

Science and Mathematics Education Centre

Gender and Test Item-Response Formats

Barry Krueger

**This thesis is presented as part of the requirements for
the award of the Degree of Doctor of Science Education
of the
Curtia University of Technology**

January, 1999

ABSTRACT

The purpose of this study was to develop a better understanding of the patterns of science achievement for 154 ninth-grade girls and boys on multiple-choice and short-answer constructed-response items. The study was guided by a model, developed from an extensive review of the literature, incorporating the dimensions of generalised self-efficacy, item-specific self-efficacy and worry. These variables were operationalised through selected or specifically developed quantitative and/or qualitative research methods, and a series of equivalent multiple-choice and short-answer constructed-response achievement items was constructed for two different unit tests. The participants in the study rated their item-specific self-efficacies on 5-point Likert-type scales immediately before answering each of the achievement items, and they completed a series of worry items from Spielberger's *Test Anxiety Inventory* halfway through each test. Qualitative data were collected by surveying all the students and by interviewing selected students. The quasi-experimental analyses revealed the absence of any practically important gender-related differences in achievement for the multiple-choice and the constructed-response achievement items. However, the boys reported more item-specific self-efficacy and less worry than the girls for each of these item-response formats, and each of these gender-related differences was judged to be practically significant. The qualitative data provided additional evidence that the girls' self-perceptions of their efficacy for answering multiple-choice and short-answer constructed-response items was lower than that of the boys. It also provided support for the model underpinning the study. Overall, there was no evidence of any practically important interactions between gender and item-response formats, for either item-specific self-efficacy, worry or achievement, indicating that neither of the item-response formats used in the study, with this group of students, advantaged one sex over the other. Additionally, the findings from this study suggested that sufficient time should be allowed during testing so that all students can complete tests to the levels of their capabilities and that, during tests, the influence of students' self-efficacies is mediated through the quality of their engagement with test items.

ACKNOWLEDGMENTS

It is a major endeavour to undertake doctoral studies while working fulltime. Therefore, I thank the following people whose support has contributed significantly towards my completion of this thesis.

I am particularly indebted to my supervisor, Professor Lesley Parker. I thank Lesley for her assistance over these last four and a half years. Lesley has introduced me to the rigours of educational research, helped me think of myself as a researcher and helped me develop my scholarly abilities. In particular, I have appreciated the quality of her advice and the constructive nature of her criticisms.

I am also indebted to the staff of the Science and Mathematics Education Centre and the off-campus services of the T.L. Robertson Library for their very practical forms of assistance, which have helped me overcome many of the difficulties associated with studying at a distance.

I am grateful to the school where I am employed, the Year 9 students of 1996 and their science teachers. None of my research would have been possible without their generosity and their willingness to accommodate my requests.

Last, though not the least important, I acknowledge the patience of my wife, Bronwyn, and our children in the seemingly endless hours that I have been estranged from them.

TABLE OF CONTENTS

ABSTRACT	ii
ACKNOWLEDGMENTS	iii
LIST OF TABLES	vi
LIST OF FIGURES	viii
CHAPTER	
1 INTRODUCTION	
Background to the Study	1
The Framework for the Inquiry	5
Overview of the Inquiry	8
Significance of the Study	12
Overview of the Thesis	13
2 REVIEW OF THE LITERATURE	
Overview	15
Boys' and Girls' Patterns of Achievement According to Test Item- Response Formats	15
Factors Influencing Girls' and Boys' Format-Related Science Achievements	26
Gender, Affective Characteristics and Achievement	43
A Theoretical and Methodological Framework for Investigating Boys' and Girls' Format-Related Science Achievements	58
Summary	64
3 RESEARCH METHODOLOGY	
Overview	65
Research Design	65
Research Instruments	76
Procedures for the Collection of the Data	94
Summary	103
4 ANALYSIS OF THE DATA ... 1	
Overview	105
Context for the Analyses	105

The Content and Extent of Girls' and Boys' Self-Efficacies and Their Worrisome Cognitions During Tests	114
The Relationships Between Self-Efficacy, Worry and Achievement.....	135
Summary of the Findings	151
5 ANALYSIS OF THE DATA ... II	
Overview	153
Context for the Analyses	153
The Content of Boys' and Girls' Worrisome Thoughts and Their Self- Efficacies During Tests	156
The Relationships Between Self-Efficacy, Worry and Achievement	171
Summary of the Findings	185
6 DISCUSSION AND CONCLUSIONS	
Overview	186
Quality of the Study	187
Discussion of the Conclusions	190
Implications and Recommendations	205
Concluding Comments	212
REFERENCES	214
APPENDICES	
A DEVELOPMENT OF THE SELF-EFFICACY SCALE	243
B DEVELOPMENT OF THE <i>ATTITUDES TOWARD TESTS SURVEY</i> ..	252
C TEST BOOKLETS AND MARKING SCHEDULES FOR THE <i>EARTH SCIENCE TESTS</i>	260
D TEST BOOKLETS AND MARKING SCHEDULES FOR THE <i>ORGANISMS AND FOOD TESTS</i>	288
E DEVELOPMENT OF THE <i>SURVEY ABOUT SCIENCE TESTS</i>	317
F INTELLIGENCE QUOTIENT SCORES	325
G INTERVIEWING GUIDE	326
H FACTOR ANALYSIS OF THE WORRY DATA	328
I RESPONSE CHARACTERISTICS FOR THE <i>SURVEY ABOUT</i> <i>SCIENCE TESTS</i>	332
J CHARACTERISTICS OF THE EQUIVALENT TESTS	333

LIST OF TABLES

Table

3.1	Overview of the Research Plan	75
3.2	Overview of the Research Instruments	77
3.3	Overview of the Data Collections	95
4.1	Numbers of Students Included in the Data Analyses	106
4.2	Internal Consistency (Cronbach Alpha) of the Quasi-Experimental Data ...	109
4.3	Check-Recheck Reliabilities for the Survey Data	111
4.4	Types of Thoughts Experienced by Girls and Boys During Tests	115
4.5	Features of Test Items Judged as Easy	116
4.6	Features of Test Items Judged as Difficult	117
4.7	Relative Importance of the Major Response Categories for Boys' and Girls' Thoughts About Multiple-Choice Items	124
4.8	Relative Importance of the Major Response Categories for Girls' and Boys' Thoughts About Constructed-Response Items	125
4.9	Item-Specific Self-Efficacies for Science Achievement Items	127
4.10	Extent of Boys' and Girls' Worrisome Cognitions for Answering <i>Earth</i> <i>Science</i> Questions	131
4.11	Extent of Girls' and Boys' Worrisome Cognitions for Answering <i>Organisms and Food</i> Questions	132
4.12	Achievement on Test Item-Response Formats	136
4.13	Students' Responses When They Lose Confidence During Tests	140
4.14	Relationships Between Item-Specific Self-Efficacy and Achievement	141
4.15	Boys' and Girls' Item-Specific Self-Efficacies for Incorrectly Answered Items	142
4.16	Girls' and Boys' Item-Specific Self-Efficacies for Correctly Answered Items	143
4.17	Relationships Between Self-Efficacy and Worry	145
4.18	Boys' and Girls' Responses to Worrisome Situations During Tests	147
4.19	Relationships Between Worry and Achievement	148

4.20 Types of Changes That are Required to Improve Girls' and Boys' Achievements	150
B.1 Item Functioning Statistics (All Students)	254
B.2 Item Functioning Statistics for Boys and Girls	255
B.3 Principal Components Analysis of Worry Items (All Students)	256
B.4 Loading of Worry Items (All Students)	256
B.5 Principal Components Analysis of Worry Items For Girls and Boys	257
B.6 Loading of Worry Items for Girls	258
B.7 Loadings of Worry Items for Boys After Rotation of the Axes	258
F.1 Comparison of Boys' and Girls' Intelligence Quotient Scores	325
H.1 Principal Components Analysis of Worry Items (All Students)	328
H.2 Loading of Worry Items (All Students)	328
H.3 Principal Components Analysis of Worry Items for Girls and Boys	329
H.4 Loadings of Worry Items for Boys and Girls	330
H.5 Loadings of the Worry Items for Girls and Boys After Rotation of the Axes	331
I.1 Response Characteristics for Questions About Negative Thoughts	332
I.2 Response Characteristics for Questions About Positive Thoughts	332
J.1 Item-Specific Self-Efficacies for Equivalent Multiple-Choice and Constructed-Response Items	333
J.2 Worry Scores for Equivalent Multiple-Choice and Constructed-Response Tests	334
J.3 Achievement Scores for Equivalent Multiple-Choice and Constructed-Response Tests	335

