000. The features of the method of teaching are identified in Section 2.4.2.1. Chapter 3 describes how the method of teaching chosen for the workshop relates to the method described by Shayer and Adey (e.g. Adey & Shayer, 1994; Shayer & Adey, 1981).

A second perspective on the development of thinking skills is concerned with identifying levels of responses using the Structure of the Observed Learning Outcome (SOLO) taxonomy developed by Biggs and Collis (1982),
and subsequently providing related activities that would allow students to build on their existing levels of response. Biggs and Collis developed their taxonomy after examining a large number of students’ written responses to open ended questions across the curriculum. The taxonomy described in Section 2.4.2.2 provides a way of evaluating qualitative material across disciplines. The taxonomy also provides a framework within which teaching can be planned to help students develop their levels of thinking.

2.4.2.1 Cognitive Acceleration

A major contribution to the field of cognitive acceleration has been made in the science area by two large-scale projects. These were Cognitive Acceleration through Science Education CASE I (1980-1983) and CASE II (1984-1987) which were under the direction of Shayer and Adey and funded by the British Social Economic Research Council (Shayer & Adey, 1981; Adey & Shayer, 1990 ; Shayer & Adey, 1992; Shayer & Adey, 1993; Adey & Shayer, 1994; Jones & Gott, 1998; Shayer, 1999). The projects were started in response to the results from a representative survey of 1400 British 10 -16 year olds based on Piagetian measures, which were reported by the Concepts in Secondary Mathematics and Science program (CSMS) (Shayer, Kuchemann & Wylam, 1976; Shayer & Wylam 1978). These showed that no more than 30% of secondary school students develop even early formal thinking by the age of 16. This was considered to be a problem since students need to be thinking at the early formal thinking stage by age 13 -14 to be able to participate successfully in higher-level science and mathematics courses.
Cognitive Acceleration through Science Education, (CASE), (Adey & Shayer, 1994) is an intervention strategy that incorporates both curriculum materials and teaching methodology based on an amalgam of the theoretical principles of learning put forward by Piaget, Vygotsky and Feuerstein. It also incorporated Piagetian tasks in the pre- and post-tests and in the learning activities. The project was located in the area of science because the project team were scientists and science teachers. The project was set up to investigate the possibility of raising general awareness of thinking amongst average students between 10 and 14 years old. The project team decided to embed an intervention within the school science programme so that the students could develop their thinking skills in a context in which they and the teacher were already involved. Regular science lessons are undertaken as well as those designed to increase the level of thinking. The programme is used in mathematics classes where it is referred to as Cognitive Acceleration in the Mathematics Classroom, (CAME).

Shayer and Adey (1993) describe the research as an intervention since “the short-term intention was to increase the cognitive development of the students, not to improve their science achievement” (p. 352). Adey and Shayer (1994) use the term intervention as related to the rate of cognitive development that in the long term increases achievement, whereas in instruction relevant thinking ability is assumed, and the aim is to maximise students’ achievements in response to appropriate learning experiences. In this thesis the terms intervention and instruction are used in this way.
One of the major differences between instruction and intervention is in the kinds of questioning strategies used. The questioning in the context of instruction clarifies the nature of the specific concept being developed and primarily depends on teacher to student interaction. However, in the intervention that Shayer and Adey used the questions were designed to be open and challenging, and two-way student-student and teacher-student discussions were an integral part of the intervention. Influenced by the work of Vygotsky (1986), their emphasis was on discussion in small group and whole class contexts. Indeed Shayer (1999) emphasises the importance of “social construction of reasoning through meta-cognition and a carefully managed use of the language of thinking” (p. 895).

The evaluation of the effectiveness of the intervention was based on Piagetian Reasoning tasks that had been used in the CSMS survey (Shayer & Adey 1993). Students were tested before the intervention and at later dates after the intervention. Jones and Gott (1998) criticised Shayer and Adey’s choice of measuring instrument because it was based on the same principles as the intervention. Shayer responded to this criticism, saying that “[i]t makes no sense to avoid using a measuring instrument because it is based on the same theory as an intervention. Such an instrument would be most scientist’s first choice to see if their use of a theory has or has not been valid” (p. 884).

This is also the case in the empirical study reported in this thesis. An analysis was undertaken to find out whether the criteria, based on the same
underlying principles as the intervention, could be used for evaluation of students’ responses in the context of reading tables.

A report published in the UK in May 2000 (Shayer, 2000) showed that there were significant gains for students in all kinds of schools. The report indicates that students who had experienced the CASE methodology achieved statistically significantly higher results in the General Certificate of Secondary Education (GCSE) science, mathematics and English exams, than those students who had not been exposed to the CASE methodology.

Although these intervention programmes were used in high schools, rather than tertiary institutions, they were found to be effective for students with wide range of ability levels, and therefore the approach to the development of higher order thinking skills is be regarded as appropriate for use with undergraduate students.

In the area of accelerated cognition, Adey and Shayer (1994) have identified five main categories that reflect the educational principles of Piaget, Vygotsky and Feuerstein: duration; concrete preparation; construction; bridging and meta-cognition. These categories are described in this section and their relevance to the empirical study is highlighted in Chapter 3.

The duration and density of the two-year time interventions of the CASE projects was based on Lipman’s and Feuerstein’s work (Adey & Shayer, 1994). The time span was considered important because the development of thinking would be related to maturation as well as external stimuli. The
regular science lessons are undertaken in the curriculum in conjunction with
the special lessons designed to increase thinking.

*Concrete preparation* involves establishing the meanings of any new technical
vocabulary to be used in the lesson so that the difficulty or challenge of the
situation is not due to students being unfamiliar with the terms. This is
slightly different from the normal instructional approach where the terms are
not introduced until the students encounter physical manifestation of the
concept. In teaching new technical vocabulary, the CASE approach maintains
that the context for the practice examples given should require processing at
no more than the upper concrete level. In addition the introduction should
be interesting and new to stimulate curiosity and motivate students’
involvement.

Cognitive conflict often occurs in the *construction* category when students
encounter an idea or concept that is discordant with their previous
conceptual framework. Teachers create cognitive conflict by providing
situations that include events or observations to provoke students to adjust
their ways of thinking by using a new framework that can incorporate new
ideas.

Fensham and Kass (1988) point out that in some cases surprising events will
simply be ignored as inexplicable or uninteresting by students. So the teacher
needs to plan so that the conflict is located within a context that is familiar to
the students and makes some cognitive demand not too far out of their reach.
In normal teaching lessons, or *instruction*, the students are learning to develop a concept and the teacher might highlight students’ misconceptions and expose them to some evidence that disputes them. In the *intervention* they are concerned with “the development of the schemata of formal operations then the cognitive conflict provided should be such as to help pupils construct these reasoning patterns for themselves” (p. 63).

In *bridging* and *reflection* the new thinking developed in the lesson is related to other examples in science or mathematics and in the everyday world. In this way the students are able to generalise the reasoning patterns and further develop them through use. Students are encouraged to think of instances where they might have used similar ideas before and new ones where this kind of thinking might be applied, and to discuss these in the lesson. This is a vital part of the intervention because, as Adey and Shayer (1994) point out:

> If bridging is the conscious transfer of a reasoning pattern from a context in which it is first encountered to a new context, then the transfer is most likely to be effective if the reasoning pattern has been made conscious and verbalised (p. 73).

They explain this further:

Transfer is possible, that it is very much a matter of how the knowledge and skill is acquired and how the individual, now facing a new situation, goes about trying to handle it. Given appropriate conditions such as cueing, practising, generating abstract rules, socially developing explanations and principles, conjuring up analogies...and the like transfer from one problem domain to another can be obtained (p. 22).

The method of teaching for this intervention is crucial if transfer is to be optimised. Teachers help students to share their experiences of the lesson
and to summarise what worked in that context. The abstraction of the essential aspects and the naming of them can help students to recall the procedures in order to use them in other situations.

*Meta-cognition* is described as thinking about your own thinking and learning, and it is another essential part of the *intervention*. Meta-cognition is advocated by most cognitive psychologists, and many others, for the development of higher-order thinking. However, there is no one definition of meta-cognition and this can cause confusion. Adey and Shayer have based much of their work on Vygotsky and Feuerstein who have also contributed to the way they view meta-cognition. Feuerstein *et al.* (1980) emphasise the role of adults as mediators who encourage meta-cognition by creating questions probing what they found difficult, and how they overcame those difficulties. Vygotsky (1978) describes the use of language as a mediator of learning. Adey and Shayer also use the work of von Wright (1992) to illustrate two levels of reflection. At the lower level the person is reflecting on features in the world and their ways of coping in familiar situations, and at the upper level the person reflects on their own knowledge or way of thinking. It is in the latter sense that Adey and Shayer (1994) are using the term meta-cognition. They stress that meta-cognition is the “conscious, reflective awareness about strategies – thinking about thinking (and action)” (p. 71).

Shayer and Adey (1993) note that meta-cognition is likely to be an important part of any programme aimed at developing general thinking skills. They
also point out that Schoenfeld’s (1989) teaching for problem solving involved having students asking themselves questions about what they were doing and whether it was getting them anywhere, and whether they should be doing something else instead.

2.4.2.2 The SOLO (Structure of the Observed Learning Outcome) Taxonomy

The second perspective that informed this study was the SOLO taxonomy. This section describes the SOLO Taxonomy and the ways in which it can be used for evaluation and teaching. The taxonomy, designed by Biggs and Collis (1982), was based on their study that analysed the written responses of hundreds of elementary, high school and college students to open-ended subtasks across a range of subject areas including history, mathematics, English, geography and modern languages. The focus of their interest was to evaluate how much and how well a student had learned, thereby involving both quantitative and qualitative aspects (Marton, 1976).

To measure how much has been learned is relatively simple as this usually consists of counting up how many facts can be recalled or procedures applied. Much of educational testing is concerned with this quantitative aspect and much includes multiple-choice tests. Although multiple choice tests are easy and quick to mark by hand, or using an electronic scanner, there are weaknesses with this kind of testing. Students can guess the answers, the correct answer may not be discernible to a student if it is in a different form, and they allow no part marks for method. The quantitative nature of multiple-choice subtasks is obvious, but the quantitative aspects of
the ways that some people mark essays is less obvious. Biggs and Collis (1982) point out that in many cases essay marking has consisted of an allocation of points for each correct fact which are then added up to give a number which is converted to a ratio, adjusted for quality and converted to a grade. Hence even the marking of such open-ended subtasks relies more on how much than on how well.

Given that in many cases the increase in length and detail of the student’s response goes alongside improvement in the quality of the integrated structure of the response, the allocation of a grade based on quantity plus a bonus for ‘quality’ may well result in a similar grade. However, it is important to be able to justify to students why they were evaluated in that way so that they can improve their work.

The construction of the SOLO taxonomy was influenced by the work of Jean Piaget, who identified stages of development: pre-operational, early concrete, middle concrete, early formal and formal based on responses to Piagetian tasks, (Flavell, 1963). It was assumed that for the classical Piagetian tasks the logic of the student needed to match the logic of the task and this determined whether the task would be completed. It is generally accepted by stage theorists that students develop through the Piagetian stages in the various domains of mathematical reasoning.
The SOLO taxonomy measures the level of a response, and not the stage of the student. Students may respond at different levels in their responses depending on the context. Biggs and Collis (1982) emphasise that:

The distinction between describing responses and describing people, is important because teachers who know their Piaget often find it difficult to understand why a child who “should: be formal operational on the basis of his age or performance on other tasks, appears to be concrete operational, or even pre-operational, in others (p. 23).

They have suggested that the variation between performance in school-related tasks and the performance on Piagetian tasks is due to “general learning process factors, and to factors that are specific to the task and the testing context” (p. 213).

The SOLO taxonomy has five levels: prestructural, unistructural, multistructural, relational and extended abstract. The characteristics of the responses at these levels given below are based on those of Biggs & Collis 1982; Biggs & Telfer 1987; Collis & Biggs 1979 and Lake 1999, 2002:

The responses are characterised as **prestructural** when the student:

- Is confused about the subtask
- Avoids the subtask and
- Closes without even seeing the problem.

The responses are characterised as **unistructural** when the student:

- Uses a single piece of relevant information
- Generalises in terms of only one aspect and
- Has no feeling for consistency and so jumps to conclusions and may close prematurely.

The responses are characterised as **multistructural** when the student:
• Uses several pieces of relevant information that do not relate to each other
• Generalises on a limited number of unrelated events and
• Shows a feeling for consistency but can be inconsistent due to focusing on separate aspects of the data.

The responses are characterised as *relational* when the student:

• Considers data as a whole with coherent structure and meaning,
• Looks at the interrelationships between the data
• Generalises within the context of the available data and
• Seeks consistency within the data but does not draw on data beyond that under consideration.

The responses are characterised as *extended abstract* when the student:

• Uses and interrelates all the available information and tests it against hypotheses suggested by the data
• Uses deductive logic to relate specific data to general rules
• Generalises about hypothetical situations, and
• Resolves inconsistencies.

At whatever Piagetian level a person is operating, the five levels of the SOLO taxonomy can be shown. Biggs and Collis (1982) point out that within each of the Piagetian stages:

a common pattern of response structure can be discerned. That individual enters each stage when he can give a unistructural response within the given mode of functioning. Next, as he becomes more and more familiar with the elements and operations associated with that mode he *adds* to his repertoire of mode-appropriate responses, which gives him a multistructural response. He then, with further experience, learns to integrate those responses, giving him the highest level of response within a mode, relational. To transcend the relational level is to enter a new mode. Thus, the extended abstract response in one mode jumps the barrier to form the unistructural response for the next higher mode (pp. 217-218).

Much of the work of Biggs and Collis was carried out with school-aged children who would have been normally working within varying levels of concrete operational thought. There were few who would have been
operating at the formal stage and therefore there were limited opportunities for observation at this level. Their detailed examples of the use of the taxonomy in subjects across the curriculum indicate the usefulness of the taxonomy for evaluation of student learning from primary to tertiary education (Biggs & Collis 1982; Pegg, 1997; Biggs, 1999; Forster, 1999; Hodges & Harvey, 2003; Watson & Callingham, 2003). The evaluation instrument used in the study that will be described in detail in Chapter 3 was based on this work.

The nature of the SOLO taxonomy makes it suitable for use as a framework in teaching. In the learning and teaching of an academic subject there are both content and cognitive process objectives. The content objectives include development of students’ knowledge of the concepts and facts of a particular subject. The process objectives relate to the strategies and skills that are appropriate to the ways of thinking about the subject (Bruner, 1960; Biggs and Collis, 1982). The SOLO taxonomy can productively be used in aligning curriculum objectives and learning outcomes, so that it is possible to find out how well students have learned what they were intended to learn. Biggs (1999) clarifies this relationship with reference to teaching at university, but the underlying principles are applicable at other levels of learning. He points out that the objectives are concerned with “student’s learning activities, not teachers’ teaching activities” (p. 44). These objectives need to contain not only the content or process aspects but also the level of understanding that the students are to achieve.
Biggs (1999, p. 46) has suggested that the levels of students’ understanding can be described by verbs that relate to the SOLO taxonomy:

- **Prestructural**: Misses the point
- **Unistructural**: Identify, do simple procedures
- **Multistructural**: Enumerate, describe, list, combine, do algorithms
- **Relational**: Compare/contrast, explain causes, analyse, relate
- **Extended abstract**: Theorise, generalize, hypothesise, reflect

These verbs are general and go towards describing the kinds of levels that might be expected in students’ responses, but in addition there would also be specialised verbs that are found in particular domains. The teacher can also decide to target specific levels when planning instruction. The SOLO taxonomy reflects the hierarchical nature of learning and is useful in the construction of the categories for assessing responses across the curriculum (Biggs, 1992). First, starting with the minimum standard acceptable and working through to what might be the best you could expect, teachers can set out different levels of response or grades which are described by different levels of verbs (Biggs, 1999). These qualitative grades tell students something meaningful about their outcomes, rather than simply allocating a mark.

The measurement of numeracy in general is an area that is in need of research and careful consideration. Given the nature of numeracy itself as being mathematics in a relevant context, described in detail earlier, the evaluation of students' numeracy skills is not a trivial matter. Numeracy
benchmarks in the UK and in Australia are often based on paper and pencil tests of basic mathematical skills which do not include any authentic contexts, or real attempts to measure numeracy. In Western Australia the Western Australian Literacy and Numeracy Assessment (WALNA) tests include a numeracy component that is a multiple choice mathematics test. This test is given to all students in grades 3, 5 and 7, and the results of the tests are sent to the parents. This practice can effectively distort the way that teachers view their responsibilities in developing students' numeracy and their mathematics skills. There appears to be the possibility of moving towards a national test of numeracy that will be of a similar nature.

Also located in Western Australia is a Monitoring Standards in Education Project (MSE) part of the monitoring and accountability of processes of the Department of Education and Training of Western Australia. The random sample assessment program involves administering that administers a test to a sample of students in order to monitor the education system and the results of which tests are not sent to parents. Teachers are employed to mark these test so they can be of a richer nature than WALNA tests and more appropriate for the measuring of numeracy.

A recent research study Numeracy Across the Curriculum DEST (2004), carried out in Western Australia using a sample of schools and students, considered it important to develop and measure students' numeracy skills. The numeracy assessment tasks developed and used in this project were generally much richer, embedded in other learning areas, had greater
emphasis on reasoning and problem solving, and had much greater contextual and language demands' (p.51) as compared to the WALNA tests. The report on this project shows that it is possible to design assessment tasks to measure students' numeracy. However the literature shows that available material on assessing student numeracy seems to offer little insight into ways of measuring students' numeracy. In this context the SOLO taxonomy offers a framework for measuring students' numeracy and the development of their numeracy. This framework is explored and developed for measuring students' numeracy in reading tables of data in connection with the empirical study reported in this thesis.

The SOLO taxonomy is not a method of teaching, but Biggs and Collis (1982) suggest “that setting of objectives at various SOLO levels in part determines the appropriate teaching method” (p.87), allowing for students to think rather than having too much content. More recently Lake (2002) pointed out that “[b]ecause the SOLO taxonomy is based on analysis of material presented rather than the individual presenting the material, it has the potential to be used as a powerful teaching tool” (p. 6). He developed a template, based on the taxonomy, for developing critical numeracy in the social, biological, environmental and other sciences (Kemp & Lake, 2001; Lake 1999, 2002; Lake & Kemp, 2001).
2.5 Learning Numeracy Through Other Subject Areas

In Section 2.4 the development of numeracy was explored in out-of-school situations where people adapt their mathematical skills to fit the context or they develop context specific strategies. It also considered the ability to use the mathematics they have learned in the mathematics learning in the service of other subject areas. In the latter it was mainly concerned with the literature on transferring mathematical skills to other subject areas.

However there is another dimension to the development of numeracy. Students also learn numeracy-related concepts in other subject areas. The ways that students develop their numeracy skills within woodwork, say, can be different from the way they develop those within science (AAMT, 1997).

This is particularly pertinent at college or university when students are not studying mathematics courses, but do encounter quantitative information in text, charts, tables and graphs in non-overtly mathematical subject areas such as sociology, psychology, education, media, marketing, science and statistics. Although often found in the same faculty many statisticians claim that statistics is different from mathematics (e.g. Moore 1997b, 1998), that although statistics is a mathematical science, it is not a subset of mathematics.

In the different areas they need to develop their ability to critically interpret that quantitative material. How the students develop their skills depends on a number of factors including their mathematical background, although even...
students with a strong mathematical background may not think to apply it in other subject areas (Chapman, 1988), the pedagogy and content of the courses, the learning activities provided as part of the courses and so on. There are differences between the ways that to students are expected to respond to quantitative material in the different subject areas, the extent to which these expectations are made more or less explicit.

The concept of critical numeracy ties together the numeracy, being able to apply the appropriate mathematics and interpret the results, with the critical reading that begins early in schooling at part of the development of literacy. There is extensive literature on literacy and critical literacy, and a review of this is beyond the scope of this dissertation. However it is helpful to refer to Luke and Freebody (1999) and an aspect of their three dimensional model of literacy:

Critically analyse and transform texts: understanding and acting on the knowledge that texts are not neutral, that they represent particular views and silence other points of view, influence other people’s ideas; and that their designs and discourses can be critiqued and designed in different ways (p.7).

This is linked to students developing an understanding that the way data is collected and compiled is not neutral, and that this underpins the messages of the tables and graphs. If students are looking at the text referring to tables and graphs it is important that they do not necessarily accept the author’s point of view but consider carefully the implications for themselves. In addition they need to question reliability and validity of the data presented.
As part of their study and their everyday lives students are accessing websites which provide a large range of data. Some sites provide structured activities centred around a variety of subject areas. For example the Australian Bureau of Statistics provides data in the form of text, graphs and tables as well as lesson plans suitable for use in areas such as mathematics, science, social studies. There is also a large section devoted to the 2006 census within which students can input data from their schools. In the USA similar resources exist such as the US Census Bureau, National Agricultural Statistical Service and the National Center for Health Statistics (Gal & Ben-Zvi, 2004).

In both the physical and the biological sciences there are numeracy demands. Students are required to construct and interpret tables and graphs and to consider their implications and meanings in the specific contexts (Kugler, Hagen & Singer, 2003; Lake, 1999, 2002). They need to be able to use formulas and understand the relationships between the variables, to develop an understanding of proportionality and to develop their understandings of relationships, of association and causality. They need to be able to make reasoned judgements, and come to valid conclusions based on their understanding of primary or secondary data. With appropriate teaching and learning strategies these attributes can be developed within the context of the subject area.

In social studies Frankenstein (1998, 1990, 2001) illustrates the need to think carefully and critically about the data with her examples of information
conveyed in tables and graphs. She points out that an author’s interpretation can mask the realities of the inequities between races, socio-economic status and genders unless the reader can interpret the data for themselves. Watson (1994), Best (2001) and Gal (2003) discuss the importance of being able to interpret data in the media such as newspapers or magazines. It is vital that students develop these skills and the context of the social sciences or media and communication studies is an ideal place to do so, given adequate guidance and opportunity for discussion. The students should become aware that sometimes arguments are based on biased interpretation of the data.

Statistics is a major area in which students are required to analyse data. They need to be able to make visual comparisons within and across graphs. They need to comprehend the essential differences between observed and projected data, and the validity of predictions. With tables they could be asked to extract simple numerical information or draw a graphic display. They might also be asked to conduct statistical tests to compare two or more data sets. For both graphs and tables they should consider the data in terms of the context and think critically about the ‘message’ of the data (Lake & Kemp, 2001).

These give just a few examples to show that it is likely that students will build their numeracy skills in the context of areas other than mathematics.
2.6 Reading and Interpreting Graphs and Tables

The critical numeracy in the various disciplines referred to in the previous Section, involves an understanding of the ways in which data is presented to communicate information to the reader. The initial collection, or production, of data is carried out to provide information to those people who are doing the research. Their first analysis of the data would normally include looking at some descriptive statistics such as the mean or median and perhaps the standard errors. They might compare totals and percentage changes and look at trends. People explore the data ‘...gaining an understanding of the data sources, inspecting the data for data integrity as well as unusual features that can guide the development of a formal analysis and staging the data for further analysis’ (Koschat, 2005).

The presentations of the results of an analysis or of selected data to an intended audience often include tables and graphs. The benefits of using graphical displays of information have been acknowledged in the literature for many years (e.g. Chambers, Cleveland, Kleiner & Tukey, 1983; Cleveland, 1994; Mosteller & Tukey, 1977; Tukey, 1977; Wainer, 1984, 1992, 1996). There is a dedicated major research journal, namely the Journal of Graphical and Computational Statistics, and articles in other journals such as Statistical Science, The American Statistician and Chance also include articles on graphical displays.
There is a rich literature investigating the ways in which graphical displays can be misleading, with or without intent, (e.g. Dewdney, 1993; Huff, 1997; Wainer, 1984, 1990) and it is common practice in introductory statistics courses at school or university for students to examine graphs and look for such aspects as incomplete, inaccurate or missing scales. It has been suggested by Wainer (1992) that ‘there is ample evidence that the ability to understand graphically presented material is hard-wired in, there is even more evidence that the ability to draw graphs well is not’ (p.18), even though there is no shortage of advice on how to construct good graphical displays (e.g. Tufte, 1983; Wainer, 1984, 1990).

There are two main kinds of tables that people encounter, both of which require numeracy skills to make sense of them. Firstly, there is the kind of table that is consulted to gain specific information. This is the kind of table referred to by Mosenthal and Kirsch (1998, p.639) as ‘reading-to-do’. These include tables like those with instructions for dosages of medicine depending on age or weight and those on food and health supplement packaging. These can be quite complex with a number of different kinds of units such as millilitres, grams, milligrams, kilojoules and calories, sometimes all in the same table. Similarly, there are bus, train and flight timetables of various kinds that may require some understanding of travel on specific day such as Sundays, and crossing of time zones. Reading and interpreting these kinds of tables is important. Although there can be some difficulties in reading these tables and people need to learn how to do so there is usually some
experience of this included in the secondary school curriculum (Kemp & Kissane, 1990).

Secondly, there is the kind of table that includes collected, produced and collated data that can be used to develop an understanding of the associated context or mental model. This is the kind of table referred to by Mosenthal and Kirsch (1998, p.639) as ‘reading-to-comprehend’. Examples of this kind of table include those with data on unemployment, traffic statistics, poverty, disease, and educational achievement that would be found in newspapers, journals and reports. These can include scientific or business data produced in specialised reports.

It is unlikely that someone would consult such tables simply for a particular value; they are more likely to want to make comparisons or look for trends over time. The reader needs to be able to decide what the data are about; to gauge the reliability of the data; to understand what the numbers represent; to look at changes over time and differences within and between categories, and to consider implications of the information in the light of prior knowledge. Many people skim over these kinds of tables, relying on the possibly biased interpretations of the authors of the accompanying text (Chapman, Kemp & Kissane, 1990).

The literature on the reading and interpretation of tables of data is much more limited than that the equivalent examination of graphical displays. The researcher made an extensive search of the literature, using computer
searches of databases such as ERIC, Proquest and Expanded Academic ASAP and examining related publications to come to the conclusion that the reading and interpretation of tables of data is only addressed to a minimal extent in the literature. There has been very little research reported regarding the development and improvement of expertise in dealing with tables of data.

This lack of research reflects the amount of attention given to these kinds of tables in the education systems. Twenty years ago Ehrenberg (1983) commented that ‘… statistics courses hardly ever teach students how to analyze and interpret tables’ (p.248). Very recently, Koschat (2005) asserts that this is still the case, ‘[T]he task of devising and interpreting tables is an integral part of statistical practice. Yet tables receive little attention as a topic of statistical research and statistical education’ (p.31). Koschat (2005) cites the early valuable work of Walker and Durost (1936) and comments on the lack of scholarly works on reading tables in statistical education since then, saying that major work in this field has been limited to that of Ehrenberg (e.g.1981, 1986), Wainer (e.g. 1990, 1992) and Tufte (e.g. 1983).

However important contributions to the field of tables have been made by the work of Kirsch and Mosenthal. Their work in the 1980’s on three interrelated constructs Prose/Document/Quantitative Literacy informed the large-scale NAAL study in the USA (OECD, 2003) and other major studies. Of particular relevance in this context is their work on document literacy (1989a, 1989b, 1990a, 1990b, 1990c, 1990d, 1990e, 1990f) which classifies tables
according to their inner structure and organization and which is dealt with in
detail in their IKIRSCH/PMOSE readability analysis (Mosenthal &

Their readability analysis is based on the organisational pattern and density
of the tables. Both of these are assigned levels and according to these the
table would be assigned a score as a whole. They accord four levels of lists:
simple, combined, intersected and nested. Simple lists are like shopping lists
as a list of objects, for example groceries. Combined lists comprise two or
more simple lists, each with a label at the top, for example this could be a list
of groceries next to a list of vegetables. Intersecting lists are a more complex
form of combined lists with grouped simple lists within the table and with
the first column incorporating categories of some sort. Mosenthal and Kirsch
suggest this could be concerned with fast foods as the categories, with the
columns comprising the lists of money spent on them in three countries
(1998, p.644). Nested lists would add to the previous intersected list the
money spent by gender within the costs (1998, p.646). A list is assigned a
document structure score difficulty from 1 to 4 from simple to nested.

Mosenthal and Kirsch (1998) also consider graphics documents which
essentially are based on lists of data. They assign a document structure score
of 2 to a pie chart. For bar charts there is a document structure score of 3 for
single bars and a score of 4 when there is more than one bar for a single label.
For line graphs the scores 3 or 4 depending on complexity and for time lines
there is a score of 2.
The density of a list (or table) is determined by the number of labels and items that make up the list. As the numbers of items and labels increases, so does the complexity of the list. The density is calculated according to a formula. Once the document structure score and the density have been calculated any dependency is considered and factored in.

The purpose of the development of the PMOSE/IKIRSCH document readability formula is two-fold. It enables the difficulty of a list (table) or graph to be determined and some comparisons can be made between documents. ‘…policy makers may not be so quick to decry U.S. adults inability to read a bus schedule with a readability value of 15!’ (1998, p.655). As well as this an ‘understanding of this formula also provides teachers with an important basis for knowing how to teach the structures of documents in an integrated fashion for students of all ages (1998, p.655).

Koschat (2005) points out three main advantages in using tables for providing information. Firstly a table presents data in their original form, or a summary of that data. Secondly, people can easily use the data and convert to other forms such as a graph or a model if they wish to do so, but this is not easy in reverse. Thirdly, it can quite often be the case that the reader wants to see and interpret the actual numbers. Therefore table reading skills are useful in three contexts.

As indicated earlier many people tend to skim over quantitative information and especially that encountered in tabular form and indeed Ehrenberg (1981)
points out that often tables are often hard to understand because they have not been well constructed. The reconstruction of a table can make a big difference to the readability and understanding of the table.). In similar words Ehrenberg (1981), Koschat (2005) and Wainer (1992) suggest that the rows and columns should be put in order that makes sense, making use of marginal averages or totals; the numbers to be compared should be put in columns with the highest at the top; rounding should be done to two effective digits and the layout should facilitate comparisons.

Some guidelines on reading a table of an informal nature were given by Wainer (1986). These are based on reflections of his own practice in reading a table without text for a colleague:

I Take in the broad subject matter and the variables, without yet worrying over details, sources etc.

II Focus first on one row and/or one column, preferably of averages. Establish the range of variation, i.e. the highest and lowest of the readings as “mental markers”. (Also note what form the intervening variation seems to take without yet trying to take the data in fully)

III Round all figures one looks at to one or two effective digits in one’s head, to facilitate mental arithmetic and make the results more memorable.

IV Compare the detailed readings in the body of the table against these patterns as norms.

V Now possibly consider the definitions, sources, the wider meaning of the results, and a more formal analysis (p.243-244)

Kissane (1991a, 1991b) raised the issue of the importance of teaching students how to read tables of data. He highlighted some of the difficulties that students have in responding to questions about tables of data. In Rubenstein
et al. (1992), a textbook for high school students of which Kissane was one of the authors, students are asked to think about the reading of a table of data concerning poverty in the USA in terms of *What is being presented?* (students need to identify the variables and the units of measurement), *Is the data trustworthy?* (they consider the source and kind of data presented) and *What conclusions can be drawn from the table?* (they look for patterns and try to form reasonable generalisations) (p.10). These questions form a guide to structure students' interpretation of the table of data. A model answer is provided but there is not a lot of detail in the text so one would need to infer that discussion would take place in the classroom for this to be a fruitful task.

Recently (during 2004) the Australian Bureau of Statistics has started developing a very useful website which provides advice and sample lessons for teachers. The lesson plans suggest readings produced by the Australian Bureau of Statistics suitable for various topics and school levels. These papers are easily accessible for teachers. Some of the lesson plans suggest that the students should examine a particular table of data but they only suggest a small number of questions designed to help students to develop an overview of the message of the table. For example, for a very complex table on Victimisation Rates of Selected Crimes (ABS, 2004), with nine columns and thirty rows of numbers, there are three questions on the worksheet for the students: *Which age groups reported the lowest and highest rates for assault? Which age groups reported the lowest and highest rates for robbery? What issues could be taken into account when investigating this data?* and then the students
are to Write a short paragraph comparing crime rates for age groups by sex for the different crimes. It is unrealistic to expect students to be able to write a paragraph after such a short investigation into the data. They might reasonably be a useful part of an examination of the table, but a more detailed investigation would be necessary before students would be able to write a paragraph. It would be helpful if teacher notes would make this clear.

Wainer (1992) was concerned with test items incorporating the interpretation of tables of data in public examinations where students are required to answer questions about specific entries and then perhaps do some algebraic manipulations. He observes that this constitutes a requirement for a number of low-level responses without any development to a higher level of inference. Wainer suggests that reading a table has three levels. Based on the work of Bertin (1973) and Peirce (1891), the levels are similar to those of reading a graph: 'we extract single bits of information (firstness); we look for trends and groupings (secondness); and we make comparisons (thirdness)' (p.18). However he focuses not on teaching students how to read tables but on the ways that tables can be constructed to display even complex information to make interpretation easier and more effective. He acknowledges the work of Ehrenberg (1977) in this regard. Koschat (2005) also emphasises the need to present data in tables in such a way that the reader can make sense of the data.

The work done by Lake (1999, 2002) mentioned earlier in connection with critical numeracy focuses on ways of building students' ability to interpret
graphs and tables. His papers suggest a series of hierarchical questions, based on the SOLO taxonomy, from low-level to high-level, to help students to develop higher-level skills of interpreting graphs and tables of data.

This Section has given an overview of the limited literature about how to read tables. Given the importance of this attribute of numeracy the researcher noted with concern the lack of evidence in the literature of the effectiveness of ways to enable students to develop their table reading and interpretation skills to include higher levels of cognitive responses than simply picking out individual values or simple trends.

Therefore an empirical study to conduct a study to test the effectiveness of a workshop specifically designed to improve table interpretation skills was proposed by the researcher. This study, reported in this dissertation, would include a workshop conducted with a number of groups of students, and the effectiveness of it would be evaluated through the use of before and after workshop table interpretation tasks.

The aim of the workshop was to make students aware of strategies that can be used to reach a good understanding of what data in a table indicates about the relationships between the variables within the table, and the need to consider these in the in a broader real-world context.

There are some reported studies of students' responses to tasks where the assessment coding rubrics (e.g. Bell 1976; Pegg, 1997; Guthrie, Anderson, Alao, Rinehart, 1999; Watson & Callingham, 2003) have been based on
identifying the level of response to various tasks, but unfortunately none of these seem to include interpretation of tables of data. Therefore, a suitable rubric, or measuring instrument, had to be designed to enable the comparison between the tasks performed before and after the workshop. The workshop incorporating a Five Step Framework and the measuring instrument to be both based on the SOLO taxonomy and the work of Lake (1999, 2002) and Kemp & Lake (2001).

The Rasch analysis would be used to put the levels of the students’ responses and difficulties of the tasks on the same continuum, thus putting ordinal data onto an interval scale. T-test would be used to compare the performances using the data from the continuum.

Through the study the researcher plans to make a contribution to the literature in this area of developing table interpretation skills.

2.7 Summary

This chapter developed a conceptualisation of numeracy using a range of views from mathematics educators. In considering the development of numeracy three perspectives from the literature were examined and discussed: mathematics skills in the mathematics classroom, mathematics for everyday life taught in the mathematics classroom, and mathematical concepts and skills developed and used across the curriculum.
It was proposed that there are two major components that can productively contribute to the development of numeracy. Firstly it was suggested that the development of mathematical concepts and skills, in appropriate environments with suitable activities, is essential to the development of numeracy. Secondly, that students need to develop cognitive skills through activities explicitly designed to do so.

The reading and interpretation of tables of data as a specific aspect of numeracy was examined. The limited extent of research in this area was highlighted and contribution of the empirical study on reading tables of data reported in this dissertation was foreshadowed.
CHAPTER 3  THE EMPIRICAL STUDY

3.1 Introduction

There were two main interrelated aims for this empirical study. The first aim was to investigate whether students' table reading and interpretation skills could be improved through a specially designed workshop. The second was to develop and refine an instrument to measure students’ responses on table interpretation tasks skills. This instrument would be available for measuring an individual’s performance and to gauge the effectiveness of the workshop.

In this study the focus was on reading and interpreting the kinds of tables of data that students would encounter in their study and everyday lives. The one-hour length of the workshop was chosen for both pedagogical and practical reasons. Whilst it would be certainly be considered more productive to give students a more extensive experience in reading a whole range of complexities of tables, one workshop would be the most likely maximum amount time that would be allotted to such a learning experience within a course outside mathematics and statistics courses. Therefore the researcher was resolved to investigate whether such a workshop could be shown to make a difference to students’ cognitive levels of response. If so, then such a model might be incorporated in other courses.

A comparison of the students’ responses, the refinement of the measuring instrument and the evaluation of the workshop involve the use of the Rasch
(1960) model to convert ordinal responses onto an interval scale continuum and the use of dependent T-test to make comparisons using that continuum.

The study was conducted with undergraduate students at Murdoch University in Western Australia enrolled degrees outside of the mathematics and statistic faculty. The selection of this group was decided on the basis that they needed to interpret tables of data in the course of their studies and would not normally be given explicit instruction on table reading.

Workshops for all of the students, in groups of about 20, were conducted by the researcher. The workshops were to be conducted in the place of a scheduled workshop in an education course that the students were studying, not as extra workshops outside of the course, so that they would not be seen as ‘add ons’ by the students. This one hour intervention was possible only at the researcher’s own university due to practical constraints. The design of the workshops was based on the literature in Chapter 2. Even though most of the literature reviewed concerned primary and secondary learning the principles were seen as appropriate and applicable to the learning of these students. Control groups were not used because of the nature of the intervention.

These students encounter quantitative material in tables, graphs and prose in lectures, tutorials and readings. Discussion with a wide range of academic staff across the university indicated that in non-mathematical subjects students were usually expected to learn how to interpret quantitative material on their own, without explicit teaching included in the curriculum.
An exception to this is found in Murdoch University’s Foundation Units. These units combine the study of interdisciplinary content with the development of learning skills. Each foundation unit tutorial has a balance of both content and learning skills, and a numeracy tutorial is included as one of the learning skills tutorials. The numeracy tutorial is made up of a variety of tasks including interpretations of tables, graphs and charts, and the consideration of numerical information in text. These tutorials are designed to encourage students to engage with quantitative material found in their readings and research across the disciplines areas.

A review of the Foundation Units in 1996 offered the opportunity to adapt the learning skills tutorials to relate them to new content and readings for the units. The tutorials were modified by the researcher to include the development of strategies, based on the SOLO taxonomy, which were aimed at improving students’ interpretation of graphs and tables. However it was felt that further work in the area was necessary to evaluate such changes.

3.2 A Pilot Study

Concern about the need to develop students’ ability and disposition to understand quantitative material in the Foundation Units led to a pilot study to develop a set of criteria to evaluate students’ responses to tasks involving the interpretation of tables of data.

This pilot study was set up after discussion with the lecturer planning a new interdisciplinary first year unit, based in the Institute for Sustainability and
Technology Policy (ISTP) at Murdoch University who indicated that he was concerned about students’ ability to read and interpret the quantitative material in tables and graphs. Liaison with the coordinator led to the integration of a numeracy workshop into the unit and collaboration between the coordinator and the researcher led to the design of an assignment involving interpretation of a table of data.

In the pilot study, the researcher conducted numeracy workshops on how to interpret a table of data and a graph and evaluated the students’ responses to tasks where they interpreted tables completed a few weeks after the workshop. The workshop was based on workshops conducted in a team teaching situation with Lake and based on his work (Lake, 1999). There were about 60 students involved in the pilot study. The students’ responses were allocated to levels according to responses for *point, trend, relational* and *message* interpretations based on the SOLO taxonomy. In addition, several students were interviewed and positive feedback was received from them and the coordinator. As a result of this work, it became evident that a well-constructed measuring instrument was needed to evaluate student responses to table interpretation tasks. Such a measuring instrument could then be used as a diagnostic tool to identify student needs and to enable the evaluation of the effectiveness of numeracy workshops, in terms of student learning.
3.3 The Aims of the Study

The pilot study focused on developing and measuring students’ table and graph reading skills. For the main study it was decided to concentrate on table reading skills given that there was only an hour available for the workshop due to practical constraints.

The two main aims for this empirical study:

1. To develop and refine an instrument to measure students’ responses on table interpretation tasks skills, and

2. To investigate whether students’ table reading and interpretation skills could be improved through a specially designed workshop.

The measuring instrument and the workshop were both developed based on the hierarchical SOLO taxonomy. The researcher felt that although there are instruments used to measure numeracy, even though this is often measured as maths skills) and applications of mathematics (PISA, TIMMS, NAAL) it was important in this case to link the teaching and the measurement to the same structure.

From a pedagogical perspective an investigation about how teaching can help to develop table interpretation skills should include a series of workshops with reflection on the processes of interpreting tables of various kinds. Ideally these would be integrated into students’ credit bearing units so they can see the relevance of developing such skills in the course of their
study. However, while a unit coordinator was prepared to allow some time for an existing unit to be used for the purpose of this study, a larger allocation of time was not available in the authentic context of an existing unit. Therefore it was only possible to conduct a single one-hour workshop for the students in the study.

A set of objectives were developed to achieve the two aims:

1.1 To construct a measuring instrument to assess levels of students’ responses to table interpretation tasks

1.2 To use the Rasch Model to refine the measurement instrument and to place students’ levels of response and the difficulty of the items onto a continuum

1.3 To measure change in the students’ levels of responses to the table interpretation tasks before and after the workshop using dependent t-tests.

2.1 To develop and conduct a one-hour intervention workshop to improve students’ reading and interpretation of tables of data using a Five Steps Framework

2.2 To design table interpretation tasks to be given to the students to complete before and after the workshop

While the generalisation of the study is limited because it is conducted on only first year students from one university, the associated design and implementation issues allow for the workshop concept to be carefully studied with a view to informing other studies of this kind.
3.4 The Subjects

The study was conducted over three years from 2000 to 2003. The data for 265 students who completed table interpretation tasks both before and after a workshop were used in comparison of performances. All of these students were first year undergraduate, pre-service primary teachers enrolled in a Bachelor of Education degree. They included male and female students, with both school leavers and mature-aged students being represented.

The students come from a wide range of career and mathematical backgrounds and many of them come to university ill-equipped to critically interpret the kinds of quantitative information that they encounter in their studies and their everyday lives as highlighted in Chapter 2. These students are studying to be primary school teachers and the curriculum they cover at university is diverse, spread across many subject areas. The development of critical numeracy skills is important to them for their study, for their future careers as teachers where they will influence their students’ attitudes, and in their own lives as citizens.

Whilst participating in the study the pre-service education students who attended workshops were enrolled in one of two of the first year units: *Introduction to Science* or *Cultural Mathematics*. The coordinators of these two units were enthusiastic about integrating this kind of workshop into their units and were amenable to the inclusion of one *intervention* workshop as an initial step. Since it required some adjustment of the existing teaching
program to accommodate the workshop it was not possible to have more
than one workshop in units of this kind. During the course of the study the
researcher conducted all of the workshops.

3.4 The Workshop

The one-hour workshop was designed to help students develop the
confidence and skills to extract information from tables of data, and to think
critically about any implications associated with the data. It introduced
students to a set of practical strategies, or steps, for reading and critically
interpreting tables. Although the workshop focuses on reading tables of data,
the strategies were designed to be also appropriate to the interpretation of
graphs. Although this was a single one-hour workshop for each individual
student, there were many students involved in the study each of whom
participated in the workshop, thus there were multiple opportunities to
gauge student involvement and responses to the material.

The development of the steps was based on the SOLO taxonomy developed
by Biggs and Collis (1982) and on the work of Lake (1999). These steps help
students to build up their understanding of a table by firstly looking at the
context and reliability of the table, then at the meaning of the numbers in the
table, and then comparisons between various values, usually maximum and
minimum. These may include changes in values over time for two or more
years, which will be referred to as trends in this thesis. Then these trends or
comparisons within or across categories are compared so as to build up a
better understanding of the variations in the data. Finally an analysis of the
table in the light of societal and environmental issues can lead to a deeper
understanding of the implications of the data presented in the tables.

The effectiveness of the workshop was to be measured by comparing the
levels of students’ responses on table interpretation tasks to be undertaken
before and after the workshop. The workshop was designed to develop the
students’ levels of thinking skills and to help achieve this aim the workshop
incorporated components of the five categories identified by Adey and
Shayer (1994) and described in Chapter 2, Section 2.4.2.1. These categories
describe important characteristics of programmes designed for cognitive
acceleration. It was not possible to integrate all the characteristics
comprehensively given the one-off intervention workshop format, but the
ways in which they were incorporated are described below.

To incorporate the characteristic, concrete preparation, involved setting the
scene and identifying language aspects. At the beginning of the workshop
there was general discussion about the importance of reading and
understanding the tables of data in their study materials and the media. It
was highlighted that an author’s interpretation of data can be deliberately or
accidentally misleading and that it is important that people are able to
interpret the data themselves and draw their own conclusions. Examples
were given of such cases from the media and students contributed their own.
Concrete preparation emphasises the introduction, or revision of appropriate language. Students used terms that describe the location of data in the table such as title, rows, columns and footnotes. They located maximum and minimum percentages or numbers. They made comparisons using words such as smaller, larger, decrease and increase, change over time, and trend.

To incorporate the characteristic, construction, of the intervention the students were given questions to guide their thinking, but they were encouraged to consider other related questions. Through this activity they were constructing their own interpretation of the table of data. They worked in small groups and were encouraged to discuss their answers to the questions included in each step, and to debate the issues surrounding the context of the data in the table.

As part of construction, cognitive conflict was initiated by the researcher who highlighted differences between different individual and group responses leading and encouraging students to sharing and comparing their answers. In addition, students’ comments would prompt others to disagree with them and subsequent discussion highlighted various issues with the potential to evoke a conflict of ideas. From the beginning of the activity students were urged to look for anything they found surprising in the table of data, and to share their findings with the students in their group. Plenary discussion in the middle and at the end of the activity enabled all students to share their responses to the questions and to think about the implications of the data in the table.
In this workshop the development of the characteristic of bridging started with students talking about the places in which tables of data are found in the media and in study materials and published journals. It was emphasised that the aim of the workshop was to introduce the students to a set of strategies that could be used to assist in the interpretation of tables in a whole range of contexts and situations. Thus the potential transferability of the skills developed through the use of the strategies was made explicit. The bridging component encouraged students to engage with tables of data in other settings.

In the part of the workshop, concerned with the development of the characteristic of metacognition, students were encouraged to think about, and comment on, the strategies that had been used in the workshop. They were given a handout that identified the kinds of activities involved with each of the five steps. Given more time it would have been more desirable for the students to reflect on the process of the steps and to construct their own checklist that could be refined through discussion in their groups.

To incorporate the characteristic, duration and density, required that the intervention is located within the curriculum, interweaving procedural knowledge with content. Ideally the intervention would take place over about two years, so that some of the lessons are dedicated to the kinds of work that accelerate cognition. In this case the time frame was very short but the workshop was conducted as an integral part of an introductory
mathematics or science unit. In this way students would not be expected to view the workshop as an ‘add on’, but rather as an essential part of the unit.

In the pilot study the researcher used the terms: *setting the scene, point interpretation, trend interpretation, relational interpretation* and *extended abstract* in the workshop and subsequently evaluated the responses accordingly. For the main study these terms were renamed to clarify the purpose of the steps as shown below:

**Five Step Framework**

**Step 1: Getting started**

This involved looking at the title, headings, footnotes and source to ascertain the context and expected reliability of the data.

**Step 2: WHAT do the numbers mean?**

In this step students make sure that they understand precisely what the numbers mean. They look for other values including the largest and smallest values in one or more categories or years to become familiar with the data.

**Step 3: HOW do they change or differ**

In this step students are encouraged to consider differences in data in a single row or column. Often this involves looking at changes over time but it can also include comparisons within categories, such as male or female, at any given time.

**Step 4: WHERE are the differences?**
In this step students are using their knowledge about categories or trends to investigate relationships within the table. They can make comparison between different trends in columns or rows.

**Step 5: WHY do they change?**

In this step students reflect on the implications of the variations in the data. They are prompted to consider reasons why there are variations, and what may have influenced these variations. They consider the effects of international, national or local policies and look for the potential impacts of social, environmental and economic factors.

The workshop activity involved students reading and interpreting the table of data shown in Table 3.1, which was concerned with risk factors of smoking, obesity, alcohol consumption and level of exercise related to gender and age. This table fits Mosenthal and Kirsch’s (1998) definition of an tested list.

Table 3.1 Health Risk Factors related to socio-economic status

<table>
<thead>
<tr>
<th>Quintile of disadvantage</th>
<th>Male  %</th>
<th>Female %</th>
<th>Males  %</th>
<th>Female %</th>
<th>Male  %</th>
<th>Female %</th>
<th>Male  %</th>
<th>Female %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st (most disadv)</td>
<td>35.9</td>
<td>27.1</td>
<td>7.3</td>
<td>3.5</td>
<td>44.4</td>
<td>33.6</td>
<td>35.8</td>
<td>37.7</td>
</tr>
<tr>
<td>2nd</td>
<td>30.9</td>
<td>22.6</td>
<td>4.4</td>
<td>2.4</td>
<td>47.1</td>
<td>33.6</td>
<td>35</td>
<td>34.6</td>
</tr>
<tr>
<td>3rd</td>
<td>27.2</td>
<td>20.1</td>
<td>5.7</td>
<td>2.6</td>
<td>44.9</td>
<td>32.1</td>
<td>35.1</td>
<td>32.7</td>
</tr>
<tr>
<td>4th</td>
<td>25.6</td>
<td>20.0</td>
<td>5.2</td>
<td>2.8</td>
<td>48.3</td>
<td>30.8</td>
<td>33.5</td>
<td>33.4</td>
</tr>
<tr>
<td>5th (least disadv)</td>
<td>18.2</td>
<td>14.8</td>
<td>4.5</td>
<td>4.0</td>
<td>46.0</td>
<td>27.4</td>
<td>25.8</td>
<td>27.9</td>
</tr>
<tr>
<td>Average</td>
<td><strong>27.0</strong></td>
<td><strong>20.6</strong></td>
<td><strong>5.4</strong></td>
<td><strong>3.1</strong></td>
<td><strong>46.2</strong></td>
<td><strong>31.2</strong></td>
<td><strong>32.7</strong></td>
<td><strong>33.1</strong></td>
</tr>
</tbody>
</table>

(a) age standardised rates
(b) For smoker status and alcohol consumption, data relate to the population aged 18 years and over. For weight and exercise, data relate to the population aged 15 years and over.
The steps and specific questions for this table of data are included in Table 3.2. These were supplemented by discussion in the workshop.