LIST OF FIGURES

Figure	
2.1	Relationship Between Self-Efficacy Beliefs and Outcome Expectancy 46
2.2	Model for Boys' and Girls' Achievements on Equivalent Multiple-Choice and Constructed-Response Test Items 59
3.1	Comparison of the Test Booklets 66
4.1	Differences Between Girls' and Boys' Self-Efficacies for Multiple-Choice and Constructed-Response Questions by Research Group 129
4.2	Differences Between Boys' and Girls' Self-Efficacies for Multiple-Choice and Constructed-Response Questions by Unit Topic 129
4.3	Differences Between the Extent of Girls' and Boys' Worrisome Cognitions for Multiple-Choice and Constructed-Response Items by Research Group .. 134
4.4	Differences Between the Extent of Boys' and Girls' Worrisome Cognitions for Multiple-Choice and Constructed-Response Items by Unit Topic 134
4.5	Differences Between Girls' and Boys' Achievement for Multiple-Choice and Constructed-Response Items by Research Group 138
4.6	Differences Between Boys' and Girls' Achievement for Multiple-Choice and Constructed-Response Items by Unit Topic 138

CHAPTER 1

INTRODUCTION

BACKGROUND TO THE STUDY

The Underrepresentation of Girls in High School Science and Women in Science-Related Careers

Until quite recently, opportunities to study science, and therefore participate in science-related careers were denied to females and limited to upper class white males (D.R. Baker, 1998). For example, D.R. Baker (1998) reports that girls were barred from academic high schools in most European countries before the 1920s, few universities in the United States enrolled women last century and that women often worked as unpaid assistants in their husband's laboratories. Further, women were prohibited from conducting scientific research in the United States when they were first employed in universities and instead, they were assigned teaching positions in 'gender-appropriate departments' (D.R. Baker, 1998, p.870).

Gender-related disparities still exist in schools, academic fields and the labour market (D.R. Baker, 1998; Sonnert, 1996; Teese, Davies, Charlton & Polesel, 1995). For example, D.R. Baker (1998) reports that world wide, girls are less likely than boys to continue studying high school science when it is no longer compulsory. Additionally, although women comprise 40-50 % of university students in the United States and Europe, a much higher proportion of males than females choose to study science at this level (Osborn, 1994). As a case in point, Kahle and Meece (1994) cite evidence from the United States that women earn 30 % of the bachelor's degrees in natural science and engineering and only 21 % of the doctorates in these fields. The importance of issues associated with females' underparticipation in high school and university science is underscored by inequities in the labour market. Crockett's (1996) analyses of the 1991 national census data showed that, in Australia, women comprised 28% of the veterinarians, 26% of the chemists and less than 10% of the professionals who were employed in the 'hard sciences' (Crockett, 1996, p.266).

Recently, philosophical arguments about gender equity and pragmatic concerns relating to the work force have called into question the numbers and contributions of women in science (Kahle & Meece, 1994; Sonnert, 1996). In this context, the underparticipation of women and girls in science has been regarded as problematic nationally, in Australia and internationally for at least the last two decades (see for example; Australian Science Teachers Association, 1984; D.R. Baker, 1998; De Laeter, J. Malone & Dekkers, 1989; Kahle & Meece, 1994; Kelly, 1985; L. Parker, 1984; Sjoberg & Imsen, 1988). In Australia, the importance of issues relating to girls' participation in the school curriculum was highlighted by the Working Party on the Education of Girls, who reported in 1984 that:

The maximisation of a society's potential demands that all individuals be equipped with the confidence and competence to contribute to their fullest extent. All individuals should have equal opportunity to realise their potential and any limitation on access to necessary knowledge and skills on the ground of sex is not only socially unjust but contrary to the interests of Australian society. (Commonwealth Schools Commission, 1984, p.7).

Of the issues that have been identified, those concerning the underrepresentation of girls in high school science when it is no longer compulsory and women in science-related careers frame this study.

In the context of these issues, interventions in Europe, the United Kingdom, Australia and the United States have been implemented to address inequities associated with the schooling of girls. In Australia, the successes of these initiatives, which have involved curricular and pedagogical reform, have resulted in girls' steady 'colonisation' (Teese et al., 1995, p.V) of the physical sciences. Moreover, in the wider school context, it is popularly reported that boys are now the disadvantaged sex. For example, the Australian press has recently reported that boys are less literate than girls and that boys are more likely to drop out of school, fail and develop behavioural problems (Arndt, 1994; Milliner, 1998; 'Time for a new boy', 1998). However, in answering the question, 'Who wins at school?', Teese et al. (1995, p.V) assert that 'everything depends on the location in the curriculum which girls take up and how well they succeed at these locations.' They argue that the senior secondary curriculum is still a 'hierarchy of unequally structured opportunities' (p.107), and that it locates vocational benefits in the science and mathematics subjects that are taken mainly by boys. In this context, recent data from the Australian state of Queensland

show that girls comprised 61% of the cohort of students studying biology, 47% of the cohort studying chemistry and only 31% of the cohort studying physics in 1996 (Queensland Board for Senior Secondary School Studies, 1997). Clearly, girls, in spite of their other recent successes at school, are still underrepresented in the physical sciences, the subjects that provide access to higher status studies at university and male-dominated careers (Teese et al., 1995).

The Influence of Testing on Curriculum and Pedagogy

As noted in the previous section, curricular and pedagogical initiatives have been implemented to redress girls' underparticipation in high school science courses. However, it must be recognised that curriculum fidelity requires the congruence of assessment methods, pedagogy and curricular content (Gipps, 1994), and that '[educational] reforms most often founder on [a] failure to match assessment techniques with the intentions of reforms' (Collis, 1992, p.36).

In this context, commentators (for example; Crooks, 1988; Frederiksen, 1984; Gipps, 1994; Madaus, 1988; Neil & Medina, 1989) assert that one of the universally accepted facts about assessment procedures and testing is the influence they have on what is taught and how it is taught. These commentators argue that high stakes tests exercise a 'legitimizing power' (L. Parker & Rennie, 1998, p.898) on curriculum and pedagogy to the extent that teachers teach to the test, a tradition of past exams defines the curriculum, and instructional practices are influenced by the nature of the responses that are required from students in tests (Gipps, 1994; Madaus, 1988; Neil & Medina, 1989). It has also been suggested that changes to teaching styles and students' conceptions of subjects may be effected by changing the way that subjects are assessed (Newton, Tymms & Carrick, 1996). At the same time, some of these commentators have drawn attention to the need for tests to be equally valid and fair for all students, irrespective of background variables such as gender, socio-economic status and ethnicity. In this context, it is of importance to this study that much of the literature and the public debate relating to the influence of assessment practices is scathing in its criticisms of multiple-choice testing, particularly as it appears in the form of standardised testing in the United States (for example, Bracey, 1990;

G. Miller 1986; Neil & Medina, 1989).

Purpose of the Study

The relative test achievements of girls and boys, together with the fairness of school assessment procedures have received much attention in recent years from educators, researchers and parents (for example; Coorey, 1998; Gipps, 1994; Gipps & P. Murphy, 1994; C. Jones, 1995; 'Parents want role in assessments', 1995). Three aspects of testing have been identified as associated with gender-related differences in science achievement; namely, (i) the context in which test questions are written (for example, Bransky & Qualter, 1993; G. Erickson & L. Erickson, 1984), (ii) test item-response formats (for example, Bolger & Kellaghan, 1990; R. Murphy, 1982; Volkoff & Hocevar, 1995) and, (iii) whether tests are school-based or externally administered (for example, P. Murphy, 1991; L. Parker & Tims, 1995). It has been hypothesised by other researchers (for example; Rennie & L. Parker, 1991; Rosser, S. Brown, Greenberger, S. Johnson, Madaus, Welsh & L. Wolfe, 1989) that an over-reliance on assessment procedures that favour boys could have a negative effect on girls' attitudes, participation and persistence in science. Moreover, assessment procedures have been identified as exerting a pivotal role in informing girls and boys about their capabilities for pursuing particular courses of study (Rennie & L. Parker, 1991), and a powerful influence on what is learned and how it is learned (Cizek, 1993; Crooks, 1988; Frederiksen, 1984; Gipps, 1994; Madaus, 1988; P. Murphy, 1996). Therefore, it is important to develop a greater understanding of the assessment procedures that are used in high school science so that recent curricular and pedagogical changes to redress girls' underparticipation in science courses can be supported and promoted by equitable ways of assessing their science achievement.