Table 3.2 Steps to Interpret a Table of Data

<table>
<thead>
<tr>
<th>Step 1: Getting organised</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Topic</strong></td>
</tr>
<tr>
<td>Q: From the title, what is the general topic being examined?</td>
</tr>
<tr>
<td><strong>Scope</strong></td>
</tr>
<tr>
<td>Q: From the title, what do all the groups being studied have in common?</td>
</tr>
<tr>
<td><strong>Comparisons</strong></td>
</tr>
<tr>
<td>Q: From the labels on the left column, how are the groups being compared?</td>
</tr>
<tr>
<td>Q: From the labels on the top row, how are the groups being compared?</td>
</tr>
<tr>
<td><strong>Reliability</strong></td>
</tr>
<tr>
<td>Q: From the footers, what evidence is there that the information is reliable?</td>
</tr>
<tr>
<td><strong>Definition</strong></td>
</tr>
<tr>
<td>Q: In this table what are the meanings of the following: socio-economic disadvantage, and quintile.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Step 2: WHAT do the numbers mean?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q: What is the highest average risk factor for females?</td>
</tr>
<tr>
<td>Q: What is the lowest average risk factor for females?</td>
</tr>
<tr>
<td>Q: Which socio-economic group of men has the most smokers?</td>
</tr>
<tr>
<td>Q: Which socio-economic group of men has the fewest smokers?</td>
</tr>
<tr>
<td>Q: Which socio-economic group of women has the most non-exercisers?</td>
</tr>
<tr>
<td>Q: Which socio-economic group of women has the fewest non-exercisers?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Step 3: HOW do they change?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q: How does the proportion of non-exercising men change with socio-economic advantage?</td>
</tr>
<tr>
<td>Q: How does the proportion of women smokers change with socio-economic advantage?</td>
</tr>
<tr>
<td>Q: How does the proportion of overweight men change with socio-economic advantage?</td>
</tr>
<tr>
<td>Q: How does the proportion of moderate to heavy drinking change with socio-economic advantage?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Step 4: WHERE are the differences?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q: Which are most likely to be the most important risk factors in determining health?</td>
</tr>
<tr>
<td>Q: Who are more seriously at risk men or women?</td>
</tr>
<tr>
<td>Q: Who are less seriously at risk the more or less socio-economically disadvantaged?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Step 5: WHY do they change?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Are you being manipulated?</strong></td>
</tr>
<tr>
<td>Is there any reason to suspect that current data would be any different?</td>
</tr>
<tr>
<td>Would the results be the same if the survey had been done 50 years ago or 50 years hence?</td>
</tr>
<tr>
<td>What additional data would you require to believe this information?</td>
</tr>
<tr>
<td>How valid are the data as indicators of what is happening in this society?</td>
</tr>
<tr>
<td><strong>How well do our conclusions fit with related data and theories?</strong></td>
</tr>
<tr>
<td>Alcoholism is an escape from poverty; overweight tycoons.</td>
</tr>
<tr>
<td>Healthy blue-collar workers; hypochondria amongst yuppies; gender stereotypes and fitness.</td>
</tr>
<tr>
<td><strong>What are the implications?</strong></td>
</tr>
<tr>
<td>What personal beliefs would I reconsider on the basis of the evidence?</td>
</tr>
<tr>
<td>-------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Is the current health system equitable; is it sensible or even equitable for the government to promote private health insurance?</td>
</tr>
<tr>
<td>Should money be going into men’s or women’s health?</td>
</tr>
</tbody>
</table>

### 3.5 The Table Interpretation Tasks

The three table interpretation tasks were chosen to be typical of the kinds of tables students might encounter in their study. They were chosen to be of approximately the same complexity to accommodate the pre-post test style of study. Although two were nested lists (table of document level 3) and one was an intersected list (table of document level 4) (Mosenthal & Kirsch, 1998) this intersected list included more items than the nested lists. They were all selected from the Australian Bureau of Statistics as it was considered reliable and accessible. Any difference in difficulty would be accounted for through the Rasch analysis which would order tasks and students on the same continuum.

The students were required to complete two table interpretation tasks, one before and one after the workshop, so that a comparison of the levels of responses could be made leading to an evaluation of the effectiveness of the workshop. The tasks required students to look at a table of data and write down their responses to a set of three subtasks. It was felt that a request such as ‘what the table tell you?’ was too open for these students and so subtasks (described below) were designed.
The students were given a worksheet with the table and subtasks, with spaces for their responses and given ten minutes in a supervised environment to give individual written responses to the task. About three weeks after the workshop a similar task with a different table was given to the students to complete under supervised conditions.

It was important to incorporate more than one table of data into the table interpretation tasks, as people would be required to complete a task on two occasions. Without some variety they could become bored and therefore not sufficiently motivated to respond fully. There would also be the danger that students would simply repeat previous responses if they were given the same table to interpret.

Three tables were chosen to be of approximately equal complexity in regards to the number of categories and the extent of data involved, but to measure any differences in difficulty of the tasks the relative difficulties were assessed using the Rasch Model for data analysis described in Chapter 4. Discussion concerning possible reasons for the differences in difficulty as found by the Rasch analysis in Chapter 8 highlights possible reasons why the intersected list was harder than the nested lists.

The tables were chosen as being of relevance to the students and in contexts about which they would be likely to have some knowledge. The three tables for the Homeworkers, Suicide and Environmental Concerns tasks were taken from publications of the Australian Bureau of Statistics.
### Table 3.3 Homeworkers Table and Text Material

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Managers and administrators</td>
<td>16.0</td>
<td>10.2</td>
<td>26.2</td>
<td>15.0</td>
<td>9.2</td>
<td>24.2</td>
</tr>
<tr>
<td>Professionals</td>
<td>25.3</td>
<td>19.9</td>
<td>45.3</td>
<td>38.3</td>
<td>28.8</td>
<td>67.1</td>
</tr>
<tr>
<td>Para-professionals</td>
<td>4.7</td>
<td>2.7</td>
<td>7.4</td>
<td>6.0</td>
<td>1.6</td>
<td>7.6</td>
</tr>
<tr>
<td>Tradespersons</td>
<td>18.7</td>
<td>12.6</td>
<td>31.3</td>
<td>24.3</td>
<td>13.3</td>
<td>37.7</td>
</tr>
<tr>
<td>Clerks</td>
<td>3.0</td>
<td>104.9</td>
<td>107.9</td>
<td>4.9</td>
<td>127.4</td>
<td>132.3</td>
</tr>
<tr>
<td>Salespersons and personal service workers</td>
<td>7.1</td>
<td>20.3</td>
<td>27.4</td>
<td>14.0</td>
<td>34.6</td>
<td>48.6</td>
</tr>
<tr>
<td>Plant and machine operators, and drivers</td>
<td>1.7</td>
<td>7.6</td>
<td>9.3</td>
<td>3.8</td>
<td>6.6</td>
<td>10.4</td>
</tr>
<tr>
<td>Labourers and related workers</td>
<td>3.7</td>
<td>8.0</td>
<td>11.7</td>
<td>6.3</td>
<td>9.2</td>
<td>15.5</td>
</tr>
</tbody>
</table>

Taken From ABS Persons Employed at Home 6275.0 September 1995

The Australian Bureau of Statistics collects information on employed persons who work at home through the Survey of Persons Employed at Home. People who work any hours at home are included in the survey, but only those who usually work more hours at home than elsewhere, in their main or second job, are classified as employed at home. Here they are referred to as ‘Homeworkers’. Farmers, unpaid voluntary workers, people who work less than one hour and people who work from home but spend most of their working time away from home, are not counted as home workers.

The tasks Homeworkers 1 & 2 are both based on Table 3.3 about people working from home. The worksheet for the task included the table and some text material.

It can be seen in Table 3.3 that students need to understand that the table is concerned with those people who work from home as their primary place of work. There are male, female and persons categories, where the category of persons includes the sum of the male and female figures, all given in thousands of people. The kinds of occupations are listed as categories, and students need to think about the kinds of work that categories such as plant and machine operators might include when people working from home are classified in those areas. The data for two years, April 1989 and September 1995, which may include seasonal variation as well as a change over time.
Table 3.5 Methods of Suicide Table and Text Material


<table>
<thead>
<tr>
<th>Leading Methods of Suicide, By Sex 1988 and 1997</th>
<th>1988</th>
<th>1997</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Males %</td>
<td>Females %</td>
</tr>
<tr>
<td>Poisoning by solid &amp; liquid</td>
<td>12.2</td>
<td>40.0</td>
</tr>
<tr>
<td>Other gases and vapours</td>
<td>20.7</td>
<td>14.5</td>
</tr>
<tr>
<td>Hanging and strangulation</td>
<td>24.5</td>
<td>19.5</td>
</tr>
<tr>
<td>Firearms and explosives</td>
<td>28.3</td>
<td>6.9</td>
</tr>
<tr>
<td>Other specified methods</td>
<td>5.6</td>
<td>6.2</td>
</tr>
<tr>
<td>Other methods</td>
<td>8.7</td>
<td>12.9</td>
</tr>
<tr>
<td>All methods</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>Total suicides</td>
<td>1 730</td>
<td>467</td>
</tr>
</tbody>
</table>

It can be seen in Table 3.5 that similarly to Table 3.4 there are also several components to this table that need to be understood. Again, there are males, females and person categories, but for this table the numbers with the methods categories are given as percentages. These percentages add up to 100% as they include categories such as *other specified methods* and *other methods* to ensure that this is the case. The *Total* is given as a number of people. As with the *Homeworkers* table the data are given for two different years, but this time the months are not specified.

The tasks *Environmental Concerns 1 & 2* are based on Table 3.4 concerned with attitudes of people about particular environmental issues. The students were given some text and the table.
Table 3.4 *Environmental Concerns* Table and Text material

People’s views and practices reflect the importance and priority given to environmental issues by society. The Australian Bureau of Statistics has conducted a household survey on environmental issues biennially since 1992. The reported levels of concern by Australians about particular environmental problems are summarised in the table.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Air pollution</td>
<td>40.2</td>
<td>34.1</td>
<td>30.9</td>
<td>32.4</td>
</tr>
<tr>
<td>Ocean pollution</td>
<td>32.3</td>
<td>26.7</td>
<td>23.8</td>
<td>24.1</td>
</tr>
<tr>
<td>Freshwater pollution</td>
<td>29.9</td>
<td>25.5</td>
<td>23.7</td>
<td>26.7</td>
</tr>
<tr>
<td>Destruction of trees/ecosystems/deforestation</td>
<td>32.8</td>
<td>25.6</td>
<td>23.6</td>
<td>21.7</td>
</tr>
<tr>
<td>Garbage/rubbish disposal</td>
<td>22.9</td>
<td>15.7</td>
<td>14.0</td>
<td>14.7</td>
</tr>
<tr>
<td>Ozone layer</td>
<td>28.6</td>
<td>17.1</td>
<td>10.9</td>
<td>13.4</td>
</tr>
<tr>
<td>Destruction of animals/wildlife/extinction of species</td>
<td>19.3</td>
<td>13.3</td>
<td>9.1</td>
<td>9.6</td>
</tr>
<tr>
<td>Other pollution</td>
<td>14.1</td>
<td>9.1</td>
<td>8.8</td>
<td>4.5</td>
</tr>
<tr>
<td>Toxic chemical/hazardous waste</td>
<td>21.3</td>
<td>11.9</td>
<td>8.6</td>
<td>11.5</td>
</tr>
<tr>
<td>Nuclear testing/weapons</td>
<td>14.6</td>
<td>6.7</td>
<td>7.6</td>
<td>7.0</td>
</tr>
<tr>
<td>Greenhouse effect</td>
<td>17.2</td>
<td>8.8</td>
<td>6.3</td>
<td>10.1</td>
</tr>
<tr>
<td>Urban development/overpopulation</td>
<td>12.6</td>
<td>7.8</td>
<td>5.9</td>
<td>8.8</td>
</tr>
<tr>
<td>Other</td>
<td>5.8</td>
<td>5.7</td>
<td>5.5</td>
<td>3.9</td>
</tr>
<tr>
<td>Uranium mining/use/radioactive materials</td>
<td>8.5</td>
<td>3.6</td>
<td>5.1</td>
<td>5.4</td>
</tr>
<tr>
<td>Use of pesticides</td>
<td>13.7</td>
<td>7.0</td>
<td>4.2</td>
<td>6.4</td>
</tr>
</tbody>
</table>

Source: Environmental Issues: People’s Views and Practices (4602.0)

The table for Environmental Concerns is different from the previous tables in that it does not have gender differences. It has 15 categories of *issues of concern* and this time the data are given for four years, each two years apart.

The data are given as percentages that represent the proportion of people interviewed who cited that the particular category is of concern to them. Therefore, the percentages do not sum to 100% because people could express concern about more than one category.

The table interpretation tasks were aimed at finding out the levels of response that students could give to the tables of data. Students are likely to respond in depth to a table of data if motivated through interest, or the need to obtain information for their research for an assignment. In this study they were being asked to read a table as part of a task out of context of that kind.
of research. From one perspective, it was thought that the students should only be given a table of data and asked to interpret that table. However, it was felt that this would not really give students the feeling that they had the license to respond in sufficient depth for the analysis to reveal the level of analysis of which they were capable. Therefore, the following three subtasks were designed to invite responses, without providing too much guidance:

- Describe the main features of the table and how they are related.
- What conclusions can you draw from the table?
- Explain and discuss the information contained in the table.

However, after these subtasks had been used with the tables for Environmental Concerns and Suicide it was decided to change the third subtask. This was done because the researcher’s supervisor observed that the existing subtask did not appear to give students permission to go outside the data in the table to give their responses. So the third subtask was replaced by:

- What societal and environmental factors might account for the data in the table?

After this change the table interpretation tasks had two versions for each table of data: one with the original third subtask and the other with the new third subtask. This made six tasks in total, all which were completed during the study. Table 3.6 shows the descriptors and abbreviations used in some of the subsequent descriptions of the tasks:
Table 3.6 The Subtasks for the Table Interpretation Tasks

<table>
<thead>
<tr>
<th>Third Subtask</th>
<th>Environmental Concerns</th>
<th>Suicide</th>
<th>Homeworkers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Explain and discuss the information found in the table.</td>
<td>Environmental Concerns 1 (EC1)</td>
<td>Suicide 1 (S1)</td>
<td>Homeworkers 1 (HW1)</td>
</tr>
<tr>
<td>What societal and environmental factors might account for the data in this table?</td>
<td>Environmental Concerns 2 (EC2)</td>
<td>Suicide 2 (S2)</td>
<td>Homeworkers 2 (HW2)</td>
</tr>
</tbody>
</table>

All of the student responses to the tasks were allocated a level according to the SOLO taxonomy so that the results could be analysed quantitatively. The levels are in Chapter 5, and the analysis is reported in Chapters 5 and 6.

3.6 An Instrument for Measuring Numeracy

Table 3.7 shows the instrument designed for measuring the levels of students' responses to the table interpretation tasks, based on the levels of SOLO taxonomy described in Chapter 2, Section 2.4.2.2.

Ratings were based on a set of criteria shown in Table 3.7 that were developed to identify the levels of the students’ responses to the table interpretation tasks. It includes sublevels for multistructural, relational and extended abstract.

Table 3.7 Criteria for Allocation to Initial Levels

<table>
<thead>
<tr>
<th>Initial level</th>
<th>Coding</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prestructural Level 0</td>
<td>0</td>
<td>No responses, or all responses inappropriate or irrelevant.</td>
</tr>
<tr>
<td>Unistructural Level 1</td>
<td>1</td>
<td>Description of what the table is about including the headings or categories but with no additional information.</td>
</tr>
<tr>
<td>Multistructural Level 2a</td>
<td>2</td>
<td>Specific values used to make a comparison at a particular time (for example the highest or lowest value for a category), or one trend over time is described.</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Two trends, or one trend and a sub-trend, or a trend and a comparison, within a category, are described.</td>
</tr>
</tbody>
</table>
A possible reason for trend(s) or comparisons made above is included, possibly related to the media, social influences, changes in technology, or includes comments on sampling procedures. More than one reason for the comparisons or trends or one reason and a comment on the sampling are given.

An understanding of the relevance of the information to overriding abstract principles is shown, and an application to a situation that is not given in the table. This response takes in more global and philosophical views concerning the context.

Ratings were based on a set of criteria shown in Table 3.7 that were developed to identify the levels of the students’ responses to the table interpretation tasks. It includes sublevels for multistructural, relational and extended abstract.

The instrument was based on an assumption that the levels would all be identifiable in similar ways for all the tasks. After the Rasch Analysis, differences in these levels led to modification of the instrument to take into account the variations in the responses to the six tasks, as described in Chapter 5.

3.7 Coding of the Tasks for Analysis

Each student’s set of written responses to the three subtasks was taken as constituting a single response which was allocated to a level according to the criteria of the measuring instrument. Samples of students’ responses to the six tasks are provided in Appendix A and Appendix B. These levels had been clarified after the first marking of the tasks by two markers familiar with the taxonomy, so that inter-rate reliability was established. Each level was
allocated a code: 0 for prestructural, 1 for unistructural, 2a and 2b for the two levels of multistructural, 3a and 3b for the two levels of relational and 4a and 4b for the two levels of extended abstract. These were subsequently allocated numerical scores from 0 to 7 as shown in Table 3.7 and the results were entered into a spreadsheet.

Each student was also allocated a coding in the spreadsheet according to whether they completed a task before or after the workshop, and the gender of each student was recorded. The analysis of these results will be reported in Chapters 5 for the qualitative analysis and Chapter 6 for the numerical analysis.

Overall, there were nine groups of students who completed the six tasks. The first four groups of students were those enrolled in the first year units *E162: Cultural Mathematics* or *M113: Introduction to Science*. These are included in Table 3.8 as Group A & B: *M113*, Group C & D: *E162*. The aim was to compare the performance on their tasks before and after the workshop.

Table 3.8 Allocation of Tasks

<table>
<thead>
<tr>
<th>Students</th>
<th>Environ Concerns 1</th>
<th>Environ Concerns 2</th>
<th>Suicide 1</th>
<th>Suicide 2</th>
<th>Home-workers 1</th>
<th>Home-workers 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group A: <em>M113</em></td>
<td><em>Before</em> (1)</td>
<td><em>After</em> (4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group B: <em>M113</em></td>
<td><em>After</em> (5)</td>
<td></td>
<td></td>
<td></td>
<td><em>Before</em> (7)</td>
<td></td>
</tr>
<tr>
<td>Group C: <em>E162</em></td>
<td></td>
<td></td>
<td><em>Before</em> (3)</td>
<td></td>
<td></td>
<td><em>After</em> (9)</td>
</tr>
<tr>
<td>Group D: <em>E162</em></td>
<td><em>After</em> (2)</td>
<td><em>After</em> (6)</td>
<td></td>
<td></td>
<td><em>Before</em> (8)</td>
<td></td>
</tr>
</tbody>
</table>

The students were given the tasks to complete and the responses were coded and recorded. The tasks completed by the groups of students are shown in Table 3.8. *Before* indicates that the tasks were completed before the workshop, and *after* denotes that the tasks were completed after the
workshop. Given that the tasks were potentially of different difficulty the Rasch Analysis was used to order the difficulty of the tasks and the ability of the students on the same scale to make it possible to compare the students’ responses on tasks completed before and after a workshop.

In general for a Rasch analysis to work properly, and to make valid comparisons made between responses, it is necessary to have linking networks between all of the items, (Linacre, 1997). This means that responses to each item (task) can be joined via a network to the responses to every other item through its positioning in the rows and columns in a matrix. Responses in the same row or the same column can be linked together for this purpose. In this case, it can be seen in Table 3.8 that a network of ‘links’ can be drawn via between the tasks to form a network joining all tasks except for Environmental Concerns 2 and Homeworkers 2, which can be connected only to each other. In this table possible links can be described in terms of the numbers in the table allocated to the table tasks. For example, the set of links can be made (1) to (2) to (6) to (8) to (7) to (5) to (1) which does not include (3) and (9).

In other words, for an analysis to be possible, the distribution of the tasks in the table should not be such that the table cannot be divided into two or more mutually exclusive tables. It can be seen in Table 3.8 that the table could in fact be separated into two separate tables on each side of Suicide 2, which had no entries because it had not been given to students at this stage.
In order to complete the network so that all responses can be linked in the Rasch analysis, the tasks were completed by another group of 88 first year students completing the first year humanities unit used in the pilot study. A workshop was not provided for these students, so all their responses were coded as being before the workshop. In Table 3.9 they are entered as ‘No w/sh’ to differentiate between those who completed before and after, and those who completed only before, the workshop. The abbreviations EC1 for Environmental Concerns 1, and so on, are used in Table 3.9.

Table 3.9 Distribution of Tasks

<table>
<thead>
<tr>
<th>Students</th>
<th>EC1</th>
<th>EC2</th>
<th>S1</th>
<th>S2</th>
<th>HW1</th>
<th>HW2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group A: M113</td>
<td>Before (1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group B: M113</td>
<td></td>
<td>After (7)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group C: E162</td>
<td>Before (5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group D: E162</td>
<td>After (2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group E: S108</td>
<td>No w/sh (3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group F: S108</td>
<td></td>
<td>No w/sh (12)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group G: S108</td>
<td>No w/sh (6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group H: S108</td>
<td>No w/sh (4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group I: S108</td>
<td>No w/sh (11)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

It can be seen from Table 3.9 that each response can now be linked to the other responses through a network such as: (1) to (7) to (8) to (14) to (15) to (2) to (3) to (12) to (13) to (17) to (16) to (5) to (18) to (19) to (11) to (10) to (4). Thus this combination forms an appropriate structure for the Rasch analysis and comparisons of responses relative to other responses can be made.

The responses to the tasks labelled before and after were used to evaluate any change in the levels of response after, compared with before, the workshop.

The statistical analysis of the change in levels is given in Chapter 4.
3.8 Summary

The study was conducted as described and the data assembled for analysis. The Rasch analysis was used to place the students’ ability and the relative difficulty of the tasks on an equal interval scale. These locations were needed to be able to compare students’ responses before and after the workshop to evaluate both the quality of the measuring instrument and the effectiveness of the workshop. The analyses for these evaluations are reported in detail in Chapters 4, 5, 6 and 7.
CHAPTER 4 QUANTITATIVE ANALYSIS

4.1 Introduction

This Chapter describes the Rasch analysis of the data and refinement of the initial measuring instrument given in Table 3.7 in Chapter 3. The analysis uses the Rasch Model to map the relative locations of the tasks, and the thresholds for the different categories of those tasks, onto the same continuum as the location of the ability of the persons (Andrich, 1985). It transforms ordinal qualitative responses onto a linear continuum. This enables comparisons to be made between students’ responses to tasks of different difficulties. The Rasch Model used for unidimensional, polytomous, ordered qualitative response categories, is a generalisation of Rasch’s Simple Logistic Model (Andersen, Reder & Simon1977; Andrich, 1978; Rasch, 1960/1980). In this thesis both the terms task and item refer to the tasks involving the interpretation of tables of data undertaken by the students.

Section 4.2 gives a brief outline of the theory of the Rasch Model to highlight the probabilistic nature of the model. A more comprehensive account of this model can be found in the literature referred to in this section. For the purpose of this analysis, the RUMM (Andrich, Sheridan & Luo, 2002) software program was used. As part of this analysis, the structure of the operation of the hypothesized ordered categories was investigated. Section 4.3 summarises the statistics and rationale used in the analysis using the Rasch Model. The results of the analysis of the data using the initial
measuring instrument for the coding are reported in Section 4.4. The process of rescoring to refine the instrument is described in Section 4.5 and the results of the analysis on the rescored data are given in Section 4.6. A summary of the chapter is provided in Section 4.7.

4.2 The Rasch Model

The common Rasch Model, known also as the Simple Logistic Model (SLM) is concerned with dichotomous responses of persons to items. Substantial literature now exists on the application of the Rasch Model for research designs in the social sciences of the kind used in this study (Rasch, 1960/80, 1961; Andrich, 1978; Wright & Stone, 1979; Wright & Masters, 1982; Andrich, 1989; Bond & Fox, 2001; Watson & Callingham, 2003). The details of the model, therefore, will not be repeated in this thesis except as a brief summary.

In 1960 Rasch developed a probabilistic measurement model for social sciences incorporating, independently, Thurstone’s requirements of invariance, undimensionality and additivity for fundamental measurement. (Wright, 1997).

Rasch articulated these principles of invariance of comparisons as follows:

The comparison between two stimuli should be independent of which particular individuals were instrumental for the comparison; and it should also be independent of which other stimuli within the considered class were or might also have been compared.

Symmetrically, a comparison between two individuals should be independent of which particular stimuli within the class considered
were instrumental for comparison; and it should also be independent of which other individuals were compared, on the same or on some other occasion (Rasch, 1961, p. 332).

The model that arises from these requirements for items with dichotomous scoring is given by:

\[
\Pr\{X_{ni} = 1\} = \frac{e^{\beta_n - \delta_i}}{1 + e^{\beta_n - \delta_i}} \quad (1)
\]

\[
\Pr\{X_{ni} = 0\} = \frac{1}{1 + e^{\beta_n - \delta_i}} \quad (2)
\]

or in general

\[
\Pr\{X_{ni} = x\} = \frac{e^{x(\beta_n - \delta_i)}}{1 + e^{(\beta_n - \delta_i)}} \quad (3)
\]

where \(X_{ni} = x\) and \(x\) takes on the values 0 or 1, \(\beta_n\) is the location of the person \(n\) on a continuum, and \(\delta_i\) is the location of item \(i\) on the same continuum.

The locations \(\beta_n\) and \(\delta_i\) are on the same continuum and these can be estimated independently of each other. In addition, an important part of the application of the SLM is the study of the fit of the data to this model.

**4.2.1 The Rasch Model for Polytomous Responses**

Rasch’s Simple Logistic Model for dichotomous data can be generalised for polytomous responses in which the categories (more than two) are ordered: that is, each type of response can be considered as demonstrating a different amount of some trait (Andrich, 1978; Wright & Masters, 1982).

Key features of this model are that
1. The successive categories are scored with successive integers as is done traditionally in elementary analyses of ordered categorical data Andrich (1978);

2. It estimates the thresholds that define the categories and no assumptions about equal distances between thresholds are made that justify the integer scoring (the thresholds are where there is an equal probability of the student at that location responding at the level below, or the level above that location of difficulty);

3. The model retains the distinctive properties of the dichotomous Rasch Model, that the person and item parameters can be estimated independently of each other.