Ample evidence has emerged showing that boys' and girls' relative academic achievements are influenced by test item-item response formats. Research in Australia and overseas shows that boys generally achieve relatively better than do girls when multiple-choice test item-response formats are used, and that the extent of these differences is either reduced or reversed when constructed-response formats are chosen (for example, Anderson, 1989; Bell & Hay, 1987; Bolger & Kellaghan, 1990;

Breland, Danos, Kahn, Kubota & Sudlow, 1991; D. Burkam & A. Burkam, 1995; Harding, 1981; Hellekant, 1994; Hoste, 1982; Mazzeo, Schmitt & Bleistein, 1991, 1993; R. Murphy, 1982; Teese et al., 1995; Volkoff & Hocevar, 1995). It is against this context that some commentators have advocated the use of performance assessments and constructed-response formats so that assessment procedures are less biased toward boys (Jovanovic, Solano-Flores & Shavelson, 1994; Mehrens, 1998). Clearly, a greater understanding is required for how these alternative response formats function, in comparison to the multiple-choice response format, so that all students can be afforded equal opportunity to demonstrate their knowledge and understanding of science content. The major purpose of this study is to contribute to that enhanced understanding.

THE FRAMEWORK FOR THE INQUIRY

Conceptual Framework

Three major hypotheses have been advanced by previous researchers (for example; Mazzeo et al., 1991) to explain boys' and girls' patterns of achievement according to test item-response formats; namely, (i) the influence of construct-relevant factors, (ii) the influence of construct-irrelevant factors, and (iii) the differential reliabilities hypothesis. This study takes the view that there are no appreciable differences between girls' and boys' cognitive skills and their cognitive styles, insofar as their achievement is concerned, on multiple-choice and constructed-response test formats. Instead, this study assumes P. Murphy's (1994) perspective that students' cognitive responses are mediated by their construct-irrelevant affective characteristics, and consequently, that boys' and girls' cognitive responses to test items are confounded with their affective responses.

Girls' and boys' perceptions of their self-efficacy and the worry component of test anxiety were identified in the review of the literature as two construct-irrelevant responses that could help to better understand their patterns of achievement according to test item-response formats. These variables have been shown to be sensitive to gender-related situational factors (L.G. Jones & L.P. Jones, 1989; Lenney, 1977;

M. Lundeberg, Fox & Puncochar, 1994; Hembree, 1988) and also related to academic achievement (Hembree, 1988; Multon, S.D. Brown & Lent, 1991; Seipp, 1991).

Moreover, social-cognitive theory attributes a powerful causal role to self-efficacy in relation to the enactment of behaviours (Bandura, 1984, 1986, 1997; Pajares, 1996, 1997). In this context, there is also some evidence that ultimately, girls' decisions to persist in science and mathematics courses and to enter scientific careers are linked to their self-efficacies (Hackett & Betz, 1989; Pajares, 1996; Zeldin & Pajares, 1997).

According to the model that was derived for the purposes of this study, boys' and girls' patterns of science achievement were conceptualised as dependent on their item-specific self-efficacies. These self-efficacy perceptions involve cognitive appraisals of situational variables at the level of the individual test item. Gender was conceptualised to have an indirect influence on girls' and boys' item-specific self-efficacies through their generalised self-efficacies for answering equivalent multiple-choice and constructed-response test items. Additionally, the extent of boys' and girls' worrisome cognitions during tests was posited to depend on the strength of their item-specific self-efficacies. Therefore, the model predicts that girls' and boys' patterns of science achievement on multiple-choice and constructed-response items may be illuminated by an examination of their generalised self-efficacies, perceptions of their item-specific self-efficacy and the extent of their worrisome cognitions for answering achievement items set in each of these response formats.

The Research Context

The participants in this study comprised all 154 Year 9 science students (73 boys and 81 girls) who attended a non-government school located in the Australian state of Queensland. The researcher was employed at this school as a teacher of science and religious education, and he was responsible for teaching one of the science classes that participated in this study. Most of the students were 14 years of age when the data were collected.

Although this school is non-selective with respect to students' 'abilities' (as demonstrated by previous achievement) the students who participated in this study were not representative of the general population of ninth-grade science students in

Queensland. Compared to Australians at large, a majority of the parents who send their children to this school enjoy higher than average household incomes, and an overwhelming majority have Anglo-Saxon backgrounds. Moreover, anecdotal evidence suggests that the parent body generally hold high expectations for the quality of their childrens' education.

Research Questions

The objective of this study was to develop a better understanding of the science achievements for the boys and girls from one school on multiple-choice and constructed-response items by addressing the following questions:

- Are there gender-related differences in the content and extent of students' self-efficacies for answering science achievement questions set in multiple-choice and constructed-response formats? And consequently: is there a gender-related interaction between girls' and boys' perceptions of self-efficacy on science achievement questions and the response format these questions require?
- Are there gender-related differences in the content and extent of students' worrisome cognitions for answering science achievement questions set in multiple-choice and constructed-response formats? And consequently: is there a gender-related interaction between boys' and the girls' worrisome cognitions on science achievement questions and the response format these questions require?
- How are the girls' and the boys' perceptions of self-efficacy and worry related to their achievement on multiple-choice and constructed-response test formats?

OVERVIEW OF THE INQUIRY

Methods of Inquiry

Research Design

This study adopted a 'binocular' (Reichardt & Rallis, 1994b) view of the research questions by using methodologies from the 'positivist' and the 'interpretivist-constructivist' research paradigms (Lincoln & Guba, 1985). As such, this study begins to address the 'missing links' (Krockover & Shepardson, 1995, p.223) in gender equity research by going beyond the reporting of quantitative differences and by moving toward the development of contextual understanding. Quasi-experimental and survey methodologies were used, in the positivist design, to test relationships between gender, test item-response formats, achievement and item-specific self-efficacy, and to describe the content of girls' and boys' generalised self-efficacies and their worrisome thoughts for multiple-choice and constructed-response items. Additionally, interviews were conducted within the interpretivist-constructivist design to clarify and extend the theory that was developed for this study by considering the subjective meanings that students attach to their generalised self-efficacies and their worrisome cognitions for answering multiple-choice and constructed-response test items. The researcher also maintained a journal throughout the inquiry, which documented the data collections and enhanced his capacity for reflection by recording his ongoing analyses of the interpretive data.

Procedures for Obtaining the Data

The methodological procedures for this study were implemented in three phases. The first phase focused on the development of research instruments for measuring the strength of boys' and girls' item-specific self-efficacies and the extent of their worrisome cognitions. The *Attitude Toward Tests Survey*, which was designed to measure the extent of girls' and boys' worrisome cognitions during testing, incorporated Spielberger's (1980) worry items from his *Test Anxiety*

Inventory within a framework of other items relating to testing. A 5-point Likert-type confidence scale was also developed to measure the strength of students' item-specific self-efficacies for answering individual multiple-choice and constructed-response test items. Each of these research instruments was piloted with students at the research site and modified in response to the pilot study data.

The second phase involved constructing and administering test booklets for the collection of the quasi-experimental data. Sets of equivalent multiple-choice and constructed-response achievement items were devised and arranged into four test booklets for each of the *Earth Science* and the *Organisms and Food* unit topics. To avoid the possible confounding influences of familiarity and order of the item-response formats, each test booklet contained only one of the equivalent forms of each item, and the order of the multiple-choice and constructed-response sections was reversed for half of the test booklets. The quasi-experimental data were collected from equivalent groups of boys and girls at the conclusion of each unit topic under standard examination conditions.

The final phase focussed on the collection of the qualitative data. The *Survey About Science Tests* was developed and piloted with a Year 10 class, and answered by all the Year 9 science students after the quasi-experimental data collections were completed. This survey instrument collected data relating to the content and influence of girls' and boys' perceptions of their generalised self-efficacy and their worrisome cognitions for answering multiple-choice and constructed-response test items, and it was also used to identify students for interviewing. The collection of the interview data was informed by F. Erickson's (1986) criteria for minimising researcher bias and ensuring the adequacy of the interpretive evidence. The interviews were conducted over two interviewing cycles according to a flexible and dynamic interviewing guide and situational cues. In all, 22 students were selected for interviewing according to their potential for reporting disconfirming evidence.

Procedures for Analysing the Data

The coarse-grained data from the positivist design and the fine-grained data from the interpretivist-constructivist research designs were analysed exhaustively.

The quasi-experimental data were aggregated and mean effect sizes were calculated for the boys and girls on each response format for worry, item-specific self-efficacy and achievement. Survey and interview data were coded and grouped into categories that emerged during the analyses. The relative importance of the response categories from the survey data was compared for girls and boys, whereas patterns of generalisations within the interpretive database were reported after a search for disconfirming evidence and the analysis of discrepant cases. Findings from each of the research traditions were discussed according to Bernstein's (1991) notion of 'dialogical encounter' (p.337), which involved a dialectical approach whereby each of the research traditions was regarded as a 'conversational partner' of the other.

Statement of the Findings

The outcome of the dialogical encounter between findings from the positivist and the interpretivist-constructivist traditions was reported as a series of seven working hypotheses, namely;

- (i) There are no substantial gender-related differences in the content of girls' and boys' self-efficacies for answering multiple-choice and constructed-response tests.
- (ii) Boys report relatively more self-efficacy for science tests than do girls, and the extent of these gender-related differences is not moderated by test item-response formats.
- (iii) There are no substantial gender-related differences in the content of boys' and girls' format-related worrisome cognitions for answering science tests.
- (iv) Girls perceive a relatively higher frequency of worrisome cognitions during science tests than do boys, and the extent of these gender-related differences is not moderated in any substantial way by test item-response formats.
- (v) There is no practically important interaction between gender and test item-response formats for boys' and girls' achievements on multiple-choice and short-answer constructed-response tests.
- (vi) The relationships between the variables in the theoretical model underpinning this study are supported.

- (vii) The quality of students' engagement with test items mediates the relationship between self-efficacy and achievement, particularly when time is limited.

Scope and Limitations of this Study

Factors associated with the gender equity debate, and in particular with gender fairness in testing are particularly complex. Elwood and Comber (1995, p.28) contend that an examination of these issues must take into account

the relative experiences of male and females, the expectations they and their teacher[s] have about what they are capable of, and the way in which assessors and examiners choose to assess and examine subjects.

The intent of this study was not to address all of these issues, but rather to focus on the ways that boys and girls are required to show their knowledge and understanding during science tests, together with factors that are associated with girls' and boys' self-perceptions of their capabilities. Although the etiology of gender-related differences in achievement is theoretically and practically important, such an undertaking is well beyond the scope of this study. Instead, this study extends the body of research knowledge in the field of gender and assessment through an examination of boys' and girls' self-efficacies and their worrisome cognitions for multiple-choice and constructed-response test formats.

This study has been limited to a comparison of multiple-choice and short-answer constructed-response formats — the item-response formats that tend to be most commonly used in high school science tests. Therefore, this study cannot comment whether multiple-choice or constructed-response formats present a relative advantage to either girls or to boys in comparison with extended-answer response formats or non-test assessments. Moreover, the data collections were subject to several limitations. First, students' perceptions of their self-efficacy and the extent of their worrisome cognitions were limited to the use of self-report measures due to the absence of suitable methods for collecting observational data. Second, the researcher acceded to requests by the cooperating teachers regarding the distribution of students to the research groups and the allocation of part marks for the constructed-response items.

This study has also been limited to all the Year 9 science students at one

particular school. As mentioned previously, these students were not representative of the general population of science students state-wide in Queensland, nation-wide in Australia, or internationally. Consequently, the typicality of the findings has not been established. Therefore, the findings from this study have been presented as a series of working hypotheses. Nevertheless, readers of this study should be able judge for themselves the applicability of this study for their context from the descriptions of the research setting, the participants and their instructional context. However, extrapolation of this study's findings to settings involving a wide cross-section of the Year 9 student population and to settings where parents do not hold high expectations for their childrens' achievement will need to proceed with some degree of caution.

SIGNIFICANCE OF THE STUDY

The significance of this study lies in the contribution that it makes toward issues that are important to teachers, educational researchers, policymakers, and ultimately, to women and girls. Recent scholarly activity has underscored the currency and importance of the issues described in this thesis. For example, in the five years since this study was conceptualised, Willingham and Cole from the Educational Testing Service in the United States of America edited the volume, *Gender and Fair Assessment* (1997). This review of the research literature incorporated several very recent studies that were completed during the currency of this research project. However, none of these recent studies investigated affective constructs as well as boys' and girls' achievements on stem-equivalent multiple-choice and constructed-response items. Moreover, in 1998, a combined symposium at the annual conferences of the American Educational Research Association (AERA) and the National Council for Measurement in Education (NCME) was convened to discuss issues relating to the use of multiple-choice and constructed-response test formats. Other recent scholarly activity has heralded advances in self-efficacy research and theory (Bandura, 1997; Pajares, 1996, 1997) and suggested ways of working with both of the research paradigms (Donmoyer, 1996, 1997; Moss, 1996).

Specifically, this present study is important from methodological, theoretical and educational viewpoints. First, this study models one way that perspectives from

the positivist and the interpretivist-constructivist research paradigms can be brought to bear on a single research problem. Although issues relating to generalisability may be problematic for other researchers who follow this approach, the faithfulness of the research to each of these paradigms and the richness of the findings, nevertheless, attests to the significance of the research design.

Second, this study is theoretically important because it provides a conceptual framework for understanding academic achievement in terms of self-efficacy and worry. It also contributes toward the research literature by providing specific descriptions for the interaction between self-efficacy and achievement-related behaviours during tests, and more generally, by relating gender to self-efficacy, test format, worry and science achievement. In doing so, this study provides additional support for the recently reported finding by Willingham and Cole (1997a) that there is no interaction between gender and item-response formats for achievement on parallel multiple-choice and short-answer constructed-response items.

Finally, this study is educationally significant because it addresses issues relevant to girls' participation in science courses and science-related careers. This study makes specific recommendations about the use of multiple-choice and short-answer constructed-response test formats for science classes, as well as for the administration of achievement tests. Perhaps more importantly, this study provides a counterpoint to the commonly held view that girls now have the upper hand at school and that boys are the disadvantaged sex. The findings suggest, that at least for the students involved in this study, interventions are required to address gender differences that continue to favour boys in science classes.

OVERVIEW OF THE THESIS

This first chapter has set the scene for the remainder of this thesis. It has located the purpose of the study within a wider gender equity and educational framework, and it has also introduced the contents of this thesis. Chapter 2 surveys the literature relevant to the research problem. It reviews girls' and boys' patterns of achievement for multiple-choice and constructed-response items, together with factors that may impinge, differentially, on their test format-related science

achievement. The review of the literature culminates in the construction of a theoretical and methodological framework, from which this study's research methodology was derived. Chapter 3 outlines the methods of inquiry. The research design, characteristics of the research instruments and the research context are described, together with the procedures that were used to collect the data. Chapters 4 and 5 present the analysis of the research data — Chapter 4 focuses on the coarse grained data that were collected in the positivist design and Chapter 5 on the fine-grained data that were collected in the interpretivist-constructivist design. The final chapter integrates and discusses the findings from each of these data analyses, and it highlights the methodological, theoretical and educational implications of this study.

CHAPTER 2

REVIEW OF THE LITERATURE

OVERVIEW

This chapter reviews the literature relating to girls' and boys' patterns of academic achievement on test item-response formats. Its purpose is to produce a theoretical and methodological base for this study.

The review is divided into four sections. The first section examines boys' and girls' patterns of achievement according to test item-response formats. Hypotheses to account for gender-related differences in achievement on multiple-choice and constructed-response test formats are identified in this section and appraised. The second section identifies differences between girls and boys relevant to the most likely of these hypotheses; section three conceptualises the relationship between these variables and academic achievement; and the final section integrates these variables into a model as well as developing a methodological base for this study.

BOYS' AND GIRLS' PATTERNS OF ACHIEVEMENT ACCORDING TO TEST ITEM-RESPONSE FORMATS

The Nature of Assessment and Testing

The under-participation of girls in science and of women in scientific careers was linked, in Chapter 1 to assessment-related issues. Assessment, in the wider sense of its meaning, refers to the collection of information on the performance of individuals, groups, programs and schools, as well as to selection, certification and accountability processes (Gipps, 1994). It is important, for the purposes of this study, to make clear the distinction between assessment and testing. Although assessments may involve tests, the epistemology and purposes of assessment and testing differ (Gipps, 1994). Essentially, formative assessments are employed to support teaching and learning whereas summative procedures, like tests, are intended predominantly for making judgements about students' achievements (Doran, Lawrenz & Helgelson,

1994; Ebel, 1979; Gipps, 1994).

As introduced in Chapter 1, test item-response formats are linked to gender-related differences in achievement, which are observed across a wide range of curricular areas (for example, Bolger & Kellaghan, 1990; Mazzeo et al., 1991; R. Murphy, 1982; Volkoff & Hocevar, 1995). The literature categorises test formats according to the type of response they require. Selected-response formats require students to choose an answer from among a series of alternatives (for example, true-false, multiple-choice and matching questions) whereas constructed-response formats require students to generate their responses by supplying a short answer such as a word or phrase, or by writing an extended-answer in the form of a paragraph, essay or multi-step calculation (Bennett, 1993; J. Powell & Gillespie, 1990). Another approach (for example, Anderson, 1989; Ebel, 1979; R. Murphy, 1982) classifies test formats as either objective or non-objective according to the amount of subjectivity involved in scoring students' responses. For example, matching, multiple-choice and short-answer questions are regarded as objective tests because they can be reliably scored from a simple answer key (Ebel, 1979; Neil & Medina, 1989; J. Powell & Gillespie, 1990). This review of the literature examines multiple-choice, short-answer and extended-answer test formats, which represent the test formats that are used most often in high school science programs.