Although there are different expressions for the model, the form

\[
\Pr(X_{ni} = x) = \exp[x(\beta_n - \delta_i) - \sum_{k=1}^{x} r_{ki}] / \{ \sum_{x'=0}^{m} (\exp[x'(\beta_n - \delta_i) - \sum_{k=1}^{x'} r_{ki}]) \}
\]

(4)

The form used in this thesis is where \( \beta_n \) and \( \delta_i \) retain the same meaning as in equation (3), \( m_i \) is the number of thresholds partitioning the continuum for item \( i \) into \( m_i + 1 \) ordered categories, and \( \tau_{ki} \) is the distance on the latent trait of the \( k \)th threshold from item \( i \)'s location \( \delta_i \), and \( \sum_{k=1}^{m} \tau_{ki} = 0 \). That is, the locations of the thresholds \( \tau_{ki} \) are deviations from the overall item location \( \delta_i \). In the estimation, the constraint \( \sum_{i} \hat{\delta}_i = 0 \) is imposed.

Although the \( \hat{\delta}_i \) are comparable from item to item, the \( \tau_{ki} \) are not. In order to compare the thresholds from task to task, and therefore the levels across the tasks, the location of the item is added to its thresholds giving \( \delta_{ki} = \delta_i + \tau_{ki} \).

The location of the thresholds \( \tau_{ki} \), which are mean deviated from the location \( \delta_i \), are referred to in RUMM as *centralised*; when the item location is added to
the thresholds to give $\delta_k$, the thresholds are referred to as *uncentralised*. Thus uncentralised thresholds can be compared from item to item, and are used throughout this thesis.

### 4.3 Analysis using the Rasch Model

This Section summarises the statistics, as listed below used in the Rasch analysis for analysing student performance.

1. The Person Separation Index
2. The Item Characteristic Curves and $\chi^2$
3. The Category Characteristic Curves
4. The relationship between the person distribution and the location of thresholds (person-item distribution)
5. The relationship between the locations of the thresholds and the six items (threshold map)

These components are described in Sections 4.3.1 to 4.3.5.

#### 4.3.1 Person Separation Index

The Person Separation Index ($r_\beta$) is equivalent to a traditional measure of reliability. This is calculated as the ratio of true variance to observed variance based on the estimate of the person’s ability, $\beta_i$ (Andrich, 1982, p. 98). It can be shown that the index is given by the following expression:

$$r_\beta = \frac{\text{observed score variance}\,\text{–error variance}}{\text{observed score variance}} = 1 - \frac{\text{error variance}}{\text{observed score variance}}$$
The separation index is also an indicator of the power of the test of fit to the model. A low value indicates that the items are not discriminating well between the people whereas a high value suggests that the discrimination is good. This separation index also indicates the power of detecting misfit to the model: the greater the index, the greater the power. Unfortunately this index becomes unstable when there are extreme scores and it is necessary to extrapolate the location estimates of persons and their standard errors (Andrich, 1982).

4.3.2. Item Characteristic Curves and Associated $\chi^2$ Values

The Item Characteristic Curves (ICC) are S-shaped curves that plot the expected score on the task on the vertical axis against the relative ability of the person on the horizontal axis. Those used in these analyses were produced by the RUMM program. They are used as a basis for studying whether or not each item functions invariantly across the trait (Hagquist & Andrich, 2004). In the analysis, the persons are divided into approximately equally sized class intervals. The researcher selects the number of intervals based on the sample size. In the examples in Figures 4.1 and 4.2, four class intervals were chosen. The program calculates the observed mean score for each class interval, and these are plotted against the theoretical Item Characteristic Curve.

The observed means for each class interval should clearly be close to those predicted for the ability within the class interval shown by the theoretical
Item Characteristic Curve. The distances between these observed means and those provided by the theoretical Item Characteristic Curve provide a test of fit between the data and the model. Since the locations of the class intervals are determined from the data as a whole some items do not have a mean score for a particular group, as no one with their ability located in that interval did that task.

The discrepancy between observed means and the Item Characteristic Curve is formalised by a $\chi^2$ test of fit on three degrees of freedom (number of class intervals -1). This test of fit is used as the main statistical indicator to complement the graphical displays. Generally a probability value of greater then 0.05 indicates a reasonable fit to the model. The Item Characteristic Curves in Figures 4.1 and 4.2 are chosen to illustrate tasks for which the fit to the model is adequate (Figure 4.1), and less adequate (Figure 4.2).

Figure 4.1 ICC for the *Initial Analysis* of *Environmental Concerns 1*

The Item Characteristic Curve in Figure 4.1 for the *initial analysis* of the task *Environmental Concerns 1* shows that the observed means fit the curve quite
closely and that the $\chi^2$ probability value of $p < 0.362$ indicates numerically a good fit to the model.

![Figure 4.2 ICC for the Initial Analysis of Homeworkers 2](image)

The high score for the observed mean for the second class interval in Figure 4.2 shows that there is some misfit. The $\chi^2$ probability value of $p < 0.001$ shows that the data do not fit the theoretical model as well as for the previous item, even though the observed means tend to follow the Item Characteristic Curve. In this case the members of this second class interval scored better on this item than would have been predicted by the model.

The table of individual item fit statistics produced by the RUMM program is a summary of information on location, residuals and $\chi^2$ probability values and standard errors for the item location estimates, which indicate the uncertainty of the estimates of the item locations.

### 4.3.3. The Category Characteristic Curves

Category Characteristic Curves (CCC) are curves that plot the probability of each score on the task on the vertical axis against the relative ability of the
person on the horizontal axis. The ordered category structure requires the thresholds to be in their natural order for each score of $x$ taking values from 0, 1, 2…7. In that structure a person located in the region between two thresholds defining the level also has the highest probability of being assigned to that level.

Two Category Characteristic Curves illustrate when thresholds are (Figure 4.3), and are not (Figure 4.4), in natural order. They are the Category Characteristic Curves for the tasks for which the Item Characteristic Curves were given in Section 4.3.2. Figure 4.3 for the initial analysis of the task Environmental Concerns 1 shows an example where all the thresholds are in the natural order.

![Figure 4.3 CCC for the Initial Analysis of Environmental Concerns 1](image)

Figure 4.3 CCC for the Initial Analysis of Environmental Concerns 1

Figure 4.4 shows an example where the thresholds are not in the natural order.
Figure 4.4 CCC for the Initial Analysis of Homeworkers 2

A symptom of the problem is that for the scores of 1 and 2 there is no ability level at which either of these scores would be the most likely outcome. The reversed thresholds provide evidence that the operation of the categories is not working as intended.

When the thresholds are reversed there are one or more levels that will never have the greatest probability, no matter where the location of the person on the continuum and then rescoring is necessary. During rescoring adjacent categories are combined and can lead to a better fit to the model.

4.3.4. The Relationship Between the Person Distribution and the Locations of the Thresholds

A person-item distribution, as illustrated in Figure 4.5, is also provided by the RUMM program. The distribution in Figure 4.5 shows the relationship between the persons and the thresholds of all the items. This graphical representation includes all the thresholds in the range shown. In Figure 4.5 ‘10005.6’ indicates that the sixth threshold for item (task) five is located in the interval from 3.20 to 4.00. The sixth threshold is the one between the levels
coded 5 and 7. The histogram shows that there were 58 persons located in the same interval. The person of ability located at the same position as the threshold would have a 50% chance of responding at the level coded as 6 and a 50% chance of responding at level 7.

Figure 4.5 Person-Item Location for the Coded SOLO Levels

This kind of representation of the thresholds and items shows the complete range of thresholds in the range, even when there are reversed thresholds.

4.3.5. The Relationships between Locations of Thresholds and the Item Distribution

A Threshold Map, which is a graphical representation of the locations of the thresholds for each item, is produced by the RUMM program similar to Figure 4.6. This enables a visual comparison of the relative locations of item thresholds to be made. These are situated at the interfaces of the regions that represent the levels of responses to the items. The RUMM program identifies those items in which the thresholds are reversed, and the threshold order
cannot be displayed in a Figure such as Figure 4.6 further confirming the need to reinvestigate the ordering of the categories.

Figure 4.6 Threshold map for the *Initial Analysis*

The threshold map produced by the RUMM program has the regions labelled 0, 1, 2 and so on appropriate to the coding for a particular analysis. In Figure 4.6 an adapted version is shown on which the SOLO levels 0, 1, 2a, 2b, 3a, 3b, 4a, and 4b have been imposed on the output rather than the coding from 0 to 7 so that the relative locations of the SOLO levels can easily be identified.

During any real analysis the statistics and related inferences in interpreting the responses of students described in this Section are considered more or less simultaneously in making substantive inferences. Therefore the summary of the statistics that follow in the *initial analysis*, and subsequent analyses, are reported in combination with each other rather than simply as the above list.
4.4 Statistics for the Initial Analysis

It will be recalled that the initial analysis was the first analysis conducted on the original data of the coded student responses according to the criteria used in the initial measuring instrument. It was conducted on all of the data for all of the students who completed table interpretation tasks, some of whom completed tasks both before and after the workshop, and others who only completed tasks either before or after a workshop. The individual item fit statistics for the six tasks are given in Table 4.1. The tasks are ordered by location with Suicide 1 as easiest, to Environmental Concerns 2 as the most difficult, with a range from -3.203 to 2.159.

<table>
<thead>
<tr>
<th>Item</th>
<th>Location</th>
<th>SE</th>
<th>$\chi^2$</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Suicide 1</td>
<td>-3.203</td>
<td>0.097</td>
<td>7.617</td>
<td>0.054638</td>
</tr>
<tr>
<td>Homeworkers 1</td>
<td>-2.819</td>
<td>0.130</td>
<td>9.230</td>
<td>0.026380</td>
</tr>
<tr>
<td>Suicide 2</td>
<td>-0.020</td>
<td>0.159</td>
<td>2.443</td>
<td>0.485706</td>
</tr>
<tr>
<td>Homeworkers 2</td>
<td>1.871</td>
<td>0.071</td>
<td>16.522</td>
<td>0.000887</td>
</tr>
<tr>
<td>Environmental Concerns 1</td>
<td>2.011</td>
<td>0.067</td>
<td>3.201</td>
<td>0.361710</td>
</tr>
<tr>
<td>Environmental Concerns 2</td>
<td>2.159</td>
<td>0.081</td>
<td>9.436</td>
<td>0.024019</td>
</tr>
</tbody>
</table>

The $\chi^2$ values and associated probabilities in the table give some indication of how well the data fits the theoretical model. The low probability values, shown in bold type, where $p < 0.05$ for Homeworkers 1, Homeworkers 2 and Environmental Concerns 2 indicate relative marginal misfit. The fit will be considered more closely following an analysis of the operation of the categories. The standard errors (SE) in Table 4.1 show the uncertainty associated with the estimations of the locations of the tasks. These are relatively small compared with the locations of the tasks.
It will be recalled from Chapter 3, that the levels of students' responses to each task were coded according to the set of criteria described in Table 3.7.

The coding was based on the SOLO taxonomy, providing scores: 0 for prestructural responses, 1 for unistructural, 2 for multistructural level 2a, 3 for multistructural level 2b, 4 for relational level 3a, 5 for relational level 3b, 6 for extended abstract level 4a and 7 for extended abstract level 4b, responses.

Table 4.2 has been constructed by location with the easiest task, as defined by the mean location, Suicide 1 (S1), at the top of the table and the hardest task, Environmental Concerns 2 (EC2), at the bottom of the table. It shows the thresholds estimates for all the tasks. The thresholds are indicated as Thresh 1: 0.1, meaning the threshold between 0 (unistructural) and 1 (prestructural), Thresh 2: 1.2a, meaning the threshold between 1 (prestructural) and 2a (multistructural), and so on, showing that for Suicide 1 the first threshold is at -27.899 and the second threshold is at -10.509.

It is evident from Table 4.2 that Homeworkers 2 and Environmental Concerns 2 show reversed thresholds (indicated in bold type) because the locations of the thresholds are not in the natural order, so that rescoring is necessary. The locations of the reverse thresholds are in bold type in Table 4.2.
Table 4.2 Uncentralised Thresholds for the *Initial Analysis*

<table>
<thead>
<tr>
<th>Task</th>
<th>Mean Location</th>
<th>Thresh 1: 0.1</th>
<th>Thresh 2: 1.2a</th>
<th>Thresh 3: 2a.2b</th>
<th>Thresh 4: 2b.3a</th>
<th>Thresh 5: 3a.3b</th>
<th>Thresh 6: 3b.4a</th>
<th>Thresh 7: 4a.4b</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>-3.203</td>
<td>-27.899</td>
<td>-10.509</td>
<td>-1.758</td>
<td>1.498</td>
<td>2.404</td>
<td>4.103</td>
<td>9.741</td>
</tr>
<tr>
<td>HW1</td>
<td>-2.819</td>
<td>-25.849</td>
<td>-9.148</td>
<td>-1.634</td>
<td>0.466</td>
<td>0.923</td>
<td>3.510</td>
<td>12.00</td>
</tr>
<tr>
<td>S2</td>
<td>-0.02</td>
<td>-9.851</td>
<td>-3.943</td>
<td>-1.542</td>
<td>-0.721</td>
<td>0.451</td>
<td>3.903</td>
<td>11.563</td>
</tr>
<tr>
<td>HW2</td>
<td>1.871</td>
<td><strong>0.543</strong></td>
<td><strong>-1.225</strong></td>
<td><strong>-1.394</strong></td>
<td>-0.236</td>
<td>1.976</td>
<td>4.968</td>
<td>8.467</td>
</tr>
<tr>
<td>EC1</td>
<td>2.011</td>
<td>-1.120</td>
<td>-0.736</td>
<td>-0.029</td>
<td>1.073</td>
<td>2.645</td>
<td>4.758</td>
<td>7.486</td>
</tr>
<tr>
<td>EC2</td>
<td>2.159</td>
<td><strong>-0.204</strong></td>
<td><strong>-0.388</strong></td>
<td><strong>-0.452</strong></td>
<td>0.068</td>
<td>1.633</td>
<td>4.707</td>
<td>9.750</td>
</tr>
</tbody>
</table>

Note: Tasks with reversed thresholds are in bold

Table 4.2 also shows that the thresholds with the same ‘a priori’ categories are at different locations across tasks, indicating that there are differences in difficulties between levels within the tasks. It appears to be at the lower end of the thresholds where the largest discrepancies lie.

Figure 4.7 shows a person-item location graph that includes the location of each threshold.
This gives an orientation of the student locations relative to the threshold locations.

The histogram for the 681 profiles shows an approximately symmetrical distribution, with the majority of person locations clustered around the mean of 0.65. The locations of the thresholds are represented by the letter I, (for item), and a numerical code. After the three zeros the first digit of the code corresponds to the task and the number after the decimal point corresponds to the level of the response. In this thesis, the codes for the tasks are as follows:

1: *Environmental Concerns 1*

2: *Suicide 1*

3: *Environmental Concerns 2*

4: *Homeworkers 2*

5: *Suicide 2*

6: *Homeworkers 1*
The codes for the levels, as given in Table 4.2, are given as 0 to 7, from prestructural to extended abstract. Thus, at the lowest end of the scale shown on the graph, I0005.2 is the second threshold for *Homeworkers 1* between levels 1 and 2a. The first threshold has a lower value but is not shown on this scale. In Figure 4.7, there are 27 out of 42 thresholds represented. The other thresholds are well outside the person distribution and are not included.

The threshold map in Figure 4.8 for the *initial analysis* shows the relative locations of the thresholds for the tasks with the levels represented as regions, in terms of SOLO levels. This threshold map makes comparison of relative locations easier than using Table 4.2 that includes the precise numerical locations of the thresholds.

<table>
<thead>
<tr>
<th>Suicide 1 Initial</th>
<th>0</th>
<th>1</th>
<th>2a</th>
<th>2b</th>
<th>3b</th>
<th>4a</th>
<th>4b</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homeworkers 1 Initial</td>
<td>0</td>
<td>1</td>
<td>2a</td>
<td>2b</td>
<td>3b</td>
<td>4a</td>
<td>4b</td>
</tr>
<tr>
<td>Suicide 2 Initial</td>
<td>0</td>
<td>1</td>
<td>2a</td>
<td>2b</td>
<td>3b</td>
<td>4a</td>
<td>4b</td>
</tr>
<tr>
<td>Homeworkers 2 Initial</td>
<td>Reversed Thresholds</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Environmental concerns 1 Initial</td>
<td>Reversed Thresholds</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Environmental concerns 2 Initial</td>
<td>Reversed Thresholds</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 4.8 Threshold Map for SOLO Levels for the *Initial Analysis*

The threshold map in Figure 4.8 shows that for *Suicide 1* and *Homeworkers 1* the corresponding thresholds are in similar locations. *Suicide 2* has quite similar threshold locations to *Suicide 1* and *Homeworkers 1* from the threshold 2a to 2b onwards, although the threshold from 1 to 2a is higher than in the first two tasks. For *Environmental Concerns 1* the thresholds start higher and are very
close together meaning that there is a small ability range within which one would expect students to achieve these scores. It is not possible to comment on the other two tasks as the map cannot represent regions where there are reversed thresholds, however, Figure 4.8 illustrates visually that achievement of SOLO levels did not require equivalent abilities for the different tasks.

### 4.5 Category and Item Characteristic Curves for Initial and Final Analyses

As previously discussed for the initial analysis, there were problems with some of the categories for the tasks. Reversed thresholds were the most obvious problem for Homeworkers 2 and Environmental Concerns 2. Changes made through the rescoring process for each task resulted in a set of categories for each task that fitted the requirements of the Rasch model, as indicated by the Category Characteristic Curves, the Item Characteristic Curves and the $\chi^2$ probability values.

In this Section two analyses are compared for each task, the initial analysis and the final analysis. The initial analysis used data from the use of the original categories. Following the initial analysis the data were rescored as a result of problems detected in that analysis. For example, some categories were combined because they did not provide reliable comparisons. After rescoring a final analysis was conducted both to check on the adequacy of the model and to determine the person abilities and the item difficulties.
The Category Characteristic Curves and Item Characteristic Curves in this Section are for the initial analysis, and the final analysis after rescoring for each task. The CCC, ICC and the probability values were used to guide the rescoring of the categories to fit the model. The tasks are shown in the following order: Environmental Concerns 1, Suicide 1, Homeworkers 1, Homeworkers 2, Suicide 1 and Environmental Concerns 2. The first three tasks, Environmental Concerns 1, Suicide 2 and Homeworkers 1, required the same rescoring, while the other three required variations of this pattern. For each task a table showing the rescoring for the final analysis is given.

4.5.1 CCC and ICC for the Initial and Final Analyses of Environmental Concerns 1

This Section examines CCC and ICC for the initial analysis and final analysis of Environmental Concerns 1.

Figure 4.9 CCC for the Initial Analysis of Environmental Concerns 1

The CCC in Figure 4.9 show that the initial categories work as intended for Environmental Concerns 1. The distance between the first and second
thresholds is very small indicating a potential problem as further data may suggest that a score of 1 is never the most likely.

![Figure 4.10 ICC for the Initial Analysis of Environmental Concerns 1](image)

The ICC in Figure 4.10 shows that the data visually fits the model, and this is confirmed by the $\chi^2$ probability value of $p < 0.362$.

Intermediate analyses between the initial and final analyses were conducted, each time rescoring until the best fit to the model was found using the probability values associated with the ICC as a guide. The resulting combinations are recorded in the table 4.3 for the final analysis. Table 4.3 shows the way that the initial categories were combined for the final analysis for Environmental Concerns 1. Corresponding rescorings are shown in Tables 4.4, 4.5, 4.6, 4.7 and 4.8 for the other tasks. The resulting qualitative combinations of the criteria are described in detail in Chapter 5.

<table>
<thead>
<tr>
<th>Category</th>
<th>0</th>
<th>1</th>
<th>2a</th>
<th>2b</th>
<th>3a</th>
<th>3b</th>
<th>4a</th>
<th>4b</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial score</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>Final score</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>
This table shows that categories 0 and 1 have been combined to form a new category now coded 0. There were almost no student responses in the 0 category, so it is logical to combine the 0 and 1. Similarly it also shows that 4a and 4b have been combined to form one category now coded 5. There were no responses in the 4b category and not many in the 4a category.

![Figure 4.11 CCC for the Final Analysis of Environmental Concerns 1](image)

The CCC in Figure 4.11 confirms that after rescoring the six categories still work as intended for Environmental Concerns 1.

![Figure 4.12 ICC for the Final Analysis of Environmental Concerns 1](image)
The ICC in Figure 4.12 shows that data fit the curve well. This is confirmed by the $\chi^2$ probability value of $p < 0.191$.

### 4.5.2 CCC and ICC for the Initial and Final Analyses of Suicide 2

This Section examines the CCC and ICC for the *initial analysis* and *final analysis* for Suicide 2.

![Figure 4.13 CCC for the Initial Analysis of Suicide 2](image)

The CCC in Figure 4.13 show that the middle categories work well but extremes of the scale should be collapsed with adjacent categories because the thresholds though ordered extend past the ability range of the respondents.
Figure 4.14 ICC for the *Initial Analysis of Suicide 2*

The ICC in Figure 4.14 show a good fit between the observed and expected means. The $\chi^2$ probability value of $p < 0.486$ shows a good fit to the model. It can be seen that small differences in the observed means from the theoretical curve can distort the regular ogive.

As described earlier Table 4.4 shows how the categories are combined to improve the fit to the model, as described previously.

Table 4.4 Rescoring for *Suicide 2*

<table>
<thead>
<tr>
<th>Category</th>
<th>0</th>
<th>1</th>
<th>2a</th>
<th>2b</th>
<th>3a</th>
<th>3b</th>
<th>4a</th>
<th>4b</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial score</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>Final score</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

Figure 4.15 CCC for the *Final Analysis of Suicide 2*

Figure 4.15 show the categories work as intended for *Suicide 2* after rescoring.
Figure 4.16 ICC for the Final Analysis of Suicide 2

The ICC in Figure 4.16 shows that second and 4th class intervals still do not fit as well as the other two. The probability value of $p < 0.587$ for the $\chi^2$ shows that the fit to the model is better than in the initial analysis.

4.5.3 CCC and ICC for the Initial and Final Analyses of Homeworkers 1

This Section examines CCC and ICC for the initial analysis and final analysis for Homeworkers 1.

Figure 4.17 CCC for Initial Analysis of Homeworkers 1

The CCC in Figure 4.17 show that the categories are in the order that was intended for Homeworkers 1. The extremes look as if they need combining.
Figure 4.18 ICC for *Initial Analysis of Homeworkers 1*

The ICC in Figure 4.18 shows that the observed mean for the second class interval is below the expected mean, although the other means fit on the curve well. The $\chi^2$ probability $p < 0.026$ shows the fit is statistically marginal.

As described earlier Table 4.5 shows how the categories are combined to improve the fit to the model.

Table 4.5 Rescoring for *Homeworkers 1*

<table>
<thead>
<tr>
<th>Category</th>
<th>0</th>
<th>1</th>
<th>2a</th>
<th>2b</th>
<th>3a</th>
<th>3b</th>
<th>4a</th>
<th>4b</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial score</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>Final score</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>
Figure 4.19 CCC for the *Final Analysis of Homeworkers 1*

The CCC in Figure 4.19 show that the categories work in the order intended but the distance between the 4th and fifth thresholds is very small.

![CCC graph](image)

Figure 4.20 ICC for the *Final Analysis of Homeworkers 1*

The ICC in Figure 4.20 shows that after rescoring the observed means fit the theoretical curve, although there is no mean for the second class interval as no student has the ability in that interval. The $\chi^2$ probability of $p < 0.093$ shows a good fit, and a considerable improvement on the initial analysis.

### 4.5.4 CCC and ICC for the Initial and Final Analyses of Homeworkers 2

This Section examines the CCC and ICC for the initial analysis and final analysis for Homeworkers 2.
Figure 4.21 CCC for the *Initial Analysis of Homeworkers 2*

The CCC in Figure 4.21 show that the categories do not work as intended, there are reversed thresholds between -2 and 1.

Figure 4.22 ICC for the *Initial Analysis of Homeworkers 2*

The ICC in Figure 4.22 shows the observed mean for the second class interval is above the anticipated value although other observed means fit theoretical curve. The $\chi^2$ probability value of $p < 0.001$ shows misfit to the model.

As described earlier Table 4.6 shows how the categories are combined to improve the fit to the model.
Table 4.6 Rescoring for *Homeworkers 2*

<table>
<thead>
<tr>
<th>Category</th>
<th>0</th>
<th>1</th>
<th>2a</th>
<th>2b</th>
<th>3a</th>
<th>3b</th>
<th>4a</th>
<th>4b</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial score</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>Final score</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

The CCC in Figure 4.23 confirm that the categories work as intended for *Homeworkers 2* after rescoring.

The ICC in Figure 4.24 shows that the observed means fit the theoretical curve. The $\chi^2$ probability value $p < 0.864$ now shows a good fit to the model.

4.5.5 CCC and ICC for the *Initial and Final Analyses of Suicide 1*
This Section examines the CCC and ICC for the *initial analysis* and *final analysis* for Suicide 1.

Figure 4.25 CCC for the *Initial Analysis* of Suicide 1

The CCC in Figure 4.25 show that the categories are in the natural order. Categories 0 and 7 could be rescored with adjacent categories because of their extreme width outside of the ability range.

Figure 4.26 ICC for the *Initial Analysis* of Suicide 1

The ICC in Figure 4.26 shows that the data fits the theoretical model quite well. The $\chi^2$ probability of $p < 0.055$ shows an acceptable fit to the model.

As described earlier Table 4.7 shows how the categories are combined to improve the fit to the model.
Table 4.7 Rescoring for *Suicide 1*

<table>
<thead>
<tr>
<th>Category</th>
<th>0</th>
<th>1</th>
<th>2a</th>
<th>2b</th>
<th>3a</th>
<th>3b</th>
<th>4a</th>
<th>4b</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial score</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>Final score</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

Figure 4.27 CCC for the *Final Analysis* of *Suicide 1*

The CCC in Figure 4.27 show the categories work as intended for *Suicide 1*.