The administration of tests and the interpretation of their scores are founded on several assumptions. The first assumption is that of universality (Gipps, 1994), whereby it is assumed that test scores have the same meaning for all individuals. However, tests do no more than sample particular aspects of students' knowledge and their understanding (Childs, 1990; Gipps, 1994; J. Powell & Gillespie, 1990; R. White & Gunstone, 1992). Consequently, testing must be regarded as a subjective procedure because it involves arbitrary decisions about what to test and how to test. The second assumption is that of unidimensionality (Gipps, 1994). Traditional methods of testing assume that students' achievement scores reflect their knowledge and understanding of the subject matter (Gipps, 1994; P. Murphy, 1988). However, the literature is unanimous in its reports of the invalidity of this assumption. For example, not all students share with their assessors the same cultural and out-of-school experiences, and the same meanings for language (see; Archenhold, 1988;

G. Erickson & Farkas, 1991; S. Johnson, 1987; Kahle & Lakes, 1983; Neil & Medina, 1989; Sjoberg & Imsen, 1988). Consequently, individual students or groups of students may perceive questions and question cues in different ways (Bracey, 1990; P. Murphy, 1988). Thus, P. Murphy (1991) asserts that valid interpretations of test scores involve understanding the characteristics of individual learners and how they interact with tasks.

In general, it is difficult to ascertain if tests are biased toward any particular group. Group differences in achievement may reflect real differences in attainment, or they may reflect differences between groups in their opportunities to gain access to educational opportunities or demonstrate their understanding (Gipps, 1994; Gipps & P. Murphy, 1994; Willingham & Cole, 1997b). Although the term 'test bias' is sometimes used in the literature to describe situations where groups with the same abilities achieve different test scores (Childs, 1990), the use of this term is avoided in this review because of its emotional connotations (Flaughter, 1978) and its simplistic approach to complex educational issues (Gipps & P. Murphy, 1994). Instead, testing procedures that highlight inequities of opportunity to show learning are best regarded as issues of validity rather than test bias (Gipps & P. Murphy, 1994; Scheuneman, 1991; Willingham & Cole, 1997b). It is important to recognise in this context that the validity of tests and assessments is not contingent on whether disparate groups have the same access to educational opportunities (Willingham & Cole, 1997b).

Achievement of Girls and Boys on Test Item-Response Formats

Evidence that boys achieve relatively higher than girls on multiple-choice test items was first reported from analyses of data from large-scale, public examinations conducted in Great Britain. For example, Wood (1978) reported that a relatively higher proportion of boys than girls passed an O-level General Certificate of Education (GCE) examination in English after multiple-choice comprehension items were introduced in 1972. This shift in achievement patterns was detected also for subjects other than GCE English and examinations conducted by the University of London Board (R. Murphy, 1980, 1982; Wood, 1978). R. Murphy (1980), for example, reported a similar finding for boys and girls on the GCE O-level Geography

examinations set by the Associated Examining Board. When multiple-choice questions were introduced into the Geography examination an additional 10% of the boys relative to the girls achieved A – C grades. R. Murphy (1982) also compared the performance of girls and boys on other large-entry GCE examinations. He reported that the male candidates achieved relatively higher scores than the female candidates on multiple-choice items between 1976 and 1979 after boys' and girls' achievements on the non-objective questions were taken into account. This pattern was consistently true for 8 of the 16 subjects over the four years studied by R. Murphy. Moreover, the relative improvement in the boys' scores for these subjects varied up to a 7% advantage over the girls. However, he detected no advantage for the males on the multiple-choice items in mathematics. Although item-by-item analyses in the other subjects showed slight variations in girls' and boys' response patterns, none of these differences could explain the relative advantage to the boys on the multiple-choice response formats.

Additionally, Harding (1981) reported gender-related differences in achievement for six GCE science examinations conducted in 1974. Although she found no significant differences between boys' and girls' pass rates, there were differences favouring boys on two of the four multiple-choice sections, no overall differences on the structured-answer sections, and an advantage for girls on the essay-type questions. Additional evidence that boys perform relatively better than do girls on multiple-choice test item-response formats, from both Australia and overseas, has been reported elsewhere (for example, Anderson, 1989, Bell & Hay, 1987; Bolger & Kellaghan, 1990; Breland et al., 1991; Bridgeman & Lewis, 1994; D. Burkam & A. Burkam, 1995; Hellekant, 1994; Hoste, 1982; Mazzeo et al., 1991, 1993; Smail & Kelly, 1984; Stumpf & J. Stanley, 1996; Teese et al., 1995; Volkoff & Hocevar, 1995) though some researchers report that this effect is small (for example, Bransky & Qualter, 1993; Whitehouse & Sullivan, 1990). However, although Stobart, Elwood and Quinlan (1992) report a weakening of the 'multiple-choice effect,' evidence from the United States and Sweden suggest that this effect is as robust as ever. Breland et al. (1991) reported that there were no gender-related differences on the constructed-response sections of the 1986 Advanced Placement history examinations, but that there were practically important gender-related differences on

the multiple-choice questions. Additionally, Hellenkant (1994) determined that multiple-choice items advantaged upper secondary school boys in Sweden by an average of 8% in tests of English proficiency between 1986 and 1993.

Overall, there is ample research evidence for the existence of gender-related differences in achievement according to test item-response formats. Further, although most of this research has utilised data generated by large-scale external examinations for senior secondary school-age and college students, there is also some evidence from the English-Language Arts examinations of the California Assessment Program to indicate that similar effects exist among representative samples of eighth grade students (Volkoff & Hocevar, 1995). Otherwise, there is little data to establish either the presence or absence of age-related trends. Furthermore, the literature does not report if these gender-related differences are moderated by the cognitive complexity of test items.

Hypotheses to Account for Boys' and Girls' Patterns of Achievement on Multiple-Choice and Constructed-Response Items.

Overview of the Hypotheses

There is some evidence to indicate that girls' and boys' achievements on multiple-choice response formats are influenced by the consistent operation of all-pervasive factors. For example, Mazzeo et al. (1993) reported that gender-related differences in achievement on the multiple-choice sections of various Advanced Placement examinations persisted when items that functioned differentially for boys and girls were removed. The literature (for example, Mazzeo et al., 1991) advances three hypotheses relating to the influence of all-pervasive factors; namely, (i) the differential skills hypothesis, (ii) the existence of gender-related differences in construct-irrelevant factors and, (iii) differences in the reliabilities of the various item-response formats.

The Differential Skills Hypothesis

The differential skills hypothesis attributes disparities between girls' and boys' achievements to gender-related differences in the academic and subject matter specific competencies required by each of the item-response formats. Debate about the equivalence of constructed-response and multiple-choice response formats has centred around the issue whether these response formats restrict the nature of the content and the processes that can be measured (Bennett, 1993; Frederiksen, 1984). One argument asserts that 'Any [form of] assessment anoints certain characteristics, ways of being, and kinds of information, while underplaying, ignoring, or outright suppressing others' (D. Wolf, 1993, p.213). Hence, it is popularly reported that questions set in multiple-choice formats measure the recognition and recall of factual information, together with test-taking skills whereas constructed-response test formats are alleged to measure higher order outcomes (for example; Bracey, 1990; Guthrie, 1984; Neil & Medina, 1989). However, Mehrens (1998) contends that this simplistic and popular view has little currency among true measurement specialists.

Instead, Ebel (1979) and Frederiksen (1984) are amongst those measurement experts who assert that not all test formats can be used with the same ease and effectiveness to measure the complete range of cognitive processes. Frederiksen (1984), for example, reports that it is easier to cast multiple-choice items into a free-response format than to cast free-response items into a multiple-choice format because multiple-choice items are generally written to test lower order outcomes. Additionally, cognitive theorists assert that multiple-choice questions contribute information on general cross-contextual components of cognitive skills, whereas constructed-response questions, which are more closely tied to their context, measure situated cognition (Bennett, 1993). However, although a theoretical distinction can be made between the requirements of these two item-response formats, there is a paucity of empirical evidence to support the practical importance of this distinction. For example, Hogan (1981) concluded from his review of the research literature that the various test item-response formats were equivalent insofar as the measurement of knowledge is concerned, but Frederiksen (1984), although concurring with Hogan's findings for the measurement of knowledge, concluded that the test item-response

formats were not equivalent for complex problems. Recently, a more rigorous and exhaustive review reported that the evidence is too contradictory and too inconclusive to give trustworthy answers (Traub, 1993). Although Traub (1993) posited that multiple-choice and constructed-response formats may measure different constructs in some domains and not in others, he concluded that if any differences were to be found that they would be small. This viewpoint is consistent with analyses of data from Advanced Placement examinations (Bennett, Rock & Wang, 1991; Lukhele, Thissen & Wainer, 1994; Thissen, Wainer & Wang, 1994) and the National Educational Longitudinal Study of 1988 (Hamilton, 1997). The findings from these analyses show that an overwhelming majority of multiple-choice and constructed-response questions require the same processes, and that where small differences exist they are subject to differing interpretations (for example, Bennett et al., 1991; Thissen et al., 1994). Hence, it seems unlikely, given this contemporary evidence that boys' and girls' patterns of achievements on multiple-choice and constructed-response test formats can be explicated by the differential skills hypothesis.