Figure 4.28 ICC for the *Final Analysis* of *Suicide 1*

The ICC in Figure 4.28 shows that the observed mean for the second class interval is a little low. The \( \chi^2 \) probability value of \( p <0.006 \) shows that the fit is not adequate. However, further rescoring was unable to improve this fit.
4.5.6 CCC and ICC for the *Initial and Final Analyses of Environmental Concerns 2*

This Section examines the CCC and ICC for the *initial analysis* and *final analysis* of Environmental Concerns 2.

![Figure 4.29 CCC for the Initial Analysis of Environmental Concerns 2](image)

The CCC in Figure 4.29 show reversed thresholds and indicates a need to rescore the categories.

![Figure 4.30 ICC for the Initial Analysis of Environmental Concerns 2](image)

The Item Characteristic Curve in Figure 4.30 shows that observed mean for the third class interval is higher than the expected mean. The $\chi^2$ probability value of $p <0.024$ shows that the data do not fit the model very well.
As described earlier Table 4.8 shows how the categories are combined to improve the fit to the model.

Table 4.8 Rescoring for *Environmental Concerns 2*

<table>
<thead>
<tr>
<th>Category</th>
<th>0</th>
<th>1</th>
<th>2a</th>
<th>2b</th>
<th>3a</th>
<th>3b</th>
<th>4a</th>
<th>4b</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial score</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>Final score</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

Figure 4.31 CCC for the Final Analysis of Environmental Concerns 2

The Category Characteristic Curves in Figure 4.31 show that the categories work as intended for *Environmental Concerns 2* after rescoring.

Figure 4.32 ICC for the *Final Analysis of Environmental Concerns 2*
The ICC in Figure 4.32 confirms that after rescoring the observed means fit the theoretical curve, although the third class interval had zero interval size. The $\chi^2$ probability value of $p < 0.624$ now shows a good fit to the model.

The summary in Table 4.9 shows the combinations of original categories combined to form new categories in the rescoring process. For all of the tasks, the two categories at both extremes were combined making six categories; 0 and 1 were combined to become 0, and 6(4a) and 7(4b) were combined to become 5, reflecting the small number of respondents with abilities around each of the two extreme thresholds.

Table 4.9 Overall Combination of Categories

<table>
<thead>
<tr>
<th>Level</th>
<th>Code</th>
<th>0</th>
<th>1</th>
<th>2a</th>
<th>2b</th>
<th>3a</th>
<th>3b</th>
<th>4a</th>
<th>4b</th>
<th>Initial $\chi^2$ Prob</th>
<th>Final $\chi^2$ Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>EC 1</td>
<td>0 0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>0.362</td>
<td>0.191</td>
</tr>
<tr>
<td>S 2</td>
<td>0 0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>0.486</td>
<td>0.587</td>
</tr>
<tr>
<td>HW 1</td>
<td>0 0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>0.026</td>
<td>0.093</td>
</tr>
<tr>
<td>HW 2</td>
<td>0 0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>0.001</td>
<td>0.864</td>
</tr>
<tr>
<td>S 1</td>
<td>0 0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>0.055</td>
<td>0.006</td>
</tr>
<tr>
<td>EC 2</td>
<td>0 0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>0.024</td>
<td>0.624</td>
</tr>
</tbody>
</table>

For the three tasks *Environmental Concerns 1*, *Suicide 2* and *Homeworkers* the rescoring required only combining of these extreme categories. *Homeworkers 2* needed one more combination of 2(2a) and 3(2b), while for *Suicide 1* the combination of 4(3a) and 5(3b), was needed. For *Environmental Concerns 2* categories 2(2a), 3(2b) and 4(3a) were combined making only 4 categories in the final version. Table 4.9 also gives the $\chi^2$ probability values to give a sense of how the tasks fitted the model before and after the rescoring of categories with a value of $p > 0.05$ indicates a reasonable fit. This is the case in the final analysis for all the tasks except Suicide 1, as discussed.
4.6 Statistics for the Final Analysis

In Section 4.4 the Individual Item Fit shown in Table 4.1 included the individual item fit for the tasks located in order of difficulty according to the initial analysis. Table 4.10 shows the statistics for items in order of difficulty for the final analysis. While the order remained the same for the harder three tasks, it changed for the three easiest tasks. For these the initial order was, from easiest to hardest: Suicide 1, the easiest, then Homeworkers 1 and Suicide 2 but that changed to Homeworkers 1, Suicide 2, Suicide 1 in the final analysis.

Table 4.10: Individual Item Fit for the Final Analysis

<table>
<thead>
<tr>
<th>Task</th>
<th>Mean Location</th>
<th>SE</th>
<th>$\chi^2$</th>
<th>Final $\chi^2$ Probability</th>
<th>Initial $\chi^2$ Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homeworkers 1</td>
<td>-1.260</td>
<td>0.109</td>
<td>4.761</td>
<td>0.092510</td>
<td>0.026380</td>
</tr>
<tr>
<td>Suicide 2</td>
<td>-1.092</td>
<td>0.162</td>
<td>1.932</td>
<td>0.586740</td>
<td>0.485706</td>
</tr>
<tr>
<td>Suicide 1</td>
<td>0.246</td>
<td>0.144</td>
<td>12.445</td>
<td>0.006007</td>
<td>0.054638</td>
</tr>
<tr>
<td>Homeworkers 2</td>
<td>0.404</td>
<td>0.115</td>
<td>0.739</td>
<td>0.863971</td>
<td>0.000887</td>
</tr>
<tr>
<td>Environ Concern 1</td>
<td>0.735</td>
<td>0.077</td>
<td>4.756</td>
<td>0.190565</td>
<td>0.361710</td>
</tr>
<tr>
<td>Environ Concern 2</td>
<td>0.966</td>
<td>0.174</td>
<td>0.944</td>
<td>0.623836</td>
<td>0.024019</td>
</tr>
</tbody>
</table>

As well as the order of difficulty changing, the difference in location of easiest to hardest decreased from 5.362 logits to 2.226 logits. This is a direct result of combining the categories and thus reducing the score range. The last two columns of the table show the $\chi^2$ probability values for the initial analysis and the final analysis. The numbers in bold type are those with values of less than 0.05, showing an improvement in the final analysis.

Table 4.10 also shows that in the initial analysis all of the table interpretation tasks with the first version of the third subtask Explain and discuss the information contained in the table (Environmental Concerns 1, Suicide 1,
Homeworkers 1) were less difficult than their corresponding table interpretation task with the second version of the third subtask What societal and environmental factors might account for the data in the table? (Environmental Concerns 2, Suicide 2, Homeworkers 2). This pattern is consistent in the final analysis for Environmental Concerns and Homeworkers although not for the Suicide table. It seems that the change of phraseology has generally reduced the difficulty of the tasks, perhaps because it has suggested more highly developed responses to students. The relative locations of the thresholds for the final analysis are shown in Table 4.11.

Table 4.11 The Uncentralised Thresholds for the Final Analysis

<table>
<thead>
<tr>
<th></th>
<th>Threshold 1</th>
<th>Threshold 2</th>
<th>Threshold 3</th>
<th>Threshold 4</th>
<th>Threshold 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>HW1</td>
<td>0/1.2a</td>
<td>-4.656</td>
<td>-1.976</td>
<td>2b.3a</td>
<td>-0.432</td>
</tr>
<tr>
<td>S2</td>
<td>0/1.2a</td>
<td>-4.412</td>
<td>-2.374</td>
<td>2b.3a</td>
<td>-1.724</td>
</tr>
<tr>
<td>S1</td>
<td>0/1.2a/b</td>
<td>-4.791</td>
<td>0.908</td>
<td>3a/b.4a/b</td>
<td>-4.620</td>
</tr>
<tr>
<td>HW2</td>
<td>0/1.2a</td>
<td>-2.627</td>
<td>-1.079</td>
<td>3a.3b</td>
<td>1.012</td>
</tr>
<tr>
<td>EC1</td>
<td>0/1.2a</td>
<td>-2.209</td>
<td>-0.502</td>
<td>2b.3a</td>
<td>0.439</td>
</tr>
<tr>
<td>EC2</td>
<td>0/1.2a</td>
<td>-2.270</td>
<td>1.531</td>
<td>3b.4a/b</td>
<td>3.638</td>
</tr>
</tbody>
</table>

The notation for Threshold 1 for Homeworkers 1 is 0/1 . 2a, that shows the location of the threshold between the combination of the categories 0 and 1 and category 2a. The second threshold for Homeworkers 1 is labelled 2a.2b which shows the location of the threshold between categories 2a and 2b, and so on for the other thresholds. The second threshold for Suicide 1 is written as 2a/b.3a/b which denotes the threshold between the combination of 2a and 2b and the combination of 3a and 3b.
The first thresholds for *Homeworkers 1*, *Suicide 2* and *Suicide 1* are very similar, as are the first thresholds for *Homeworkers 2*, *Environmental Concerns 1* and *Environmental Concerns 2*. The patterns are different for the other thresholds that are more easily compared in the threshold map in Figure 4.33.

Figure 4.33 is the threshold map for the final analysis showing the thresholds in terms of the SOLO levels.

![Figure 4.33 Threshold Map for the SOLO Levels for the Final Analysis](image)

This shows that the thresholds are within a much smaller range than for the initial analysis shown in Figure 4.8 and correspond closely with the ability range of students. In Figure 4.33 the range is from -5 to +5 whereas in Figure 4.8 for the initial analysis the thresholds are within the range -28 to +13. This change is as a result of combining the extreme categories at each end. The lower end of the category 4a/b is similar for the tasks except for *Homeworkers*.
1. Given the small range for the scale the location of the middle thresholds are quite similar for the corresponding SOLO levels.

The Person Item Fit distribution for the final analysis is given in Figure 4.34. The person distribution is quite symmetrical, as with that of the initial analysis, and the spread of persons is narrower than that for the initial analysis due to the combination of the extreme categories. The graph includes 22 thresholds out of 2. The others, namely 1.5, 2.3 and 4.4, are out of the range of the scale.

![Figure 4.34 Person-Item Fit for the Final Analysis](image)

**4.7 Summary**

In this chapter an analysis of the data, using the Rasch Model, was described from the initial analysis until the final analysis stage. Rasch modelling has the advantages that it accounts for the relative difficulties of tasks, and does not assume that identical coding of different tasks necessarily implies equal difficulties between adjacent levels of response. The Rasch model permits the
comparison of the categories and the tasks so that student abilities can be
validly and reliably compared regardless of which tasks are involved. For the
data obtained in this study appropriate comparisons could not be made
without the use of this model. The interpretation of the qualitative effects on
the criteria for the instrument consequences of the collapsing of the
categories is pursued in Chapter 5.

During the analysis, the instrument underwent refinement through rescoring
to obtain properly ordered thresholds and the best possible fit. At the same
time the students’ abilities were located on a common scale and so could
subsequently be examined for change in location corresponding to the
change in position on the scale after the workshop compared with before the
workshop. An evaluation of the change of location is described in Chapter 6.

This Chapter has shown that the combination of categories used in the final
analysis produces a measuring instrument for the six tasks used in the study
that will distinguish well between the students responding to these tasks. For
three of the tasks it was seen that there were common categories. For the
others there was some variation. One aim of the study was to construct an
instrument that would work in general for tasks of this kind. To investigate
the possibility of constructing such an instrument with common categories
for all tasks an instrument with only four categories for all tasks was
undertaken, and reported in Chapter 6.
CHAPTER 5 QUALITATIVE ANALYSIS

5.1 Introduction

In this Chapter the qualitative affects of the collapsing of the categories described in Chapter 4 are considered. This entails examining the changes to the criteria in the initial analysis (with seven distinct categories for all tasks), to the final instrument in the final analysis (that uses varying numbers of levels depending on the task). Comparisons are made between the thresholds for the six tasks in the final analysis and some observations are given concerning possible reasons for the differences in locations of the thresholds.

5.2 The Measuring Instrument

The initial version of the measuring instrument is reproduced in Table 5.1 to provide for comparison with the refined instrument for the six tasks. The rescoring of the categories and subsequent coding for all of the tasks was described in detail in Chapter 4. In this rescoring the two extreme categories for each end of the scale, (0 and 1) and (4a and 4b), were combined for all of the tasks. These were the only rescoring required for Environmental Concerns 1, Suicide 2 and Homeworkers 1. The analysis using the Rasch Model showed that further rescoring was necessary for Homeworkers 2, Suicide 1 and Environmental Concerns 2.
### Table 5.1 Criteria for Allocation to Initial Levels

<table>
<thead>
<tr>
<th>Initial Category</th>
<th>Coding</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prestructural</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level 0</td>
<td>0</td>
<td>No responses, or all responses inappropriate or irrelevant.</td>
</tr>
<tr>
<td>Unistructural</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level 1</td>
<td>1</td>
<td>Description of what the table is about including the headings or categories but with no additional information.</td>
</tr>
<tr>
<td>Multistructural</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level 2a</td>
<td>2</td>
<td>Specific values used to make a comparison at a particular time (for example the highest or lowest value for a category), or one trend over time is described.</td>
</tr>
<tr>
<td>Level 2b</td>
<td>3</td>
<td>Two trends, or one trend and a sub-trend or a trend and a comparison, within a category, are described.</td>
</tr>
<tr>
<td>Relational</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level 3a</td>
<td>4</td>
<td>A possible reason for trend(s) or comparisons made above is included, possibly related to the media, social influences, changes in technology, or a comment is made on the sampling procedure.</td>
</tr>
<tr>
<td>Level 3b</td>
<td>5</td>
<td>More than one reason for the comparisons or trends or one reason and a comment on the sampling procedure are given.</td>
</tr>
<tr>
<td>Extended abstract</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level 4a</td>
<td>6</td>
<td>An understanding of the relevance of the information to overriding abstract principles is shown, and an application to a situation that is not given in the table.</td>
</tr>
<tr>
<td>Level 4b</td>
<td>7</td>
<td>This response takes in more global and philosophical views concerning the context.</td>
</tr>
</tbody>
</table>

### 5.2.1 Final Categories for Environmental Concerns 1, Suicide 2 and Homeworkers 1

The combination of the levels coded 0 and 1 during rescoring of all the items produces a category that includes both incorrect or irrelevant answers and some low level answers. It was necessary to include the lowest level,
prestructural (0), in the initial table of criteria for completeness but it was found that very few students responded at that level. Therefore collapsing the levels corresponding to 0 and 1 produced a more useful category. Similarly, there were very few students who responded at the 4b level and the combination of 4a and 4b for all items produced a category that catered for any responses at the more abstract level.

Table 5.2 Final Categories for Environmental Concerns 1, Suicide 2, and Homeworkers 1

<table>
<thead>
<tr>
<th>Coding &amp; Categories</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 (0,1)</td>
<td>No correct responses or a description of what the table is about including the headings or categories, but with no additional information.</td>
</tr>
<tr>
<td>1 (2a)</td>
<td>Specific values used to make a comparison at a particular time (for example the highest or lowest value for a category), or one trend over time is described.</td>
</tr>
<tr>
<td>2 (2b)</td>
<td>Two trends, or one trend and a sub-trend, or a trend and a comparison, within a category, are described.</td>
</tr>
<tr>
<td>3 (3a)</td>
<td>A possible reason for trend(s) or comparisons made above is included, possibly related to the media, social influences, changes in technology, or a comment is made on the sampling procedure.</td>
</tr>
<tr>
<td>4 (3b)</td>
<td>More than one reason for the comparisons or trends, or one reason and a comment on the sampling procedure are given.</td>
</tr>
<tr>
<td>5 (4a, 4b)</td>
<td>An understanding of the relevance of the information to overriding abstract principles is shown, and an application to a situation that is not given in the table. This response may take in more global and philosophical views concerning the context.</td>
</tr>
</tbody>
</table>

For the three tasks Environmental Concerns 1, Suicide 2 and Homeworkers 1 the sublevels 2a and 2b as well as 3a and 3b are retained as separate categories. The differences between the criteria of a response at level 2a, and the criteria for a response at level 2b are defined by the number of trends or comparisons given in the response. This looks like a question of quantity rather than
quality of response. However, the increase in the number of trends or comparisons in the responses may also be attributed to a more thoughtful and considered response, in which case the responses at 2a and 2b can be seen as qualitatively different. The same can be said for the levels 3a and 3b of the relational level, which are then seen as separate qualitative categories.

5.2.2 Final Categories for Homeworkers 2

The only difference between Homeworkers 2 and the previous three tasks in Section 5.2.1 is the combination of 2a and 2b because they were not distinguishable, even though they were distinguishable for Homeworkers 1. The criterion for the combination requires that the students identify at least one comparison or trend. The analysis showed that the probability of 2a, that is the identification of one comparison or trend, was never the most likely. Thus students were more likely to give more than one comparison or trend in the context of this task. So the rescoring for Homeworkers 2 produced five categories (0, 1), (2a, 2b), (3a), (3b), (4a, 4b), which are shown in Table 5.3.

Table 5.3 Final Categories for Homeworkers 2

<table>
<thead>
<tr>
<th>Coding &amp; Categories</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 (0, 1)</td>
<td>No correct responses or a description of what the table is about including the headings or categories, but with no additional information.</td>
</tr>
<tr>
<td>1 (2a, 2b)</td>
<td>Specific values used to make at least one comparison at a particular time (for example the highest or lowest value for a category), or at least one trend over time is described.</td>
</tr>
<tr>
<td>2 (3a)</td>
<td>A possible reason for trend(s) or comparisons made above is included, possibly related to the media, social influences, changes in technology, or a comment is made on the sampling procedure.</td>
</tr>
<tr>
<td>3 (3b)</td>
<td>More than one reason for the comparisons or trends or one reason and a comment on the sampling procedure are given.</td>
</tr>
</tbody>
</table>
An understanding of the relevance of the information to overriding abstract principles is shown, and an application to a situation that is not given in the table. This response may take in more global and philosophical views concerning the context.

5.2.3 Final Categories for Suicide 1

The rescoring for this task produced only four categories, corresponding to (0, 1); (2a, 2b); (3a, 3b); (4a, 4b). These categories distinguish between the SOLO levels multistructural (2a, 2b), relational (3a, 3b) and extended abstract, but not between their sublevels.

Table 5.4 Final Categories for Suicide 1

<table>
<thead>
<tr>
<th>Coding &amp; Categories</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 (0, 1)</td>
<td>No correct responses or a description of what the table is about including the headings or categories, but with no additional information.</td>
</tr>
<tr>
<td>1 (2a, 2b)</td>
<td>Specific values used to make at least one comparison at a particular time (for example the highest or lowest value for a category), or at least one trend over time is described.</td>
</tr>
<tr>
<td>2 (3a, 3b)</td>
<td>At least one possible reason for trend(s) or comparisons made above is included, possibly related to the media, social influences, changes in technology, or includes a comment on the sampling procedure and a reason for the comparison or trend.</td>
</tr>
<tr>
<td>4 (4a, 4b)</td>
<td>An understanding of the relevance of the information to overriding abstract principles is shown, and an application to a situation that is not given in the table. This response may take in more global and philosophical views concerning the context.</td>
</tr>
</tbody>
</table>

The argument for the combination of the criteria for 2a and 2b is the same as for Homeworkers 2. Thus, students could be expected to offer at least one comparison or trend. The combination of 3a and 3b for this task produces a category that requires at least one reason for differences within a set of data
or change in the numerical value of the variable over time. In the initial analysis the distance between the thresholds corresponding to the level of 2a spanned a much shorter distance on the scale than that for level 2b. This may mean that wherever students were aware of the importance of giving reasons they could access more than one from this context.

5.2.4 Final Categories for Environmental Concerns 2

The rescoring for Environmental Concerns 2 produced four categories (0, 1), (2a, 2b, 3a), (3b) and (4a, 4b). The second category includes multistructural and the lower level of relational. The combination of 2a and 2b with 3a is needed only for this task; the other tasks distinguish between multistructural and relational responses.

Table 5.5 Final Categories for Environmental Concerns 2

<table>
<thead>
<tr>
<th>Coding &amp; Categories</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 (0,1)</td>
<td>No correct responses or a description of what the table is about including the headings or categories, but with no additional information.</td>
</tr>
<tr>
<td>1 (2a, 2b, 3a)</td>
<td>Specific values used to make at least one comparison at a particular time (for example the highest or lowest value for a category), or at least one trend over time is described. A possible reason for trend(s) or comparisons made above is included, possibly related to the media, social influences, changes in technology, or includes comments on sampling procedures.</td>
</tr>
<tr>
<td>2 (3b)</td>
<td>More than one reason for the comparisons or trends or one reason and a comment on the sampling procedure are given.</td>
</tr>
<tr>
<td>3 (4a, 4b)</td>
<td>An understanding of the relevance of the information to overriding abstract principles is shown, and an application to a situation that is not given in the table. This response may take in more global and philosophical views concerning the context.</td>
</tr>
</tbody>
</table>
The Characteristic Category Curves for the initial analysis shown in Figure 4.29 show reversed thresholds, so there is no part of the scale for which the probability of responding at a level 1 or 2a was ever the highest so whenever respondents made comparisons they would not respond at these levels. The Category Characteristic Curves for the initial analyses of Environmental Concerns 1 and Environmental Concerns 2 are fairly similar in shape but for Environmental Concerns 2 there was little likelihood of getting a 2a or 2b. It is likely that the change in wording for the third subtask made it more usual for students to include a reason with their comparisons in this context. This influence of the language used for the subtasks is discussed in Chapter 8.

5.3 Comparison of Related Tasks

As described in section 5.2 the rescoring produced an instrument that did not have exactly the same categories for all the tasks. In order to investigate further the results of the rescoring the categories for the final analysis for the pairs of tasks concerned with the same table of data were compared to identify similarities and differences in the collapsing of categories. These categories are given in Table 5.6.

Table 5.6 Comparison of Pairs of Tasks

<table>
<thead>
<tr>
<th>Tasks with version 1 for the third subtask: Explain and discuss the information contained in the table</th>
<th>Tasks with version 2 for the third subtask: What societal and environmental factors might account for the data in the table?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Environmental Concerns 1: 0/1 2a 2b 3a 3b 4a/b</td>
<td>Environmental Concerns 2: 0/1 2a/b 3a 3b 4a/b</td>
</tr>
</tbody>
</table>
In Table 5.6 the differences in the ways that the levels have been combined for the pairs of tasks are highlighted in bold type. There are no common patterns to the differences for all the tasks but for both Environmental Concerns and Homeworkers the categories 2a and 2b combine to give 2a.b.

It was anticipated by the researcher that the second version would be more likely to give students the impression that they should think of reasons for variations in the data in the table that would lead to higher levels of responses. Thus it was expected that the tasks with the second version of the third subtask would be easier than the tasks with the first version, since the revised wording would act as a kind of prompt. However, when using the mean locations (the mean of the thresholds for that task) of the tasks to compare their difficulties it can be seen from Table 4.10 that for the final analysis of both the Environmental Concerns and the Homeworker tasks the location of version ‘1’ is easier than version ‘2’; for Suicide the reverse is true so:

- Environmental Concerns 1 at 0.735 is easier than Environmental Concerns 2 at 0.966;
- Suicide 1 at 0.246 is harder than Suicide 2 at -1.092 and;
• *Homeworkers 1* at -1.260 is easier than *Homeworkers 2* at 0.404.

The standard errors range from 0.077 for *Environmental Concerns 1* to 0.174 for *Environmental Concerns 2* sufficiently low to suggest suggest that the changes in location are not merely the result of random error.

### 5.4 Summary

The rescoring process resulted in some variation in the combinations of categories for the different tasks. This meant that the criteria established for the initial instrument required modification to fit the refined instrument. The pairs of tasks that used common tables of data were examined to find similarities and differences in the combinations and no consistent pattern was found for the two versions of the tasks. The thresholds for the pairs were also examined to see if the tasks with the first version of the subtasks appeared to be consistently harder than those with the second version of the subtask but this was not the case.

Although there are variations in the difficulties of the tasks as located by the Rasch analysis, these differences are not large ones. Overall the measuring instrument has some specific differences for the six tasks but on the whole the criteria for evaluating the responses are similar for all the tasks. The criteria use to locate the responses in terms of the qualitative levels of the SOLO taxonomy have been found to be appropriate for distinguishing between students’ responses.
CHAPTER 6  FOUR-CATEGORY ANALYSIS

6.1 Introduction

It was shown in Chapters 4 and 5 that a measuring instrument based on the levels of the SOLO taxonomy, including sublevels, could be refined using the Rasch analysis to produce a measuring instrument to distinguish between students' performances on table interpretation tasks. The instrument developed in the final analysis involved variations in categories depending on the table interpretation task undertaken.

In the construction of the initial instrument the intention had been to construct a single instrument that would be applicable to all of the six table interpretation tasks. The researcher was guided by the literature on the SOLO taxonomy (Biggs & Collis, 1982) which proposed the existence of levels: prestructural, unistructural, multistructural, relational and extended abstract, with sublevels for multistructural, relational and extended abstract, and so criteria appropriate to these were included in the measuring instrument. However, it was found during the rescoring process that whilst for three table tasks the categories were consistent and included sublevels of the SOLO levels, for the other three tasks this was not the case.

Given this problem the researcher decided to investigate the possibility of an alternative instrument that would be simpler and possibly more efficient. For the new instrument the researcher decided to use only four categories that combined the unistructural and prestructural levels and used the
multistructural, relational and extended abstract levels, each without
sublevels. It was noted that the rescoring process had produced a successful
set of four categories for Suicide 1 structured in this way. This new
instrument (Table 6.1) was constructed by the researcher and subsequently
analysed, and the results of this four-category analysis are reported in this
Chapter.

Table 6.1 Categories for Four-Category Analysis

<table>
<thead>
<tr>
<th>Coding &amp; Categories</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 (0,1)</td>
<td>No correct responses or just a description of what the table is about, including the headings or categories, but with no additional information.</td>
</tr>
<tr>
<td>1 (2a, 2b)</td>
<td>Specific values used to make at least one comparison at a particular time (for example the highest or lowest value for a category), or at least one trend over time is described.</td>
</tr>
<tr>
<td>2 (3a, 3b)</td>
<td>At least one possible reason for trend(s) or comparisons made above is given which is possibly related to the media, social influences or changes in technology. Or a comment is made about the sampling procedure and a reason for the comparison or trend is given.</td>
</tr>
<tr>
<td>4 (4a, 4b)</td>
<td>An understanding of the relevance of the information to overriding abstract principles is shown, and an application to a situation that is not given in the table. This response may take in more global and philosophical views concerning the context.</td>
</tr>
</tbody>
</table>

6.2 Statistics for the Four-Category Analysis

The data were rescored for the six tasks into the four categories and the
Rasch model used to analyse this data. The analysis produced summary
statistics and CCC and ICC, as for the initial and final analyses, and these were
used to evaluate the four-category instrument.
In Chapter 4 the CCC and ICC were reported according to commonalities in the rescoring. In the *four-category analysis* the tasks all have the same number of categories, and the order for reporting the tasks is chosen so that related pairs of tasks using the same table of data are considered together. The CCC and ICC for *Environmental Concerns 1 & 2* followed by *Suicide 1 & 2* and then *Homeworkers 1 & 2* are reported in Sections 6.2.1, 6.2.2 and 6.2.6. It will be recalled that for all the tasks four class intervals were selected and that these are allocated by the RUMM program, according to the data as a whole. In some cases only three observed means are plotted because, for the absent interval no student in that ability range attempted that particular task in this new analysis.

**6.2.1. CCC and ICC for the *Four-Category Analysis of Environmental Concerns 1***

The CCC in Figure 6.1 show that the four categories work as intended 1.
Figure 6.2 ICC for the *Four-Category Analysis of Environmental Concerns 1*

The ICC in Figure 6.2 shows that the observed mean for the third class interval is lower than expected and students in the second class interval performed as well as those in the third interval. However, the fit to the model given by $p < 0.032$ is reasonable.

**6.2.2 CCC and ICC for the *Four-Category Analysis of Environmental Concerns 2***

Figure 6.3 CCC for the *Four-Category Analysis of Environmental Concerns 2*

The CCC in Figure 6.3 show that the four categories work as intended for *Environmental Concerns 2*. 

172
Figure 6.4 ICC for the *Four-Category Analysis* of *Environmental Concerns 2*

Figure 6.4 shows that the observed means for the first two class intervals fit the model well, but the third one is higher than predicted, so students in that ability range are achieving a higher score than expected. The CCC for *Environmental Concerns 2* (Figure 6.3) is similar to the one for *Environmental Concerns 1* (Figure 6.1) with the probability of a score of 1 (responding at a multistructural level) being lower than the probability for the other scores.

For *Environmental Concerns 2* the span between the first two thresholds is smaller than for *Environmental Concerns 1* showing a smaller ability range.

6.2.3 CCC and ICC for the *Four-Category Analysis* of *Suicide 1*
Figure 6.5 CCC for the *Four-Category Analysis of Suicide 1*

The CCC in Figure 6.5 show that the four categories work for *Suicide 1*.

Figure 6.6 ICC for the *Four-Category Analysis of Suicide 1*

However, the ICC in Figure 6.6 shows the observed mean for the third interval is much lower than expected. Other means show a reasonable fit. The observed means for the first three class intervals are very similar showing that on this task less able students can score as high as the middle level students.

**6.2.4 CCC and ICC for the *Four-Category Analysis of Suicide 2***
Figure 6.7 CCC for the *Four-Category Analysis of Suicide 2*

The CCC in Figure 6.7 show that the four categories work as intended for *Suicide 2*.

Figure 6.8 ICC for the *Four-Category Analysis of Suicide 2*

However, the ICC in Figure 6.8 shows observed means for only the first three class intervals. The observed means for the second and third intervals are too high and low respectively; students with lower ability are likely to get higher scores than more able ones. The CCC for *Suicide 1* (Figure 6.5) and *Suicide 2* (Figure 6.7) are quite similar. The ICC are different with *Suicide 1* (Figure 6.6) having observed means below the curve, and *Suicide 2* (Figure 6.8) having an observed mean above the curve. The fits to the model are given by $p <0.000$ and $p <0.049$ respectively which shows that the observed means for *Suicide 2* fit the theoretical curve better than those for *Suicide 1*. 
6.2.5 CCC and ICC for the *Four-Category Analysis of Homeworkers 1*

The CCC in Figure 6.9 show that the four categories work for *Homeworkers 1*.

![Figure 6.9 CCC for the Four-Category Analysis of Homeworkers 1](image1)

However, the Item Characteristic Curve in Figure 6.10 shows that again there are observed means for only three of the class intervals. The first observed mean is lower than expected, the second mean is higher and the third quite close to the curve. The second and third observed means show that students of two different ability intervals are likely to gain the same score.

![Figure 6.10 ICC for the Four-Category Analysis of Homeworkers 1](image2)
6.2.6 CCC and ICC for the *Four-Category Analysis of Homeworkers 2*

The CCC in Figure 6.11 show that the four categories work as intended for Homeworkers 2.

The ICC in Figure 6.12 shows that the observed mean for the second interval is higher than would be predicted from the model; here students in the second interval for ability are scoring the same as those in the fourth interval.

The CCC for Homeworkers 1 (Figure 6.9) and Homeworkers 2 (Figure 6.11) are different between the first and second thresholds. For Homeworkers 1 the span is approximately six logits, for Homeworkers 2 the span is only about two
logits. For *Homeworkers 2* there is a smaller ability range expected to respond at the multistructural level. The ICC (Figures 6.10 & 6.12) are quite similar with one observed mean above the curve located around -1 logits on the scale.

The Information provided with the CCC and ICC for the six tasks is assembled in Table 6.2 to give an overview of the characteristics of the tasks.

**Table 6.2 Individual Item Fit for the Four-Category Analysis**

<table>
<thead>
<tr>
<th>Item</th>
<th>Location</th>
<th>Chi Square</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Environmental Concerns 1</td>
<td>1.715</td>
<td>0.0320</td>
<td></td>
</tr>
<tr>
<td>Environmental Concerns 2</td>
<td>0.197</td>
<td>0.0000</td>
<td></td>
</tr>
<tr>
<td>Suicide 1</td>
<td>-0.172</td>
<td>0.0000</td>
<td></td>
</tr>
<tr>
<td>Suicide 2</td>
<td>-3.155</td>
<td>0.0150</td>
<td></td>
</tr>
<tr>
<td>Homeworkers 1</td>
<td>0.771</td>
<td>0.0000</td>
<td></td>
</tr>
<tr>
<td>Homeworkers 2</td>
<td>0.644</td>
<td>0.0000</td>
<td></td>
</tr>
</tbody>
</table>

Table 6.2 his shows that this instrument distinguishes between all the categories of response, because there are no reversed thresholds.

![Figure 6.13 Threshold Map for the Four-Category Analysis with SOLO levels](image)
The threshold map in Figure 6.13 shows the thresholds in terms of the SOLO levels, and arranged in order of the mean locations. The respondent ability range is between negative four and positive four logits.

6.3 Summary

The aim of the study was to develop a common scale for rating qualitative responses to table interpretation tasks. The use of the Rasch Model led to the combinations of the initial categories that resulted in appropriate sets of categories for measuring the responses to all of the six tasks. These were found to be common for three tasks but different for the other three. An examination of the categories of the final analysis revealed that for one of the tasks, Suicide 1, the use of only four categories was appropriate. These appeared to be logical combinations based on the SOLO taxonomy based on the combination of unistructural and prestructural and with multistructural, relational and extended abstract without sublevels. The combination into four categories for all the tasks was completed and subsequently analysed using the RUMM program. It was found that this combination distinguished between responses without any reversed thresholds even though the discrimination between persons was not as good as combination of the categories for the instrument produced by the final analysis.

Thus it was shown that this four-category instrument could successfully be used across all of the tasks used in this study. Therefore, it is proposed that it could also be used to analyse students' responses to tables of similar
complexity in other contexts. It can also detect change in performance and this is reported in Chapter 7.
7.1 Introduction

One of the main aims of the study was to develop a workshop that would develop students' ability to read and interpret tables of data. Due to practical constraints outlined earlier in Chapter 3 the workshop was of one-hour duration. This was a short intervention, and thus unlikely to produce large changes in levels of response to the table interpretation tasks. However, the study was based on the assumption that students would respond to such tables in more depth after the workshop, and that this would be reflected in the higher levels of response as defined by the SOLO taxonomy. This expectation was based on the experience in the pilot study and the literature on learning mathematics and developing numeracy strategies cited in Chapter 2.

Another major aim was to construct an instrument to measure the levels of students' responses to the table interpretation tasks and to refine the instrument using the Rasch analysis. This instrument could then be used to measure the changes in the levels of the students' responses before and after the workshop. In this Chapter the changes in levels of responses are reported, to illustrate the extent to which the methodology developed can be used for this sort of purpose.

The initial construction of the instrument, consisting of criteria describing typical responses for the particular levels assumed, as is usually the case in
such a taxonomy, that the location of the levels on a common scale would be consistent across the tasks. This assumption was not supported from the empirical data obtained. The Rasch analysis showed that it was necessary to combine some of the categories tasks in order to achieve an internally consistent and valid instrument to distinguish between the students’ levels of responses that resulted in an instrument which did not have common locations on the scales for the six items.

The development and refinement of the initial instrument to become the final instrument was described in Chapters 4 and 5 including both the quantitative and qualitative adaptations to the original classification that make this possible. In addition, in Chapter 6, the development of a four-category instrument using a common set of four categories for all the tasks was reported.

The three analyses that located the ability of the students and the difficulty of the tasks on a single scale were called the initial analysis, final analysis and four-category analysis. These corresponded to the use of the initial classification of categories, the final classification of categories after combining various ones differently for different tasks, and the classification using four common categories for all tasks. Each of the three analyses gave estimates of students’ locations before and after the workshop.

In this Chapter comparisons of the before and after estimated locations are made for individual student locations and for the overall mean locations of
the tasks. These are made using the final analysis and the four-category analysis, but not the initial analysis, as this had two tasks with reversed thresholds. The comparisons of the before and after student locations is reported in Sections 7.2.1 and 7.2.2.

Firstly, for each analysis the comparisons are shown graphically in two dimensions referenced to a line of unit slope. Those points above the line show improvement, those on the line have not changed and those below the line represent students who have responded at a lower level on the after task. The large number of superimposed data points makes it difficult to see how many students were represented at each point on the scatter graph. However, the number of actual responses corresponding to locations above and below the line, were calculated from the data and are reported with the graphs.

Secondly, statistical comparisons of means using dependent t-tests for each analysis are made, with an hypothesised mean difference of zero. A comparison of the overall mean locations for the tasks is reported in Section 7.2.3.

The conditions for the completion of tasks before and after the workshop were kept as alike as possible for all the students. Therefore, although there could be other factors that influenced the responses, it was theorised that any change in the students’ ability scores would give at least some measure of the effectiveness of the workshop.
7.2 Measuring Change

In the initial stages of the evaluation each student response was allocated a score between 0 and 7, corresponding to the initial codes given in Section 3.6. The scores for each student were entered into a spreadsheet to be analysed using the RUMM software. Each student response was additionally coded according to whether they completed a task before or after a workshop and these data were used for the development of the measuring instrument. For the students who had completed both before and after tasks their scores before and after were entered on separate rows in the spreadsheet to allow a person estimate for each occasion. The RUMM program produced estimates of student locations and these were used to evaluate the change in responses and the effectiveness of the workshop.

It will be recalled that the students’ locations take into account the relative difficulties of the items so that the change in location of ability of a student represents a real change in the level of response. The responses of all students with a before and after response set are included in this analysis, including those extrapolated estimates for extreme scores, that is, a minimum score of 0 or the maximum score available on the particular tasks completed by the person.
7.2.1 Scatter graphs for the *Final and Four-Category Analyses*

Scatter graphs in Figure 7.1 and Figure 7.2 for the *final analysis* and the *four-category analysis* both have more points above the line of unit slope than on it or below it showing improvement using both classifications.

![Figure 7.1 Scatter Graph for the Final Analysis](image)

![Figure 7.2 Scatter Graph for the Four-Category Analysis](image)

Table 7.1 The Proportion of Students Who Improved After The Workshop

<table>
<thead>
<tr>
<th></th>
<th>Proportion who increased</th>
<th>Proportion who stayed the same</th>
<th>Proportion who decreased</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Final Analysis</strong></td>
<td>155/265 = 58%</td>
<td>0/265 = 0%</td>
<td>110/265 = 42%</td>
</tr>
<tr>
<td><strong>Four-Category Analysis</strong></td>
<td>181/265 = 68%</td>
<td>0/265 = 0%</td>
<td>84/265 = 32%</td>
</tr>
</tbody>
</table>
Table 7.1 show the proportions of students for whom the locations increased and decreased. None stayed exactly the same. The four-category analysis showed a greater percentage of students who improved but it is not possible to make direct comparisons between the two analyses as the range of scores available for each one is different. For the final analysis the maximum possible scores for different tasks are 3, 4, 5 or 6 and for the four-category analysis the maximum possible score is 3 for all tasks. For both analyses there was a majority of students who showed improvement.

It was surprising that 32%-42% showed a decrease. A small change in the response related to contextual or other factors is likely to have taken the response to a lower level of coding. Detailed investigation of the pairs of responses would identify reasons for the change.

In the scatter graphs in Figure 7.1 and Figure 7.2 it can be observed that for each of the scores for before the workshop there are several distinguishable scores, seen as separate dots in a vertical line, on the graph after the workshop. The students completed different tasks before and after the workshop that may have had different numbers of categories and when processed have a wider range of possible outcome locations which could account for some of the variation. In addition there is uncertainty associated with the estimates giving a range of location estimates.
7.2.2 Statistical Tests to Compare Location Estimates

The before and after estimated student locations calculated using the final analysis and four-category analysis were compared. A one-tailed, dependent t-test was conducted on the paired data comparing estimated locations for each individual student who had completed tasks both before and after the workshop (Table 7.2).

Table 7.2 Results for Dependent t-tests for Final and Four-Category Analyses

<table>
<thead>
<tr>
<th></th>
<th>Final Analysis</th>
<th>Four-Category Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Location Before</td>
<td>-0.213</td>
<td>-2.000</td>
</tr>
<tr>
<td>Mean Location After</td>
<td>0.418</td>
<td>-1.029</td>
</tr>
<tr>
<td>t statistic (df =263)</td>
<td>6.543</td>
<td>5.984</td>
</tr>
<tr>
<td>P-Value</td>
<td>&lt;0.000</td>
<td>&lt;0.000</td>
</tr>
</tbody>
</table>

For both of the analyses the difference between the before and after scores was statistically significant, which indicates an overall improvement in the level of response.

7.3 Changes in the Overall Mean Locations of the Tasks

The changes in the mean locations for the final analysis and the four-category analysis were examined using threshold maps because these allow a visual comparison which is easier than looking at a table. The thresholds maps for both analyses described in Chapters 4 and 5 are reproduced in Figures 7.3 and 7.4, each with vertical lines to show the locations of the means for the before and after locations, as used in the t-tests.
In Figure 7.3 the vertical lines are drawn at the before (-0.213) and after (0.481) mean locations.

Figure 7.3 Threshold Map for the Final Analysis

It will be recalled that in the threshold map the locations that are labelled 0/1 are for unistructural, those with 2a or 2b are multistructural, those with 3a and 3b are relational and those with 4a and 4b are extended abstract.

For all of the tasks the mean location before the workshop passes through regions that represent levels of at least multistructural, with three clearly in the relational. After the workshop the mean location is 0.694 logits further across on the scale. Two tasks, Homeworkers 1 and Environmental Concerns 1, are now located in a higher region and four have stayed within the same region. This indicates improvement in students’ average responses.
In Figure 7.3 the vertical lines are drawn at the before (-2.000) and after (-1.029) mean locations for the four-category analysis. The before line for the three tasks Suicide 1, Environmental Concerns 1 & 2 passes through the multistructural regions (2a/2b), for Homeworkers 1 & 2 it is borderline multistructural and relational (2a/b, 3a/b), and for Suicide 2 it passes through relational (3a/3b). For the location of the mean for after the workshop for all tasks the line passes through relational (3a/3b) except for Suicide 1 and Environmental Concerns 1 for which it still passes through multistructural.

The mean location of responses before the workshop is located in the regions of multistructural or relational. For all the tasks, except for Suicide 1 and Environmental Concerns 1, the mean responses are located in the relational category after the workshop.

Both analyses show an improvement in the estimates of students' ability to interpret tables. This illustrates that after the workshop students were
engaging more with the data in the tables; looking for trends and making comparisons, as well as looking for reasons for those variations. However, the analyses showed that not many students achieved the extended abstract level. Reasons for this finding are discussed in Chapter 8.

7.4 Summary

In this Chapter the changes between the before and after scores for the final analysis and four-category analysis were reported in three ways. Firstly, the pairs of scores were plotted on a scatter graph and the line of unit slope through the origin was superimposed. For both the final analysis and the four-category analysis it was observed that the students’ locations and the overall locations for the tasks showed improvement.

Secondly, dependent t-tests were conducted for the final analysis and the four-category analysis and the differences between the before and after scores for the students’ locations were found to be significantly higher after the intervention workshop, in both cases.

For the final analysis and the four-category analysis the mean locations of the majority of the student locations for before and after the workshop shifted from multistructural to relational levels. That is, that more students were proposing reasons for variations in the data after the workshop than before the workshop.
The analyses of the changes in the responses after compared with the responses before the workshop indicate that there is a significant improvement in the performance of the students after the workshop, despite the comparative brevity of the experience. This is an indicator that the workshop plays an important role in developing students’ skills in interpreting tables of data. The variations in the results can be attributed to a number of factors associated with the contexts of learning about the interpretation of tables of data and students’ responses carrying to the tasks. A discussion of the influence of these contexts is provided in Chapter 8.

The study shows that even with only a one-hour intervention a significant change in students' average responses can be achieved. In practical terms this is likely to be the extent of time that can be dedicated to this kind of explicit numeracy development within an existing unit, so confirming that this kind of intervention is of importance.
CHAPTER 8  UNDERSTANDING OF THE CONTINUUM OF DEVELOPMENT BY ANALYSING RELATIVE TASK DIFFICULTY

8.1 Introduction

The Rasch analysis of students’ responses to the tasks involving the interpretation of tables of data and the allocation of a mean score for difficulty for each task indicated that the six tasks had different levels of difficulty. It confirms also that in making comparison, it was important to account for those differences in difficulty when different tasks were used before and after the intervention. However in selecting the tasks these differences were not expected. This gives an opportunity to consider reasons for their differences and further develop the continuum of development, that is, what it means to have more of the trait. This Chapter identifies and discusses characteristics of these table interpretation tasks that are likely to have contributed to the difficulty of those tasks. It highlights the complex qualitative nature of the components that determine the difficulty of a task.

8.2 Review of the Study

It will be recalled that there were two main goals of the study. The first was to investigate the feasibility of developing an instrument to measure numeracy. The second was to investigate whether students’ ability to interpret tables of data could be improved by providing experience in an intervention workshop incorporating practical strategies based on the SOLO
taxonomy. An evaluation of the effectiveness of the workshop was to be based on the change in students’ performance on table interpretation tasks completed before and after the workshop.

The change in performance was reported in Chapter 7 where it was shown that there was a significant statistical difference between the students’ scores before and after the workshop. However, it is important not to over-interpret this result in terms of the effectiveness of the workshop, and to be aware that there are a number of factors that contribute to the performance of a student on a particular table interpretation task.

As described in detail in Chapter 3, there were six table interpretation tasks using three different tables of data, each with three subtasks. Each table of data featured in two tasks with two versions of the third subtask distinguishing them. The version of the third subtask was identified by the number with the title; Homeworkers 1 was the task using the table about Homeworkers using version one of the third subtask.

The students’ performances on these tasks was analysed using the Rasch Model. The Rasch analysis assigns a mean location of difficulty for each of the tasks and an ability location for each student. This chapter discusses the difficulties of the tasks in terms of their properties and the personal attributes of students that may have affected their success in reading the tables.
8.3 Conceptual Framework

Learning in any discipline involves both content and process components. The content component includes concepts and factual knowledge relevant to a domain and the process component relates to the strategies and skills needed to make sense of the concepts and factual knowledge. The aim of the workshop in this study was to develop critical numeracy skills that are important for analysis of quantitative content during study at university by providing students with a framework to develop table-interpretation skills. Reid (1998) reported that critical literacy is not adequately developed for students in many university courses. The direct links between critical literacy and critical numeracy are highlighted by Chapman and Lee (1990) who discuss the transformations between text and mathematical representations, and highlight the socially-learned nature of reading and writing.

To complete the tasks in this study the students were expected to understand how information is conveyed in a table of data. Tables are mathematical constructions, and arguably harder to understand than graphs because the trends are less evident (Wainer, 1992). In order to make sense of a table of data the students need to know how to work out what the numbers mean, how the components of a table are related to each other, and how to consider the data in a wider context. For some students this may be difficult, especially if they have not previously examined a table of data in depth. People tend to retrieve specific values from a table, as is appropriate in
looking at tables intended for access to particular information, and tend not to make comparisons or look at trends across time.

Chapman (1988) found that when university students encountered mathematical ideas embedded in non-mathematical readings the students read those mathematical ideas less precisely than they did for similar ideas placed in an obviously mathematical context, or did not read them at all. She concluded that there is a tendency for people to skim over or avoid tables of data in text, unquestioningly taking the author’s interpretation of the primary data.

The tasks in this study were designed to gauge students’ ability to interpret a table of data at a series of levels: starting with identifying the meaning of the numbers in the tables, then making comparisons and looking for trends, and finally thinking about the data in a wider context. The tasks were not intended to measure students’ recall of knowledge about specific content. Therefore, unlike many tasks that students are expected to complete in other parts of their university studies, the level of response was not directly dependent on learned knowledge in a discipline area. However, the students’ levels of responses could have been influenced by a variety of factors related to previous experience, the workshop structure and the tasks themselves. Chapter 2 highlighted how the extent to which people learn and transfer their knowledge, and the ways that they respond to mathematical or other tasks, is related to the context of their learning.
The Rasch analysis allocated a mean difficulty location on the scale for each task that was different for each of the six tasks. There are a number of possible factors that may influence the difficulty of the items and these need to be considered in the light of two meanings of the term context as identified by Wedege (1999). She refers to task context and situation context, where task context is the linguistic nature of the task and its relationship to its meaning. Situation context, by comparison, is concerned with historical, social and psychological circumstances.

In this thesis task context embodies the linguistic and structural aspects of the table interpretation tasks. These tasks required students to show their ability to interpret a table of data, with some accompanying text, by responding to three subtasks that are ordered by an increase in the level of thinking required in the response. The relationship between the language and meaning of a task is highly complex, based on shared understandings and knowledge, so students’ reactions to the tasks vary accordingly. The aspects of the task context of the tasks that might influence students’ responses in this study will be discussed in detail in Section 8.4.

The term situation context is used here to refer to a complex collection of factors that influence learning and performance. As suggested by Wedege (1999) these include the historical, social and psychological components, that students bring to a situation, as well as the context of the situation in which the students are learning or responding. Section 8.5 focuses on the cognitive
and affective components of the situation context that might influence the student performance on the table interpretation tasks.

### 8.4 Task Context

The purpose of this section is to analyse the task context contributing to the difference in difficulty of the six tasks, as shown by the final analysis. The tables and tasks were described in Chapter 3. The tables of data were selected by the researcher as being of apparent similarity in structural complexity to make the tasks approximately equivalent in difficulty. However, it was anticipated that the use of two versions of the third subtask would make a difference to the difficulty of the corresponding tasks. The final analysis showed that for version 1 the order of difficulty from least to most difficult was Homeworkers, Suicide, Environmental Concerns, while for version 2 the order was Suicide, Homeworkers, Environmental Concerns. Thus the Environmental Concerns task was the most difficult in both cases. The difficulty of the Suicide and Homeworkers tasks relative to each other varied. For version 1 the Suicide task was harder than the Homeworkers task while for version 2 the Homeworkers task was harder than the Suicide task. Therefore it was concluded that it was not only the versions of the subtask that made a difference between the tasks, but that there were other variables involved as well.

Consideration of other contributing factors was necessary to understand why the tasks are different in difficulty. In this Section the impact of the structure
of the tables (8.4.1) the accompanying text (8.4.2) and the language of the subtasks for both versions (8.4.3) are examined.

8.4.1 Structure of the tables

When tables of data used in this study were chosen by the researcher they appeared to be of similar complexity, they included longitudinal data with no nested information. However, the mean locations of the tasks as measured by the final analysis show that the six tasks vary in difficulty as shown in Table 8.1, where a lower number on the number line indicates a less difficult task.

Table 8.1 Mean Locations of the Tasks

<table>
<thead>
<tr>
<th>Task</th>
<th>Version 1</th>
<th>Version 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Environmental Concerns</td>
<td>0.735</td>
<td>0.966</td>
</tr>
<tr>
<td>Suicide</td>
<td>0.246</td>
<td>-1.092</td>
</tr>
<tr>
<td>Homeworkers</td>
<td>-1.262</td>
<td>0.404</td>
</tr>
</tbody>
</table>

It can be seen in Table 8.1 that the two Environmental Concerns tasks were the most difficult of the six tasks. The difference between the locations for the two versions is only 0.231 logits. For the other two tasks, Suicide 1 is more difficult than Suicide 2 by 1.338 logits and Homeworkers 2 is more difficult than Homeworkers 1 by 1.666 logits. This variation is likely to be associated with whether the tasks were completed before or after the workshop, as discussed earlier.

In the table for Environmental Concerns, the data were collected from an attitude survey, whereas the data for the other two tables were collected by other means. The students were expected to realise that the data represented
the percentages of the people surveyed who were concerned about a particular issue, and that the survey respondents could express concern about more than one issue. From students’ responses it was clear that some thought that the percentages represented people’s views on the relative severity of the environmental issues. This error may have been avoided with the inclusion of more explicit text with the table, although this would have changed the students’ experience and made it less authentic.