The Influence of Construct-Irrelevant Factors

Whereas the previous hypothesis focussed on cognitive skills relevant to the solution of multiple-choice and constructed-response test questions, this hypothesis proposes that multiple-choice and constructed-response test scores are influenced by sources of variation that are related to gender but unrelated to academic and subject-matter competencies. Construct-irrelevant factors, which include cognitive, stylistic, affective and conative aspects of task performance are as important for test score interpretation as construct-relevant factors because they also contribute toward the difficulty of tasks (Scheuneman, 1991; Snow, 1993). Specific construct-irrelevant factors identified in the literature include neatness of handwriting, verbal skills, information-processing strategies, guessing and item omission, together with students' attitudes, motivation, anxiety, expectations for success and feelings of self-efficacy toward the various test item-response formats (Bolger, 1984; Linn, De Benedictus, Delucchi, Harris & Stage, 1987; Mazzeo et al., 1991; R. Murphy, 1982; Scheuneman, 1991; Snow, 1993).

One popular explanation (for example, Breland et al., 1991; Hoste, 1982; R. Murphy, 1982; Volkoff & Hocevar, 1995) attributes girls' relatively greater success on questions requiring written answers to reports of their verbal superiority (Maccoby & Jacklin, 1974; Sincoff & Sternberg, 1987, 1988). R. Murphy's (1982) data appear to support this hypothesis. The only subject where he did not detect any interaction between girls' and boys' achievements and item-response formats was mathematics, which does not require the construction of written answers. Additionally, evidence from the 1986 Advanced Placement history examinations shows that English skills and verbal aptitude are significant predictors of constructed-response scores on document based questions (Breland et al., 1991). Moreover, various researchers (for example, Breland & Griswold, 1982; Breland et al., 1991; Volkoff & Hocevar, 1995) have interpreted their data to be consistent with reports of girls' relative superiority in verbal production and of boys' superiority in verbal comprehension (Sincoff & Sternberg, 1987, 1988). More recently, D. Burkam and A. Burkam (1995) have shown that girls' relatively higher scores on the constructed-response sections of science examinations for Grade 4 and 8 students are ameliorated when the girls' greater exposure to reading and writing are taken into account. However, other researchers have reported gender-related differences in achievement on both the objective and non-objective sections of mathematics examinations (for example, Anderson, 1989; Bolger & Kellaghan 1990), which suggests the operation of factors other than verbal skills. Nevertheless, the hypothesis that boys' and girls' differing patterns of achievement on multiple-choice and constructed-response test formats are attributable to gender differences in verbal abilities is the subject of ongoing research (Breland et al., 1991; Volkoff & Hocevar, 1995). It is important to note in this context that verbal skills have traditionally been regarded as irrelevant constructs insofar as science achievement is concerned.

There is some evidence that the neatness and presentation of girls' work is superior to boys' work (Massey, 1983; Spear, 1989) and that neatness of handwriting influences girls' and boys' achievement scores (Briggs, 1980). Consequently, girls' relatively higher achievements on written examinations have been attributed to the greater neatness and presentation of their work (Bolger & Kellaghan, 1990; Breland et al., 1991). However, Massey (1983) reports no support for the view that untidy

candidates are penalised by examiners in A-level English Literature examinations. Furthermore, analyses of the 1986 Advanced Placement examinations show that neatness of presentation and quality of handwriting did not predict constructed-response scores on United States history and European history papers (Breland et al., 1991). Methodological issues could also be important in terms of the detection of such differences. Briggs (1980), for example, used subject teachers to score students' scripts, unlike Massey (1983) and Breland et al. (1991) who used trained examiners. On the other hand, Spear (1984) provides an interesting counterpoint to the view that girls are advantaged by their handwriting on constructed-response items. Instead, she found that science tasks that were attributed to girls from the nature of the handwriting were graded lower than identical work that was attributed to boys. Overall, it seems that gender-related differences in neatness and the presentation of work are unlikely to help explain boys' and girls' patterns of achievement on multiple-choice and constructed-response test formats.

Students fail to achieve on multiple-choice questions, either because their answers are incorrect or because they fail to answer questions. It has been well documented that girls tend to omit more multiple-choice questions than boys (Anderson, 1989; Ben-Shakhar & Sinai, 1991; Hanna, 1986) even when no penalties are applied for guessing multiple-choice answers (Ben-Shakhar & Sinai, 1991). Hanna (1986), for example, reported an omission ratio of 3:2 for Grade 8 girls compared to Grade 8 boys. It has been presumed that boys are advantaged on multiple-choice questions because girls are less willing than boys to guess answers (Bolger & Kellaghan, 1990). Although the correction of raw scores for guessing reduces the relative advantage enjoyed by boys (Ben-Shakhar & Sinai, 1991), these gender-related differences in guessing tendencies appear to account for only a small portion of the overall advantage to boys on multiple-choice items (Ben-Shakhar & Sinai, 1991; Chopin, 1975). Moreover, female undergraduate students have been reported to achieve relatively worse than their male counterparts on multiple-choice test papers even when they responded to a similar proportion of the items as the male students (Anderson, 1989). Consequently, gender-related differences in item omission and guessing tendencies are unlikely to explain girls' and boys' patterns of achievement according to test item-response formats.

Others have postulated hypotheses relating to boys' and girls' cognitive styles. It has been suggested that differences between girls and boys for set breaking and convergent thinking may help explain their patterns of achievement on multiple-choice and constructed-response items (R. Murphy, 1982; Snow, 1993). P. Murphy (1988, 1991) and Harding (1994), for example, argue that boys and girls learn to respond to the world in different ways, and that multiple-choice items favour a male's way of thinking. These commentators draw on object-relations theory (Harding & Sutoris, 1987) to assert that females tend to reflect on the broader context of questions, and that they are confused by the ambiguities contained in multiple-choice distractors. Furthermore, they assert that males are more able to disembed relevant information from questions. Little research has tested hypotheses relating to differences in girls' and boys' cognitive styles and their achievements on the various test item-response formats. However, Walding, Fogliani, Over and Bain (1994) have reported a greater tendency for Year 11 and 12 girls to select true but irrelevant multiple-choice distractors, and to select distractors that satisfy some but not all the requirements of multiple-choice items.

The remaining hypotheses relate to a series of complex affective factors such as confidence, anxiety and motivation. It appears that these factors are related to boys' and girls' socialisation, preferred response styles and their attitudes toward test formats. P. Murphy (1994), for example, draws on findings from the Assessment and Performance Unit Language Surveys to report that girls and boys differ in their out-of-school reading habits and that these reading preferences result in gendered styles of expression. In this context, the Assessment and Performance Unit Language Surveys (Gorman, White, Brooks, Maclure & Kispal, 1988) report that boys tend to produce short, factual and episodic responses whereas girls prefer to generate extended pieces of reflective writing. Additionally, P. Murphy (1994) alleges that these gendered styles of expression are reinforced by differential out-of-class and in-class experiences through the expectations exerted by parents, peers and teachers (P. Murphy, 1994, 1996). Hence, she (P. Murphy, 1994) writes:

[S]tudents' cognitive responses are mediated by affective characteristics i.e. students' expectations of themselves, others and the subject being assessed, their self image, values and confidence in themselves and their abilities ... The impossibility of separating the cognitive from the affective in students' responses has to be considered if interpretations of assessment outcomes are to be valid and therefore just. (p.7)

Research on students' preferred test formats has produced findings that are relevant in this context. Zeidner (1987) reported that junior high school students from Israel prefer multiple-choice tests over essay tests, and that they experience less anxiety and have a greater expectancy for success on multiple-choice test formats. Additionally, Gellman and Berkowitz (1993) report that female university students expressed strong preferences for essay tests whereas their male counterparts indicated a slight preference for multiple-choice response formats. It is important to note in this context that confidence affects students' ability to engage with assessment tasks as well as the nature of their engagement (P. Murphy, 1991). Moreover, confident students are more likely to experiment with strategies during tests that they have not rehearsed in class (Linn & Hyde, 1989). Little research appears to have investigated the relationship between these factors and boys' and girls' achievements on multiple-choice and constructed-response items. Hence, the presence of gender-related differences in these affective characteristics may help to better understand girls' and boys' patterns of achievements on multiple-choice and constructed-response test formats.