To help identify differences in the structure of the tables that could influence the difficulty of the tasks a summary of the features of the tables

*Environmental Concerns, Homeworkers* and *Suicide* was compiled (Table 8.2).

<table>
<thead>
<tr>
<th></th>
<th>Environmental Concerns</th>
<th>Homeworkers</th>
<th>Suicide</th>
</tr>
</thead>
<tbody>
<tr>
<td>Longitudinal time</td>
<td>4 ('92, '94, '96, '98)</td>
<td>2 ('89, '95)</td>
<td>2 ('88, '97)</td>
</tr>
<tr>
<td>Categories</td>
<td>15</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>Totals included</td>
<td>No</td>
<td>No</td>
<td>Yes: % and Number</td>
</tr>
<tr>
<td>Male, female, persons</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Data Points</td>
<td>60</td>
<td>48</td>
<td>48</td>
</tr>
</tbody>
</table>

It can be seen from Table 8.2 that the numbers of columns in the tables of data varies. The data for *Environmental Concerns* is given for the years 1992, 1994, 1996 and 1998, making four columns of data, whereas for *Homeworkers* and *Suicide* data was given for only two years of collection for each one. Therefore the students had to examine data for more years in *Environmental Concerns* than in the other two tables. However, the *Homeworkers* and *Suicide*
tables had added complexity with the inclusion of gender (males, females, persons) so there are six columns in all for both of those tables.

The categories are given in the rows of the tables and these also vary. The *Environmental Concerns* table has 15 categories of environmental issues starting with *Air Pollution* at the top of the table and going down to the *Use of Pesticides*. These are ordered according to the percentages in 1996, not to those in the first column for 1992. Some of the categories refer to several related environmental issues. The table has more than twice as many categories as each of the other two tables for *Homeworkers* and *Suicide*, but overall only about fifty percent more data points than each of the other two tables.

The two tables for *Homeworkers* and *Suicide* have some structural similarities both have data collected in two different years, even though only *Homeworkers* has the month as well; both have males, females and persons as column headings; and both have several categories (*Homeworkers* has eight categories while *Suicide* has six). Some non-common properties also contribute to the difficulties of the tables. When reading the *Homeworkers* table students need to recognise that ('000) at the top of the table means that the numbers in the table have to be multiplied by 1000. In the *Suicide* table most of the values are given as percentages but the column totals record the actual numbers of people. In the workshops, while completing the given tasks, some students asked for the meaning of *persons*, not realising that they
constituted the sum of the figures for males and females, suggesting an unfamiliarity with the kinds of tables of data used in the tasks.

In summary, the *Environmental Concerns* table, which the study found to be more difficult, has a more complex structure both in terms of data categories and time frames to be considered, than that of *Homeworkers* and *Suicide* tables. This is likely to contribute to making the *Environmental Concerns* task the most difficult for both versions of the third subtask. The difference in difficulty between the *Homeworkers* and *Suicide* tasks cannot be directly attributable to differences in structure because one is not consistently more difficult than the other.

### 8.4.2 Linguistic Components of the Tasks

Forman and Steen (1995) use the term *context* to describe aspects of word problems, examples, textbooks and teaching materials that are of a linguistic nature and which affect the meaning. Gal (2003) also highlights the influence of language on a task and the possible misunderstandings when words commonly used in everyday speech are used in a mathematical context. This also occurs in other situations where assumptions are made about how students will interpret the language of tasks or problems they encounter.

In this Section the linguistic components of the table interpretation tasks, that is the words used in the table of data, the accompanying text (8.4.2.1) and the language of the subtasks (8.4.2.2) are examined in the light of the potential
impact they have on the way students interpret the tasks and consequently the difficulty of the tasks.

8.4.2.1 Text of the Table of Data and Accompanying Text

An examination of the text accompanying the tables of data shows that the texts varied in the number of words and the type of information provided.

The text with the Environmental Concerns table stated:

People’s views and practices reflect the importance and priority given to environmental issues by society. The Australian Bureau of Statistics has conducted a household survey on environmental issues biennially since 1992. The reported levels of concern by Australians about particular environmental problems are summarised in Table 14.3.

A source of potential misunderstandings with this table was that the text did not state that people could identify more than one area of concern. Students could discern this if they add up the percentages but may not have thought to do so. The categories were clear except for other where it might have included all the other responses, or not.

The students found the task with the Suicide table easier than the task with the Environmental Concerns table. The text with the Suicide table stated:


There was no information about the categories, however they were easy to understand with the exception of three of them. These were other gases and vapours, when no mention of any gases or vapours had been made
previously, *other specified methods* and *other methods* which lacked clarity and could confuse students and affect their responses.

The text provided with the *Homeworkers* table, which was also easier then *Environmental Concerns*, was crucial to understanding the requirements to be a classified as a *Homeworker*. The text was longer than for the other two tasks.

The Australian Bureau of Statistics collects information on employed persons who work at home through the *Survey of Persons Employed at Home*. People who work any hours at home are included in the survey, but only those who usually work more hours at home than elsewhere, in their main or second job, are classified as employed at home. Here they are referred to as Homeworkers. Farmers, unpaid voluntary workers, people who work less than one hour and people who work from home but spend most of their time working away from home are not counted as Homeworkers.

The text with this table, taken from the article from which the table was extracted, seems to give more meaning and context than that from the text accompanying the other tables, but the detail is not easy to grasp on a first reading.

Some of the categories of *Homeworkers* were surprising; such as *plant and machine operators*, *and drivers* and *labourers and related workers* because they do not intuitively fit into the category of *Homeworkers* without some lateral thinking. Perhaps *machine operators*, with a higher proportion of females, includes sewing machinists and similar, and *labourers*, also with a higher proportion of females, could be packers or assemblers of some kind.

Another illustration of how information can be interpreted in different ways is related to the source of the data. Each of the three tables of data used in the
tasks had been produced from data produced by the Australian Bureau of Statistics and this source acknowledged for each one. The researcher interpreted this to imply that the data had been collected in a valid, unbiased way, but there were students who questioned the sampling methods in their responses to the tasks. However, this may not have been calling the validity of the data into question but showed an ignorance of how the sampling would have been designed to ensure a representative sample across the nation. The kinds of questions concerning the sampling centred around issues of where the sampling had taken place and the kinds of people who had been sampled.

The researcher had considered giving the students an article to read with a table embedded within it, but decided against it for two reasons. Firstly, the time constraints on the completion of the table-interpretation tasks within their normal workshops meant that it was not really a feasible option. Secondly, and more importantly, it is was thought likely that students would be influenced by the author’s comments and would lack confidence to draw their own conclusions. It is common that when people encounter a table of data in an article they are influenced by the potentially misleading written opinions of the author (Chapman, 1988). Gal (2003) emphasises the need for statistical literacy in such instances, especially when people are interpreting press releases where they need to decide what kinds of critical questions to ask.
8.4.2.2 The Language of the Subtasks

In this section the language of the subtasks are compared and this discussion does not involve a comparison between the tables of data. The subtasks were written by the researcher to give the students the opportunity, with some hints, to ‘tell the story of the table’. The way that the students interpreted and responded to the subtasks varied. As MacLachlan & Reid (1994) point out, ‘interpretation is clearly more than just an application of linguistic knowledge necessary to decode words and make connections between them and the sentences in which they were embedded’ (p.3). They also note that whenever people read they ‘draw on accumulated knowledge of the world, both experiential and textually mediated’ (p.3). Thus an analysis of the responses must acknowledge a wide range of experiences and not over-interpret the results in terms of the language of the tasks.

It will be recalled that the responses to the subtasks were given an overall score, not three scores for the separate subtasks. This was because it was assumed that in terms of responses the students who would respond at a level 3a, say, would include also responses covering levels 1 and 2. This was found to be the case with very few exceptions.

The first two subtasks were the same for all six tasks. The first subtask, Describe the main features of the table and how they are related, requested a descriptive level of response for information about the table, and all but a couple of students were able to give a reasonable response to this subtask.
Students referred to the subject matter of the table, and the time frames and categories.

The second subtask, *What conclusions can you draw from the table?* required information at a higher level of response. The word *conclusion* was included to signal that students should look for something more than in the previous subtask, something at a higher than descriptive level. It was expected that students would make some comparisons over time or within categories to gain some knowledge from the table, and to communicate this in their responses. The use of the plural, *What conclusions*, was intended to imply that they should look for more than one. A reasonable response to this subtask was achieved, at the multistructural level, for most students. This can be seen in Figures 7.3 and 7.4, where the mean location of students before the workshop is at a level 2a, or above.

The third subtask was provided in two forms that allowed the direct investigation of the influence of language on the level of response obtained. The first version of the third subtask was *Explain and discuss the information contained in the table* and the second was *What societal and environmental factors might account for the data in the table?* These subtasks are linguistically different.

The first version uses the term *explain*. The Australian Oxford Dictionary (1999, p.501) defines *explain* as (i) make clear or intelligible with detailed information, or (ii) by way of explanation or (iii) account for (one’s conduct,
etc). In this wording of the subtask it was intended that students should give reasons for the way the data changed over time for particular categories. (that is across a row) or the values of the categories differed within a particular year (that is down a column). The term *explain* is used by Biggs and Collis (1982) as one of the typical words used in university assignments to direct students to give a response at a relational or extended abstract level. Its use indicates a need to make links and give reasons. In this study many students who did not reach the relational level simply repeated their response to the second subtask as their response to the third subtask, or gave no response at all to the third subtask. Thus they were coded at the multistructural level of 2a or 2b rather than achieving 3a or 3b.

When the subtask was changed to the second version, *What societal and environmental factors might account for the data in the table?* the researcher hypothesised that this wording would change the difficulty of the subtask considerably. In the new form it suggests the students might give a list of environmental and societal factors in their responses. The term *What* in the subtask could imply that a list of examples is adequate, without the linkages and reasons required by the term *explain*. On the other hand, starting out with a list could have been the trigger for some students to give some reasons for the changes over time, thus making it less difficult than the first version. More students repeated the response to the second subtask in the response to the third subtask when they were asked to *Explain and discuss the information contained in the table* than when they were asked *What societal and
environmental factors might account for the data in the table? From this it can be inferred that students did not understand the difference between the requests for them to draw conclusions from the data and the request to explain and discuss the information contained in the table, so they repeated their response or did not respond to the third subtask. However when the third subtask was changed to What societal and environmental factors might account for the data in the table? they were able to provide a response.

The mean locations for the tasks are the means of all of the thresholds for each particular task. To investigate the situation more closely and to look for differences due to the change in the third subtask, the thresholds from the final analysis were examined and the location of the threshold between multistructural and relational was identified for each task. It was proposed by the researcher that this threshold is located where a student changes from thinking only within the table to also considering factors outside of the table. Therefore, a comparison of the locations of the thresholds between multistructural and relational for the different tasks would indicate an ordering of difficulty of the tasks at that threshold.

These thresholds were ordered by difficulty with the locations shown in Table 8.3. The standard error for each threshold gives an indication of the variation of the location of that threshold. For example, if one standard error is used to show variation this threshold for Environmental Concerns ranges between 0.362 and 0.516. The locations of the thresholds can be considered to be different if the ranges, using one standard error above and below the
mean threshold, do not overlap. It can be seen in Table 8.3 that the ranges show no overlap for the tasks for each version. The results (Table 8.3) show that for the pairs of tasks for *Homeworkers* and *Suicide* the threshold for the tasks with the third subtask version 1 were harder than the corresponding task with the third subtask version 2.

Table 8.3 Tasks Ordered by Location of Thresholds between Multistructural and Relational Categories

<table>
<thead>
<tr>
<th>Task</th>
<th>Threshold 2b to 3a: Version 1</th>
<th>Range: Version 1</th>
<th>Threshold 2b to 3a: Version 2</th>
<th>Range: Version 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Environmental Concerns</td>
<td>0.439±0.077</td>
<td>0.362 to 0.516</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Homeworkers</td>
<td>-0.432±0.109</td>
<td>-0.541 to -0.343</td>
<td>-1.079±0.115</td>
<td>-1.194 to -0.964</td>
</tr>
<tr>
<td>Suicide</td>
<td>0.908±0.144</td>
<td>0.764 to 1.052</td>
<td>-1.724±0.162</td>
<td>-1.886 to -1.562</td>
</tr>
</tbody>
</table>

For *Environmental Concerns* 2 the threshold between 2b and 3a is not identifiable because the rescoring process placed 2a and 2b and 3a into a category together.

### 8.5 Situation Context

The situation context encompasses historical, social and psychological aspects. All of these factors impact on the way people learn. The literature on situated learning referred to extensively in Chapter 2 explores the relationships between the context in which knowledge is acquired and the ability to operate effectively in other contexts (Lave, 1988, Masingila, 1993; Masingila, Davidenko & Prus-Wisnioska, 1996; Noss, Pozzi & Hoyles, 1999; Nunes, Schlieman & Carraher, 1993; Pozzi, Noss, & Hoyle, 1998). The literature also acknowledges the influence of the processes by which the
knowledge is gained (Boaler, 2000a, 2000b). These factors subsequently impact on a person’s ability to respond effectively to novel tasks.

In this Section factors of the *situation context* are identified as cognitive (8.4.1) or affective (8.4.2) and are highlighted to indicate potential influences on the performances of students on the table interpretation tasks.

### 8.5.1 Cognitive Domain

Cognitive factors that are brought to the tasks by the students are related to students’ prior knowledge and the ways that it has been acquired in different learning environments and situations. Given the diverse population of students in a university it is not possible to identify all the different ways in which students have gained their prior knowledge. Most of the Australian school leavers, those who come directly from school to university, taking part in this study, would have experienced similar approaches to the curriculum in the secondary schools they attended. There would have been some group work as well as whole class teaching, and a variety of assignment types. However, some would have studied a higher mathematics curriculum than others, perhaps giving them more confidence and familiarity with numbers. Some students would have studied subjects that made them more politically and socially aware of the influence of media and government on people’s lives. For the international and mature-aged students there would probably be a wider range of experiences that would contribute to their performance on the table.
For students to be able to respond to the table interpretation tasks at a relational level they would need to have some knowledge of:

- The linguistic meanings of the tasks and the structures of a table of data;
- The domain, or context, of the data in the table used in the task;
- The impact of the media on current issues;
- The ways in which social, environmental and economic pressures can affect people’s lives and
- The effect federal and state policies can have on the context of the data in the tables.

To be able to respond at an extended abstract level students need to be able to think in more abstract and theoretical ways, making connections between the local data to the global situation, and putting forward some theories based on their knowledge of the world.

The structure of the tables and the language of the tasks were discussed in 8.3, as part of the task context. The domains of the tables were selected as having some relevance to students, but the extent of the students’ knowledge and awareness of the domain of the table could be quite an important contributor to performance. For example, those students responding to the Homeworkers tasks who are parents of young children might be conscious of a reduction in childcare provisions in Australia, and so be aware of the widespread need for parents to combine work and childcare. Their responses would probably include aspects of childrearing as reasons for the increase in the number of Homeworkers in recent years. Other students might cite the
trend towards home businesses made possible through the increase in technology and access to the World Wide Web.

Students responding to the Environmental Concerns tasks would also have different levels of knowledge. There are students, perhaps those studying for University degrees in Sustainability, who would be aware of the change of levels of public concerns concerning environmental issues. They are likely to be knowledgeable about the Kyoto Accord and other international agreements. Those students who are members of protest groups might have specific knowledge of the potential dangers of some kinds of pollution over others, and be interested in the implications of the data in the table. Other students may be less interested or indeed perceive environmental issues as someone else’s problem.

Thus, the students brought their different knowledge and experiences to the tasks completed before the workshop. These differences surely contribute to the range in levels of the responses to the tasks, but as these background differences were not investigated no conclusions can be made about their influence on students’ performance.

The workshop provided some common, shared experiences and development of strategies for interpreting tables. Therefore it might be expected that there would be some similarities in the students’ responses. It should be noted that, even though all the workshops were conducted by the researcher, the students came with a wide range of backgrounds and
experience and interactions, and discussion varied from group to group. Given the overall improvement in the levels of the responses it can be inferred that the students were at least more aware of what was expected of them. There were similarities in the language of the responses but these were not analysed.

The situation is complex because some students were given the first version before the workshop and some were given the first version after the workshop. It could be that for some students the first version was too hard for them to interpret if they encountered it before the workshop. The second version gave more cues for the students to get started on the kind of response required. After the workshop students would have had more idea of what was expected from them. However, the Rasch model should sort this out as it orders the responses and the tasks independently.

8.4.2 Affective Domain

The literature examines affective variables from practical and psychological perspectives. Variables such as anxiety, motivation, confidence, emotions and so on have been investigated in a wide range of fields. In this study the impact of affective variables was not measured directly and so no judgements can be made without the data to support them. The comments here are made only after reflection on the results of the study and considering the differences in responses and will be confined to the factors of anxiety, confidence and disposition that are likely to directly impinge on students’ performance on the table interpretation tasks.
The table interpretation tasks in this study involve mathematical thinking; deciding which numbers to look at, understanding the magnitude of the numbers and the purposes of percentages, comparing relative sizes within categories and across time, and knowing where the changes are large or small. Large or unexpected changes in a series of numbers should alert the reader to look for some reasons for the change. The reasons for the change may not be part of the readers repertoire of knowledge but some suggestions of possible causes could be made, for example, relating to social change or media influence. The reader needs the disposition to examine the data to be able to interpret the table of data in any depth.

In this study those students completing tasks and the workshop were all pre-service primary teachers. Research in the USA has found that mathematics anxiety is very common amongst such students (Hembree 1990). Many mathematics educators (e.g. Sloan, Daane & Giesen, 2002; Evans, 2000; Stolpa, 2004; Stuart, 2000; Tobias, 1978, 1993) have found that a large number of people have mathematics anxiety and some even go on to develop math phobia. Therefore it seems very likely that some of the students participating in the study would suffer from mathematics anxiety to some degree.

The relationship between performance and anxiety is a complex issue. The OECD/UNESCO (2004) report on the results of the 2003 PISA study on secondary school students states that ‘[a]nxiety about mathematics was negatively related to performance in mathematics’ (p.12). Researchers (Hendel & Davis, 1978; Sarason 1987; Anton & Klisch, 1995) have found
connections between math anxiety and test anxiety that are quite complex and related to pressure from various sources to succeed. Furner and Berman (2003), in their review of the literature on math anxiety, ‘believe that as students become less anxious about and more confident in their abilities to do math, their performance on standardized test scores will improve and they will be better prepared for the future’ (p.170).

Although the students in this study were told that the tasks did not comprise a test it is possible that for some students at least there was an element of test anxiety in the process of responding to the novel mathematical task. As Evans (2000) points out some level of anxiety can be productive in helping students to rise to a challenge but too much anxiety is counterproductive.

When working with adults Gillespie (2003) found that for a wide range of skills their performance was related to their ease with the tasks. People’s responses to tasks ranged quite considerably for different contexts. Nunes, Schlieman and Carraher (1993) found that those people prone to error on tests were able to operate without error in practical, familiar contexts.

The development of confidence is often related to the social nature of their experience of learning. Many educators (Jaworski, 1994; Forster, 2000; Siemon, 2000) propose that understandings are socially determined, and that collaboration and interaction influence learning and performance. Students who have experienced negative interactions inside and outside classrooms may be less inclined to tackle novel tasks. In the workshop, the students were
encouraged to look closely at the data, to compare values and discuss the implications and develop their table interpretation skills in a non-threatening environment.

It is not possible to tell from the data collected to what extent the students were anxious in completing the first task because it involved some mathematics. However, by the time the students were asked to complete the second task, after the workshop, the students should have developed some familiarity with interpreting a table of data and should have felt more comfortable with completing the task. Again, using the Rasch model should pick this up with a better model fit for the after cases than for those before.

Another affective influence on students’ responses could be their reaction to the nature of the data that they were interpreting. The Suicide tasks include a table of data that shows changes in the numbers of people using the different methods of suicide over about a decade. There is an overall increase in cases over the ten years and this appears to reflect increased stresses in society. It was found through post task interviews that for some students this topic was particularly disturbing due to their experiences with the suicide of friends or family. A reluctance to engage with the data could have influenced the levels of their responses.

8.5 Summary

This Chapter identified components of the task context and the situation context for this study that were likely to have influenced students’ responses
to the table interpretation tasks. An examination of *task contexts* for the tasks indicated that the mean locations of the difficulty of the tasks would be influenced by the complexity and domain of the tables of data, the language of the subtasks and in particular the differences in language of the two versions of the third subtasks.

The *situation context* is a complex construct with both cognitive and affective components. In the cognitive domain prior knowledge and experience of learning were identified as contributing factors with anxiety, confidence and disposition playing a part in the affective domain.

The analysis of the data reported in earlier Chapters established that both versions of the tasks about *Environmental Concerns* were more difficult than the other tasks. Possible reasons for this have been suggested in this Chapter were concerned with the complexity of the table and the lack of information concerning the nature of the survey. The other two tasks had showed no consistency in order of difficulty, as this varied with the version of the third subtask. The differences in responses may have been due to affective factors, or to knowledge about the domains of the data. In a future study more information concerning students’ understanding of what they are being asked for would help to elucidate the linguistic components of the tasks.

The tables of data were selected for the students by the researcher who thought that the tables would be of about the same difficulty to interpret by the students. This post-facto analysis in this Chapter offers some hypotheses
on possible reasons for the relative difficulties of the tasks. This study did not isolate these structural and linguistic components, nor did it attempt to control cognitive and affective variables but the analysis suggests some variables worthy of further study.
CHAPTER 9 CONCLUSION

9.1 Introduction

In Chapter 1 of this dissertation identified many of the numeracy demands of people in their everyday life, community life, learning, work and recreation. In Chapter 2 the literature review outlined the ways in which the term numeracy is used in different contexts internationally, and identified other related terms used to describe the ability to interpret quantitative material. The review also highlighted the limited extent of published research in relation to developing expertise in interpreting tables of data and thus the need for more research in this area, and hence the appropriateness of the empirical study reported in this dissertation.

The empirical study was reported and discussed in Chapters 3 to 8. The main aims of the study were:

1. To construct a measuring instrument to evaluate students’ responses to tasks requiring interpretation of tables of data, and

2. To develop and conduct a one-hour workshop to improve undergraduate students’ reading and interpretation of tables of data

This chapter summarises the findings of the study and draws some conclusions in terms of teaching and learning at the tertiary level and further research.
9.2 Main findings of the study

This Section provides an overview of the findings of the study related to the measuring instrument, the tasks and the workshop.

9.2.1 The Measuring Instrument

The design of the measuring instrument was fundamental to this study for two main reasons. Firstly, it was necessary to investigate whether levels of responses to numeracy tasks could be consistently understood and reliably distinguished using a set of qualitative criteria. The literature review showed that whilst some studies had incorporated the use of levels based on a hierarchical taxonomy there was no evidence of a system of this kind being used to evaluate responses to table interpretation tasks.

Secondly, the measurement of change in performance on table interpretation tasks before and after the workshop, as measured by the instrument, was used to make some evaluation of the effectiveness of the workshop. Factors beyond the classroom between the students' completion of tasks before and after the workshop might have influenced the levels of students' responses but these were considered to be unlikely to have made a big impact as the time lines were only a few weeks. In a larger study it would perhaps be appropriate to consider such factors.

The measuring instrument consisted of a set of qualitative criteria developed for analysing students’ responses to the interpretation of tables of data tasks described in Chapter 3. The criteria were developed in conjunction with
designing the tasks and were based on the five main levels of the SOLO taxonomy. It included sublevels for the multistructural, relational and extended abstract levels of the taxonomy, making eight criteria in all. The levels were coded from 0 as the lowest to 7 as the highest level of thinking in the responses. These criteria were used to allocate a code to each of the students’ written responses to the table interpretation tasks, and the Rasch model was applied to the coded data. This analysis on the initial instrument has been described in the thesis as the initial analysis.

It was found in the initial analysis that for two of the table interpretation tasks there were inconsistent responses indicated by the presence of reversed thresholds, and that for all of the tasks there were very few responses at the lowest and highest levels. Therefore the two sets of extreme criteria were combined together and the students’ responses rescored. For three of the tasks this set of six levels of criteria worked well, but for the other three tasks additional rescoring, and combining of criteria were necessary to form the final instrument which could distinguish between the levels of response. The analysis on the final instrument has been referred to as the final analysis.

The final analysis of the data showed that the final instrument fitted the Rasch Model well. This demonstrated that a qualitative instrument based on the SOLO taxonomy could successfully distinguish between the responses of the students to the critical numeracy tasks. Whilst the development of this instrument was a strong achievement it was unexpected, and contrary to what might be inferred from the available literature, that the categorisation
of the criteria in the SOLO levels was not consistent across all the tasks. Whilst this instrument fitted the Rasch Model well it has its limitations in that there were differences in the qualitative categories for the different tasks.

An instrument using only four categories was used to investigate the possibility of constructing an instrument that had a consistent classification of criteria for all six tasks, and more potential to generalise to other tasks. This instrument used only the basic SOLO levels as criteria: prestructural and unistructural together, multistructural, relational, and extended abstract. A Rasch analysis of this instrument, called the four-category analysis and described in Chapter 6, showed that the instrument distinguished between the students and the thresholds were in the expected order. Thus, the instrument could be used successfully to gauge the responses of the students to all of the six tasks. It does not fit the model as well as using the final analysis but has the potential advantage of more general use.

Earlier discussion highlighted some of the issues associated with distinguishing between the students' responses at the sublevels of the SOLO taxonomy. It was pointed out that in the initial analysis the differences in the sublevels of the instrument were related to the number of distinct points made in response to the subtask. For example, at the relational level the difference between 3a and 3b is the number of reasons given for the variations in the table. Thus, a response at 3b is distinguishable from a response at 3a by there being more reasons, because they are quantitatively
rather than qualitatively different. Therefore, although more reasons may well imply a more thoughtful response it does not in itself mean a higher-level response. The combination of the sublevels makes sense when considered in this light confirming the suitability of the four-category analysis. Given that the four-category instrument worked for all six tasks in this study it should be transferable to other similar situations where this kind of measurement is needed, perhaps for diagnostic purposes.

The research showed that the criteria at the extended abstract level in the original set of criteria, used in the initial analysis, appeared to be at more sophisticated level than could be expected from most of the first year undergraduate students in a test situation. Few of them achieved at the level of extended abstract before the workshop although there was some improvement after the workshop. However, it was clear in the workshops, where there was an opportunity to discuss the inferences and implications of the data, that students were able to operate at the extended abstract level when the context of the material was familiar to them and there was more time to think about the data.

The researcher proposed that the complexity and wording of the tasks could make a difference to the ways that students respond to table interpretation tasks. The use of the two versions of the third subtask and the subsequent analysis indicated that the language of the tasks could make a difference in various ways. When developing criteria for use with such first year undergraduate students it is important to bear this in mind and to carefully
consider the purpose of the tasks and the context and language of those tasks.

9.2.2 The Tasks

The six tasks used in this study required students to read and interpret tables of data and were designed to elicit the highest levels of response that the students could present. The tables of data were selected as being of potential interest to the students and though to be of approximately equivalent complexity. They were selected so that interpretation would be neither trivial, nor excessively demanding.

There was a common format to the six tasks. For each task the students were required to interpret a table of data, with three subtasks to aid the process. There were two subtasks that were common for all of the tasks. These were: Describe the main features of the table and how they are related and What conclusions can you draw from the table? One subtask had two versions: Explain and discuss the information contained in the table and What societal and environmental factors might account for the data in the table?

The subtasks were constructed to encourage the students to read the tables and think about the implications of the data. They were intended to act as a guide without being too prescriptive, without giving too much direction. The contextual analysis of the tasks described in detail in Chapter 8 considered the potential effect of the language of the tasks and the context of the material on the students' responses. It was concluded that there were some
factors in the tasks themselves, as well as affective variables and students' backgrounds, that could make a difference to the levels of the students' responses.

9.2.3 The Workshop

In Chapters 1 and 2 included some definitions and descriptions of numeracy, with the need to develop table interpretation skills being highlighted. In this empirical study using intact groups in existing undergraduate units, it was only possible to design and implement a single workshop for the students in which they experienced processes aimed at developing their skills in reading and interpreting a table of data. The components of the workshop described in Chapter 3 were based on the research literature and were incorporated to make the best use of the single intervention. The study was not designed to measure the effects of these components directly so only the effect of workshop as a whole could be evaluated.

This evaluation was achieved by considering the change in individual students' levels before and after the workshop. The dependent T-test conducted on the paired before and after workshop scores indicated that there was a significant overall improvement in students' scores (p < 0.01). This result demonstrated that the workshop helped students to develop strategies to interpret tables of data.

The Rasch analysis on the students' responses showed that there was a range in the levels of response both before and after the workshop and that not
many students performed at the level of extended abstract. Hence the T-test results tend to mask the considerable variation in the changes in the levels of response. The changes were illustrated in Chapter 6 where the scatter graphs displayed the variation in the second tasks, given a particular response for the first task.

The differences in cognitive variation could have been due to a range of factors, including the context of the data in the table, the structure of the tasks and the wording of the subtasks, and the physical situations in which the tasks were undertaken. These influences were discussed in Chapter 6 where it was emphasised that the study did not measure these influences directly and that further research would be needed to investigate their impact on students' responses.

9.3 Limitations of the Study

There were limitations to this study. Some are because of the type of study and others because of the scope of the study. A list of these will lead into recommendations for further research.

A control group was not factored into the design because the workshop was incorporated into a core unit. The researcher considered that from a pedagogical perspective all students should be entitled to participate in the workshop. The absence of a control group as part of the study could cause concern in terms of validity and reliability. In order to compensate for this every workshop was conducted by the researcher, students were required to
complete the tasks in the same length of time and the coding scheme
associated with the measuring instrument was checked by cross marking of a
sample of the responses. In future research the possibility of a control group
should be explored.

There are a range of complexity of tables as defined by Mosenthal and Kirsch
(1998) and to be complete as a study students' responses to different kinds of
table tasks could be examined. In this study it was only feasible to use tables
of similar complexity but further research in this area would add more depth
to the field.

In considering the responses of students who recorded a lower response
level after the workshop than before there was no detailed examination and
comparison of the responses. An examination of the responses to identify
possible reasons for the different levels of response would be informative. In
a follow up study this could be done through careful examination of the
scripts and some interviews with the students.

An analysis of the levels of response in the light of gender differences or
maths background was considered at the time of the study and deemed to
beyond the scope of the study. However it is proposed to be part of
future/ongoing research.
9.4 Further Research

This section identifies areas of further research. There are a number of areas which merit further research, which have become evident as a result of the study. The discussion in Chapter 8 has indicated the potential effects of the language of the subtasks on the students' responses. The results of the study suggest it would be possible to identify the kinds of subtasks that elicit responses that most closely reflect the true competence of the student.

The workshops were productively conducted with a range of ages and backgrounds of first year students in the Social Sciences. There is the potential for follow up research incorporating the use of the Five Step Framework with primary and secondary students, and for students studying in the Biological and Physical Sciences at university.

In addition research undertaken to identify ways in which students can overcome their anxiety of mathematics in the context of reading tables of data would contribute to this field of study. It is important that students have the confidence to choose to engage with tables of data. Certainly it is also important that students can read other kinds of tables such as those found on medicine charts or food packets, but as was pointed out earlier this is a skill which is more often developed at the school level. The concern in the context of this study is with the development of higher order thinking needed to confidently and effectively interpret tables of data. This study has started the process of identifying the kinds of activities that could be
included in a series of lessons or workshops to build students' confidence in interpreting data presented in tabular form.

An investigation and analysis of gender differences in the responses to the interpreting tables of data tasks was beyond the scope of this study. However, the literature (e.g. Carey et al, 1995; Frankenstein, 1995, 1997; Furner & Berman, 2003) suggests that there are differences in learning styles and the kinds of activities that favour males and females. Given that many females exhibit anxiety with mathematics (Tobias, 1978, 1993) this is an area in which research would be particularly fruitful.

There is potential research to further refine the measuring instrument. The initial instrument fitted the Rasch Model well and it was evident that a measuring instrument based on the SOLO taxonomy was appropriate for evaluating responses to table interpretation tasks. However, the inconsistency across tasks made it less generalisable than was predicted initially.

The four-category instrument appears to be a reasonable prototype for a transferable instrument that measures students' ability to read tables of data. In addition, there is also the potential to modify the measuring instrument so that the links between the teaching using the Five Step Framework and the levels of the responses could be strengthened.

An investigation to examine whether the refined instrument is usable over a wide range of situations using tables of different complexity would contribute to the field. This work would be informed by the work of
Mosenthal and Kirsch (1998) considering the different difficulties of the tables in terms of structure and the density of the tables. This research would take into account the levels at which one might expect the students to respond depending on their mathematical background.

The next phase in the research would be to investigate the use of the Five Step Framework and measuring instrument in the area of graph interpretation. The literature, as outlined in Chapter 2, is more extensive than that for table interpretation but such work would make a contribution to this field.

9.5 Implications for Teaching

A significant achievement of this study is the demonstration that qualitative instruments can be constructed to measure levels of responses to the table interpretation tasks. These instruments serve several purposes. Firstly, they can be used as diagnostic instruments. Secondly, their structure would provide some insight for teachers and students into the kinds of responses that can be made when engaging with a table of data, and thirdly they could be used to evaluate different kinds of teaching and learning strategies aimed at improving students’ numeracy.

The results of this study suggest that formal instruction on strategies for interpreting tables of data leads to an increase in the ability to read complex tables. This ability can be learned, and it should not be assumed that people
would naturally acquire the appropriate skills without being exposed to suitable learning experiences.

Although this study focussed on the use of the Five Step Framework for reading and interpretation of tables of data, the steps are equally applicable to the interpretation of graphs, as illustrated in Lake & Kemp (2001). Therefore, it is proposed by the researcher that an effective approach for teaching critical numeracy would incorporate a series of workshops which examined a range of tables and graphs from different contexts, would give students valuable experience dealing with tables and graphs and develop their confidence in dealing with quantitative information in general. The students could initially be given a table or graph with the Five Steps Framework, and questions already constructed for the particular data. They would then be encouraged to develop their own analysis, making up their own questions for another graph or table, so that the process of interpretation would become a natural one.

The literature (e.g. Boaler, 1997, 1998, 2000a; Masingila et al, 1996, Adey & Shayer, 1994) have shown how important it is to develop students' skills using relevant, challenging and open ended activities. Whilst the study reported here incorporated as much of these as was possible in a single workshop there would be more scope in a larger study to investigate how students could develop their skills in reading and interpreting data in a range of ways given such kinds of activities. The present study suggests that further work of this kind is likely to be worth the effort.
APPENDIX 1

Samples of Students’ responses for Environmental Concerns 1, Suicide 1 and Homeworkers 1 to the following:

Describe the main features of the table and how they are related.

What conclusions can you draw from the table?

Explain and discuss the information in the table.

The samples are given for each level for 0/1, 2a, 2b, 3a, 3b, 4a where available. There were no responses at level 4b
<table>
<thead>
<tr>
<th>Level</th>
<th>Environmental Concerns 1</th>
<th>Suicide 1</th>
<th>Homeworkers 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>The table outlines environmental concerns in Australia during the time period of 1992 to 1998. These effects which vary are recorded in percentages. The information in the table allows one to come to the conclusion that environmental concerns and effects on the environment are very prominent and not very stable. It also highlights that people are becoming more aware that their actions may have an effect on the environment in some areas and not in others. The environmental concerns are listed on the left hand side. On the right in chronological order (with spacing of two years) we have the time period over which the information has been collected. Under this the data for each 2 year period has been recorded.</td>
<td>The table is showing comparisons of suicides between males and females and doing a total of average of female and male and representing it as persons. I draw the conclusion that more women committed suicide in 1998 and 1997. The table names a few methods of suicide such as poisoning, gases/vapours, hanging, firearms, and other methods the most popular was other methods, coming second hanging and then gases and vapours third. (note: these are incorrect conclusions)</td>
<td>One of the main features is the numbers of (m) and (f) persons who work at home. These relate to the persons column which gives the sum of (m) and (f) workers. The other main feature is the occupation they have. The number of people who work at home in 1989 in comparison to 1995. The information in the table informs me of how many (m) and (f) work at home in 1989 and 1995.</td>
</tr>
<tr>
<td>2a</td>
<td>The main features consist of what environmental concerns Australians are faced with a study every 2 years to see the effects of this. That since 1992 all of the percentage rates have dropped but some have fluctuated between the 2 year gap. It is a summary of concerns that householders were asked about and what is effecting what they are concerned about.</td>
<td>More males than females commit suicide</td>
<td>The table shows a variety of occupations and relates to the number of people who are employed to work at home for these occupations. The number of people per occupation is further specified, showing the numbers of males and females plus a totals column. The table give you the ability to analyse how figures have changed between 1989 and 1995. The majority of persons working at home has increased between 1989 and 1995.</td>
</tr>
</tbody>
</table>
**Level 2b**
The main features are different environmental concerns, and the degree of the concern (as a percentage) over time. Basically that people are less concerned about environmental issues in 1998 than they were in 1992. Although the types (areas of) concern have remained virtually in the same priority area. (For eg.) In 1992 40.2% of the responses were mostly concerned about air pollution as an environmental concern. A six percent drop was recorded in 1994, a further 4% by 1996 and concern rose by 1.5% by 1998. People are marginally less concerned about the use of pesticides than the destruction of trees.

**Suicide 1**
Compares the percentages of males and females that have suicided in 1988 and 1997. It also tells of the percentages that use a particular method of suicide. Much larger amount of males suicide than females. The suicide rate has risen from the 1988 to 1997 figures. The rate for hanging and strangulation has risen dramatically and in 1997 was the most favoured method. The information in the table shows the figures of males and females that have suicided in 1988 and 1997. There has been a significant increase between the two years with both the amount of males and females increasing. In 1988 the preferred method was hanging & strangulation and firearms however in 1997 the preferred method was both hanging & strangulation.

**Homeworkers 1**
The main features are: lists of varied occupations, specific times of different years, the two genders. We can compare the numbers of persons employed at home, males and females in different months of different years. From April '89 to Sept '95 the number of males employed at home has increased, the number of females fluctuates up and down according to the occupations. The overall percent of persons employed at home has increased within the six years.

**Level 3a**
Environmental Issues are broken down into 15 categories & the percentage of persons in Australia bi-annually that are concerned about the particular problem are listed in the 4 columns from 1992 to 1998. In the early 1990s people in Australia were more concerned with environmental issues & have become less so over the last decade. Not one issue has a higher concern over it from 1992 - 1998. The areas of concern seemed to follow a pattern though. Between 1996 and 1998 there was an upward trend in concern, people started to be concerned than they had been but not as much as in 1992 and 1998. In the early 1990s

**Suicide 1**
The main features are the source or type of method used to commit suicide, the percentage of male/female/persons who used that method and the years they did it. The main features relate to give an indication of what, who, when. The rate for suicide for males and females has increased. The method used has changed from firearms and explosives as the main method for males in 1988 to hanging and strangulation in 1997- this could be attributed to tighter gun control. For females the method has changed slightly from poisoning to hanging and strangulation in 1997.

**Homeworkers 1**
The table displays data about who works at home. Comparing 1989 data with 1995, and separating the information for males and females, plus into different categories of workers. More people are working from home in 1995 than in 1989. Clerks are the biggest homeworkers and are almost entirely female. Sales people have shown the largest % increase followed by professionals. It categorises people in the same boxes in 1989 and 1995, though there will have been movement of tasks across category. It shows an increase in homeworking in line with
<p>| Environmental and green issues had a lot of media coverage which could amount to concern - back in 1998 the effect were once again in the news with nuclear mining so may have caused the upward trend. | marked increase in suicide by hanging and strangulation and a reduction in firearms and explosives. Higher percentages of males commit suicide compared to females. There is a decreasing in poisoning in females. | improvements in telecommunications and computer use at home which enables more flexible manner with lower overheads. |</p>
<table>
<thead>
<tr>
<th>Level</th>
<th>Environmental Concerns 1</th>
<th>Suicide 1</th>
<th>Homeworkers 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>3b</td>
<td>Main features—how different things affect our environment. Pollution playing a very big part and although levels seem to have dropped substantially over time—they are still a concern with society. Obviously society has taken a more controlled interest in their environment—This could be through media, newspapers, schools etc but also because as a “whole” we seem to be much more concerned as to how our environmental issues can also be connected with future health of our society. Surveys have been conducted over time (’92 to ’98) about society’s concerns with the environment—it would appear our main concern is still to do with the environment in the area of pollution of all types.</td>
<td>Shows how male and female victims committed suicide, comparing 1988 with 1997. Also shows the total number of suicides for each sex group and complete group for each year. More males commit suicide each year than females. In 1988 a higher percentage of males committed suicide with firearms or explosives, more females committed suicide with poison. In 1997 the higher percentage of males and females committed suicide by hanging or strangulation. More people committed suicide in 1997 than 1988. Maybe there are more pressures on people in 1997 than 1988—resulting in people thinking that there is nothing better fro them than to die. Cut down in firearm/explosives suicides maybe due to the hard accessibility to get guns due to the new firearm holding regulations/permits.</td>
<td>List of various types of workers, 2 sets of dated data-comparable to each other, columns broken down 3 ways (male, female, person), title of table (explains what it is for). There has been an overall shift to more people working at home. A lot more female clerks and salespersons working from home than males and maybe far more women are employed in such positions in the first place. Technology has improved dramatically in the 6 years between ‘89 and ‘95 and as such has enabled more people to work from home (ie computer accessibility &amp; power). Manual type work has not changed dramatically to aid homeworking, evidenced by the only small increase in people.</td>
</tr>
<tr>
<td>4a</td>
<td>All these concerns are of man doing. Air pollution is the main concern for Australians as many live in cities. People rank their concerns in the order of how it affects them and media coverage. Air pollution, ocean and freshwater pollution affect people where they can see it and lower concerns like pesticides and uranium mining people can’t see the affects for many years. Australians concern for the environment has decreased over the years 1992-1998. In 1992 people’s concern for the environment was higher than in 1998 (and slowly dropped in 1994</td>
<td>No responses at this level</td>
<td></td>
</tr>
</tbody>
</table>
& 96). People are of the belief that any slight improvement in the environment is fine they then think everything in this area is OK & direct their energies elsewhere.
APPENDIX 11

Samples of Students’ responses for

Environmental Concerns 2, Suicide 2 and

Homeworkers 2 to the following:

Describe the main features of the table and how they are related.

What conclusions can you draw from the table?

Explain and discuss the information in the table.

The samples are given for each level for 0/1, 2a, 2b, 3a, 3b, 4a where available. There were no responses at level 4b.
<table>
<thead>
<tr>
<th>Level</th>
<th>Environmental Concerns 2</th>
<th>Suicide 2</th>
<th>Homeworkers 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>The main features are that this is all pollution but they are different types of pollution found in different quantities and they are all related in that they are all pollution methods.</td>
<td>No student responses in this category.</td>
<td>The dates 11/89 &amp; 9/95, the split males/ females &amp; combined columns. The occupations. They relate as a difference in (‘000s)</td>
</tr>
<tr>
<td>2a</td>
<td>The table is divided into two parts – the horizontal axis lists the years that the biennial have been conducted with corresponding levels of concern listed underneath for a range of environmental issues listed on the vertical axis. Air pollution rates highly on people’s concern for the environment with uranium/radioactive materials rating least in terms of definite problems.</td>
<td>The table records the total suicide rate for males, females and persons (as a percentage) in the years 1988 and 1997. The records given for male, female and persons suicide are all separate. The totals for suicide for each category ie males, females and persons was obtained and presented in the table. The total amount of female male and person suicide increased over the years from from 1988 to 1997.</td>
<td>Persons employed at home. Amount of females resp males and total amount of persons in the survey. Data taken from 2 time periods. Its more women home employed than men in most of the categories.</td>
</tr>
<tr>
<td>2b</td>
<td>The environmental concerns are largely based around different types of pollution. Each type of concern could be a result of another in which more environmental concerns are created. Pollution is a large concern to Australians in all its forms but as the years pass the concerns seem to reduce and then rise slightly. Air pollution is the greatest concern in people’s minds.</td>
<td>The table is broken up into six sub categories males, females and persons for both the years 1988 and 1997. Each method of suicide is shown in percentages for each group and the total number of deaths by suicide is also shown. The table shows that from 1988 to 1997 suicide by persons (males and females) has risen. Both the males and female groups increased in number. The table shows that on average poisoning, firearms + explosives + other methods decreased in number as the method of suicide. Suicide by gases &amp; vapours and hanging and strangulation increased.</td>
<td>It shows figures (in thousands) of the various occupations of people who are considered home workers. It has data for the years 1989 and 1995 which may be compared. At a glance more people work from home in 1995 than 1989. Far more female then male clerks are employed at home in both cases.</td>
</tr>
</tbody>
</table>
### 3a The table shows the different sorts of pollution and the percentage of Australian households concerned by each particular problem every two years beginning in 1992 and ending in 1998. Compared with 1992, in 1998 Australians are less concerned about environmental problems. This for all different types of environmental problems. The government must have taken steps since 1992 to try and control environmental problems. This must have worked and that is why Australians are not concerned about these problems in 1998 as much as they were in 1992.

### Suicide 2

The table shows methods of suicide that use these methods for males, females and a total in 2 time periods. That poisoning and firearms across all groups are less popular in 97 than 88. That hanging is much more popular in 97 than 88. That firearms and explosives are not as accessible in 1997 than in 1988 and therefore not used in as many suicides.

### Homeworkers 2

Data obtained from surveying males and females who work from home is provided from 10 different occupations both professional and non-professional. It compares the No’s of people working from home in these occupations in 1989 with the No’s in 1995. There has been both increase and decline in the No’s of persons working at home depending on the occupation. Professionals working from home have increased since 1989 whereas managers and Admin have decreased since 1989.

The fact that the high levels of technology in society today now allow people to have constant contact and interaction without attending the same workplace means that work can now be done from home.

<table>
<thead>
<tr>
<th>Level</th>
<th>Environmental Concerns 2</th>
<th>Suicide 2</th>
<th>Homeworkers 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>3a</td>
<td>The table shows the different sorts of pollution and the percentage of Australian households concerned by each particular problem every two years beginning in 1992 and ending in 1998. Compared with 1992, in 1998 Australians are less concerned about environmental problems. This for all different types of environmental problems. The government must have taken steps since 1992 to try and control environmental problems. This must have worked and that is why Australians are not concerned about these problems in 1998 as much as they were in 1992.</td>
<td>The table shows methods of suicide that use these methods for males, females and a total in 2 time periods. That poisoning and firearms across all groups are less popular in 97 than 88. That hanging is much more popular in 97 than 88. That firearms and explosives are not as accessible in 1997 than in 1988 and therefore not used in as many suicides.</td>
<td>Data obtained from surveying males and females who work from home is provided from 10 different occupations both professional and non-professional. It compares the No’s of people working from home in these occupations in 1989 with the No’s in 1995. There has been both increase and decline in the No’s of persons working at home depending on the occupation. Professionals working from home have increased since 1989 whereas managers and Admin have decreased since 1989. The fact that the high levels of technology in society today now allow people to have constant contact and interaction without attending the same workplace means that work can now be done from home.</td>
</tr>
<tr>
<td>3b</td>
<td>Main features are pollution, deforestation, waste disposal, &amp; resource use. All are combined as major factors which are either depleting or destroying the environment. Generally people are not as concerned in 1998 as they were in 1992 at the environmental issues posed above. This could be because the media is not hammering these problems in ’98 as much as ’92. People have become more blasé about the many issues of pollution and waste disposal. Increased population numbers or less media information could be a reason for this factor.</td>
<td>The table compares the percentages of suicides between males and females between 1988 and 1997 and also shows the preferences for different methods of suicide for males and females. The main conclusion we can draw from this table is that in 1997 more people suicided than in 1988; the number of men suiciding is higher than the female’s number and the preferred method have changed between the 2 dates. The number of suicide may have raised because of growth of population, more people living alone today; men suicide more than women</td>
<td>The main features of the table are the occupations and the dates. Males dominated the upper occupations in the table in 1989 but females slowly gained on males in 1995. There is a high proportion of clerks who are homeworkers and a low proportion of para-professionals. Overall there are a higher number of females in the homeworkers category. Kids I believe are the main factors, as the female (mother) would wish to stay at home and look after the children. Technology is also a factor as in 1989 the number of homeworkers was less than in 1995 where technology is improving and</td>
</tr>
<tr>
<td>Perhaps by recycling waste and cleaning up waterways people feel they have done all they can do.</td>
<td>because they are more exposed to the society’s judgements—e.g., a man without a job may feel distressed and ashamed.</td>
<td>the workforce has a greater choice.</td>
<td></td>
</tr>
<tr>
<td>Level</td>
<td>Environmental Concerns 2</td>
<td>Suicide 2</td>
<td>Homeworkers 2</td>
</tr>
<tr>
<td>-------</td>
<td>--------------------------</td>
<td>-----------</td>
<td>---------------</td>
</tr>
<tr>
<td>4a</td>
<td>The general trend is that the concern for the environment decreases every 2 years, however between 1996 and 1998 the concern increases – bar other pollution, other, nuclear weapons. The Australian public were obviously influenced in some way by a dramatic environmental problem and could more than likely be toxic chemical and hazardous waste disposal – this is because the difference between 1996 &amp; 1998 is much greater than any other issue. The factors could be the media, global warming, the ozone layer hole over Australia, overpopulation in other countries? Or Australia being used as a dumping ground for these wastes.</td>
<td>Table shows the number of suicides and the chosen methods of suicide by percentage in both male and females in a study taken by the ABS in both 1988 and 1997. The number of suicides, significantly by males, has increased from 1988 to 1997 and the methods in which these people choose to take their own lives have changed slightly. The choice of using a firearm/explosive has almost halved, possibly due to the introduction of a ban on holding firearms. Although this has meant that people have found another way of suicide as the total number of suicides has increased. The pressures to ‘do well’ in life are greater today than ever. We as a society have more choices and opportunities than ever and making these choices and taking these opportunities can be a daunting thing since there is so much pressure to succeed. No longer do we just have to graduate from school, marry, have 2 kids and settle in a job for a lifetime but we have to do Post grad studies, marry, own 3 properties by the age of 50, apply for a new job every 12 months or so permanent fulltime job is a thing of the past.</td>
<td>No student responses at this level.</td>
</tr>
</tbody>
</table>
BIBLIOGRAPHY


equity in mathematics education (pp. 93-125). New York: Cambridge University Press


DEST. (2004). *Numeracy across the curriculum*: Commonwealth of Australia. A project funded under the Australian Government's Numeracy Research and Development Initiative and conducted by Western Australian Department of Education and Training, Catholic Education Office and the Association of Independent Schools in Australia.


DEST. (2004). *Teachers enhancing numeracy*.


DETYA. (2000). *Numeracy, a priority for all: Challenges for Australian schools*: Commonwealth numeracy policies for Australian schools.


Dickersen, A., & Green, F. (Spring 2002). *The growth and valuation of generic skills: SKOPE research paper no. 26*: University of Kent at Canterbury.


Education Department of Western Australia. (1994). *Student outcome statements*. East Perth, Western Australia: Education Department of Western Australia.


Hill, P. (1999). Numeracy education: What do we know and what can we learn from the literacy experience. In *Numeracy, a priority for all: Challenges for Australian schools - Contextual papers*: AAMT/DETYA.


Morony, W. (2001). *Numeracy... a lot can be learned from our experiences with literacy, BUT...* Paper presented at the AAMT Virtual Conference.


Willis, S. (1996). Application for an Australian Research Council grant for collaborative research between the Education Department of Western Australia and Murdoch University.