The Differential Reliabilities Hypothesis

The final hypothesis asserts that boys' and girls' patterns of achievements on multiple-choice and constructed-response items are due to their disparate reliabilities (Mazzeo et al., 1991). Constructed-response questions have much lower reliabilities than selected-response formats (Mehrens & Lehmann, 1984; Sax 1989) because the reasons for giving and withholding credit when scoring answers are less obvious (Ebel, 1979), and also because fewer constructed-response questions can be answered in the same time as multiple-choice questions (Mazzeo et al., 1991). As a case in point, Blum and Azencot (1986) report an interscorer correlation of 0.99 for a series of multiple-choice questions and just 0.36 for equivalent essay-type questions.

The differential reliabilities hypothesis asserts that differences in academic and subject matter specific competencies between groups on constructed-response formats are attenuated by measurement error (Mazzeo et al., 1991). However, analysis of data from a range of Advanced Placement examinations casts doubt on the

differential reliabilities hypothesis as accounting for much of the discrepancy between girls' and boys' relative achievements on multiple-choice and essay-type questions (Mazzeo et al., 1991).

FACTORS INFLUENCING GIRLS' AND BOYS' FORMAT-RELATED SCIENCE ACHIEVEMENT

Introduction

The preceding section surveyed boys' and girls' patterns of achievement according to test item-response formats. The evidence was quite clear that boys achieve relatively higher than do girls on multiple-choice test formats in comparison to constructed-response test formats. This section examines gender-related differences in the construct-irrelevant factors; namely, verbal skills, cognitive styles and affective characteristics, which were identified to hold the greatest promise for explaining girls' and boys' differential patterns of achievement according to test item-response test formats. Additionally, boys' and girls' spatial and quantitative skills are compared, as a matter of completeness and also as part of a larger argument that attributes gender-related differences in achievement to non-cognitive factors.

The Gender Difference Literature

Research in the area of gender differences has been, and remains, of considerable interest to educators. A voluminous amount of literature has been generated in this area over the last thirty years or so. Among the early works were the classic reviews by Anastasi (1958) and Maccoby and Jacklin (1974). Scholars still credit Maccoby and Jacklin's 1974 treatise, *The Psychology of Sex Differences*, as the best starting point for data-based discussions of gender differences (Wilder & K. Powell, 1989; Willingham & Cole, 1997c).

Maccoby and Jacklin (1974) accumulated and reviewed more reports than anyone else before them on gender differences. They reported their findings by a vote-counting method. Differences between the sexes were deemed to exist when a large number of studies reported gender differences in one direction. However, their

review has been criticised on methodological grounds. Oft-cited criticisms relate to the unrepresentativeness of their samples, the large disparities between the studies they integrated and the generalisations they made from studies of young children (Gipps & P. Murphy, 1994; Hyde, 1986; Wilder & K. Powell, 1989). Furthermore, critics maintain that narrative reviews, such as Maccoby and Jacklin's review, are unsystematic, subjective, and prone to error because they divert attention away from the extent of the differences between groups (Halpern, 1992; Hyde, 1986; Hyde & Linn, 1988; Willingham & Cole, 1997c). More recently, quantitative reviews involving the use of meta-analytical techniques have become increasingly popular for integrating research studies on gender differences.

Meta-analysis is a research method that integrates studies on a single topic to calculate an effect size. Recent meta-analyses (for example, Linn & Petersen, 1985a; Hyde & Linn, 1988; Hyde, Fennema & Lamon, 1990; Hyde, Fennema, Ryan, Frost & Hopp, 1990) partition research studies until homogeneity, or near homogeneity is achieved. This procedure allows the calculation of effect sizes for studies that are essentially replications of each other. Although meta-analyses are able to draw conclusions from seemingly inconsistent and contradictory data (Hyde, 1986), meta-analytical techniques have not been without their critics (for example, H. Eysenck, 1978). It has been argued, for example, that findings from meta-analyses lack rigour because they are contingent on the use of small samples, publication biases, and poorer quality research in the case of unpublished studies (Burnett, 1986; Halpern, 1992). Consequently, this review examines reports from large-scale testing programs, where these data are available, as well as findings from meta-analyses.

The reporting and interpretation of statistical data yielded by meta-analytical procedures has been much debated (Cohen, 1969; Eagly, 1987; Rosenthal & Rubin, 1982b). Among the statistics reported by meta-analyses are the effect size, d , and the proportion of variance, ω^2 . The effect size, d , is a measure of how far group means are apart in standard deviation units whereas ω^2 measures the variance in the entire population that can be attributed to differences between the groups. Although it is difficult to determine practical significance from the proportion of variance (Burnett, 1986), Cohen (1969) suggests a series of benchmarks for small ($d = 0.20$), moderate ($d = 0.50$) and large ($d = 0.80$) effect sizes. As a guide, Keeves (1992) reports that an

effect size of 0.30 is approximately equivalent to the amount of science learning that takes place during a year at the lower secondary school level, and Cohen (1969) reports that an effect size of 0.80 corresponds to the difference in height between 13-year old and 18-year old girls. Moreover, Rosenthal and Rubin (1982a) have demonstrated that a small effect size of 0.20, with a between the groups variance of 4%, corresponds to a situation where 60% of one group and 40% of the other group scores above the mean. Although Rosenthal and Rubin's Binomial Effect Size Display (BESD) provides a statistic that assists in the clear communication of research findings (Eagly, 1987), the determination of practical significance involves context-dependent value judgements, which are not easily defined in statistical terms (Halpern, 1992; Kirk, 1996). Nevertheless, Cohen's (1969) benchmarks represent a 'good beginning' (Kirk, 1996, p.756) for the interpretation of practical significance. This review quotes effect sizes, and the BESD statistic where possible, to ascertain the practical importance of gender differences. Effect sizes are reported according to the convention that positive values reflect higher male scores and that negative values reflect higher female scores.

Cognitive Skills

The Nature of Cognitive Skills

Intelligence, that is, the ability to think, learn and remember (Halpern, 1992) is not a single homogeneous construct. Instead, three clusters of factors; namely, verbal, number (quantitative) and perception (visual-spatial) have been identified. Although other models for intelligence have been described (Halpern, 1992), this review follows the traditional classification by which research has been conducted and reviews in this field integrated.

Verbal, quantitative and visual-spatial skills are theoretical constructs, which represent unobservable variables that help explain results on certain types of tests (Halpern, 1992). Each of these constructs is comprised of a collection of specific subskills. For example, quantitative skills include computation, conceptual understanding and problem solving (Hyde, Fennema & Lamon, 1990). However,

measurement of gender differences in cognitive skills is problematic because cognitive skills are often confounded with each other on tasks (Gipps & P. Murphy, 1994) and also because males and females may differ in the processes they select for solving tasks (Linn & Petersen, 1985a). It is also important to note that scores on tests of cognitive skills do not represent innate ability, and that these scores are influenced by context (Linn, 1991) and prior experience (Fennema & J. Sherman, 1977; Linn, 1991).

A Comparison of Girls' and Boys' Cognitive Skills

Verbal Skills

Verbal skills are central to virtually all intellectual pursuits (Hyde & Linn, 1988). They relate to all aspects of language usage, and they include specific subskills of word fluency, comprehension of written material, spelling, punctuation and the use of analogies (Halpern, 1992; Hyde & Linn, 1988; Maccoby & Jacklin, 1974). Verbal skills are important to the study of high school science, and it has been hypothesised that gender differences in these skills are responsible for the interaction between gender and test item-response formats (Breland et al., 1991; Hoste, 1982; R. Murphy, 1982; Volkoff & Hocevar, 1995). As noted earlier in this chapter, gender-related differences in verbal skills is the subject of ongoing research in the field of gender and assessment.

Early reviewers unanimously reported that females have superior verbal skills. Anastasi (1958) reported that these differences were well-established and consistent through out the life span. However, Maccoby and Jacklin (1974) reported that the sexes perform similarly on verbal tasks until the ages of 10 or 11, and that girls begin to develop their verbal superiority from these ages. Additionally, Maccoby and Jacklin reported a female advantage of about 0.25 standard deviation units across all types of verbal skills.

Hyde's (1981) meta-analysis of Maccoby and Jacklin's data yielded a median effect size of -0.24, which corroborates Maccoby and Jacklin's estimates. Hyde calculated that gender differences contribute no more than 1% of the population variance, and consequently, she argued that gender differences in verbal skills are

small. However, the BESD statistic (Rosenthal & Rubin, 1982b) indicates that 57.5% of the girls and 42.5% of the boys in her study achieved scores above the mean (Rosenthal & Rubin, 1982a). Hence, gender does not appear to be the poor predictor of verbal performance as claimed by Hyde (1981).

More recently, Hyde and Linn (1988) employed the techniques of meta-analysis and process analysis to synthesise research investigations involving approximately 1.4 million students. The direction of the gender differences within these studies was not unanimous; 24% of the effect sizes showed a male advantage and 75% of the effect sizes showed a female advantage. The weighted mean effect size ($d = -0.11$) revealed a very slight female advantage with 52.5% of the females and 47.5% of the males achieving scores above the mean. Partitioning the effect sizes yielded five cognitive processes; namely, retrieval of word definitions, retrieval of the names of pictures, analyses of the relationships among words, selection of information from written text and verbal production (speaking and essay writing). Effect sizes for these cognitive skills ranged between +0.12 and -0.09, and differences of -0.19 were reported for tasks that required a mixture of cognitive processes. Importantly, for the context of this study, gender differences on essay writing tasks ($d = +0.09$) were trivial. Partitioning the studies according to age yielded trivial gender differences for pupils aged between 6 and 11 years ($d = -0.06$), and for pupils aged between 11 and 18 years ($d = -0.11$). Moreover, Hyde and Linn's results show that the extent of gender differences in verbal skills has diminished with time. Whereas Hyde and Linn calculated effect sizes of -0.23 favouring females for studies conducted before 1974 (the time of Maccoby and Jacklin's review), they reported effect sizes of -0.10 for studies conducted after 1974. None of these gender-related differences for verbal skills that are reported by Hyde and Linn is likely to be practically important. This finding challenges the popular assumption that females have superior verbal skills.

Halpern (1992), however, argues that Hyde and Linn's (1988) effect sizes are artifacts of their sampling, and that larger gender differences would be found if their sampling was more representative. For example, Hyde and Linn's (1988) average effect size for essay writing was influenced by scores from the College Board's English Composition Test, a single large-scale study involving a highly selected group of students. Nevertheless, analyses of well-sampled, nationally administered tests

from the United States lend support to Hyde and Linn's conclusions, that there are no psychologically and educationally important differences between boys' and girls' verbal skills. The 1980 Differential Aptitude Tests (DAT) show a female advantage for the basic English skills of spelling ($d = -0.50$) and language ($d = -0.43$) and essentially no differences ($d = +0.02$) between males and females on the higher order skills of verbal reasoning (Feingold, 1988). Similarly, the 1983 Preliminary Scholastic Aptitude Test-Verbal tests (PSAT-Verbal) show essentially no gender differences ($d = +0.02$) among junior high school students (Feingold, 1988). Feingold also reports a pattern of declining female advantage on the DAT and PSAT-Verbal tests over the years of their administration. Hence, the closure of the gap between girls' and boys' verbal achievements in recent years to the extent where they are not practically important suggests that boys' and girls' differential patterns of achievement on multiple-choice and constructed-response items cannot be attributed to gender-related differences in verbal skills.

Spatial Skills

Spatial skills are important components of cognition (Linn & Petersen, 1985a; Linn, 1991) and they are associated with science achievement (Linn & Petersen, 1986). Moreover, it has been hypothesised that gender differences in spatial skills contribute toward gender differences in science and mathematics achievement (for example, Benbow & J. Stanley, 1980; Fennema & J. Sherman, 1977). Spatial skills, despite their theoretical importance, are hard to define (Caplan, McPherson & Tobin, 1985; Maccoby & Jacklin, 1974) and their nature requires clarification (Linn & Petersen, 1985a). However, spatial skills generally refer the skills that are required for the generation, representation, transformation and recall of non-linguistic information (Linn & Petersen, 1985a). Linn and Petersen (1985a) identify three broad categories of spatial skills; namely spatial perception, mental rotations and spatial visualisation, to which Halpern (1992) adds the recently identified spatio-temporal skills. Spatial visualisation is the most relevant of these spatial skills (Linn, 1991) in educational contexts because it involves the analytic processing of spatially-presented information (Linn & Petersen, 1985a). As such, spatial visualisation is logically related to the content of science and mathematics courses (Fennema & J. Sherman,

1977; Linn & Petersen, 1986).

Gender differences in spatial skills have been widely reported. Maccoby and Jacklin (1974) concluded that no gender differences are found until adolescence, and that boys develop their superior spatial skills after this age. Meta-analysis of Maccoby and Jacklin's studies (Hyde, 1981) yielded a median effect size of +0.45 for visual-spatial skills ($\omega^2 = 0.043$), and Hyde concluded from the between-groups variances that gender differences in spatial skills are small. However, her effect sizes indicate the presence of practically important differences, which correspond to 62% of the boys and only 32% of the girls achieving scores above the median.

More recently, Linn and Petersen (1985a) reviewed a larger group of studies and they detected the presence of very small mean effect sizes for spatial visualisation ($d = +0.13$). Linn and Petersen's (1985a) results are supported by data from the 1980 Differential Aptitude Tests, which show an average effect size of +0.15 for high school students on the space relations test (Feingold, 1988). Although, Linn and Petersen (1985a) report larger gender differences favouring boys on spatial perception and mental rotations, it must be remembered that these spatial skills are unrelated to the content of science courses.

Hence, results from recent meta-analyses and standardised tests show the presence of inconsequential gender differences for educationally important spatial skills. Therefore, gender differences in spatial skills are unlikely to explain girls' and boys' differential patterns of achievement on multiple-choice and constructed-response items.

Quantitative Skills

Quantitative skills are important to the study of school science, particularly in the upper high school years, because science is reliant on mathematics to formulate its problems and to find its solutions (Maccoby & Jacklin, 1974). Quantitative skills have been categorised according to their cognitive complexity into subskills involving computation, conceptual understanding and problem solving (Hyde, Fennema & Lamon, 1990).

Early reviewers (for example, Maccoby & Jacklin, 1974) reported that boys begin to develop superior quantitative skills in adolescence and that these differences

are robust. Although Hyde's (1981) meta-analysis yielded medium sized effects ($d = +0.43$) she concluded from the between-groups variance ($\omega^2 = .01$) that gender differences in quantitative skills are small. However, the BESD statistic (Rosenthal & Rubin, 1982a) indicates that approximately 58% of the boys and 42% of the girls in her study achieved scores above the mean. Recently, Hyde, Fennema and Lamon (1990) used more sophisticated meta-analytical procedures to integrate a larger and updated group of studies relating to boys' and girls' quantitative skills.

Approximately half of their studies reported effect sizes favouring the boys and the other half of their studies reported effect sizes favouring the girls. Overall, the average effect size for gender differences in quantitative skills was -0.05 for samples drawn from the general population, which corresponds to 48.5% of the boys and 51.5% of the girls scoring above the mean. Additional analyses indicated the absence of any practically important gender differences for computation and conceptual understanding, and small differences ($d = +0.29$) favouring high school boys for problem solving. However, gender-related differences in problem solving skills may be artifacts of differential course taking patterns in mathematics and science — there is some evidence that gender differences in problem solving skills are reduced when mathematics experience is taken into account (Fennema & J. Sherman, 1978; Linn, 1991). Moreover, Hyde, Fennema and Lamon (1990) reported that the overall male advantage has decreased in mathematics from $+0.31$ for studies published before Maccoby and Jacklin's (1974) review to $+0.14$ for later studies. It is important to note in this context that the decreasing gap between girls' and boys' mathematics achievements in recent years is paralleled by girls' increasing participation in mathematics courses (Linn, 1991), and that gender differences in quantitative skills can be regarded as inconsequential.

Evidence from large-scale testing programs supports the findings reported by Hyde, Fennema and Lamon (1990). Feingold (1988), for example, reports the absence of any appreciable gender differences ($d = -0.10$) among high school students on the numerical scale of the Differential Aptitude Tests and on the 1983 Preliminary Scholastic Aptitude Test in Mathematics ($d = +.12$). Although Feingold reported a small to moderately sized male advantage ($d = +0.42$) on the 1983 Scholastic Aptitude Tests in mathematics, these data are consistent with other scores for highly

selected populations (for example, Benbow & J. Stanley, 1980).

Hence, findings from recent meta-analyses and large-scale testing programs indicate the absence of any practically important gender differences in quantitative skills for the general population of school students. Therefore, it is unlikely that boys' and girls' differential patterns of achievement on multiple-choice and constructed-response items can be attributed to gender-related differences in quantitative skills.

Implications of Research Findings About Girls' and Boys' Cognitive Skills

The research literature indicates that there are no practically important gender differences in any of the cognitive skills that underscore science achievement. Accordingly, gender differences in cognitive skills are unlikely to yield insights relevant to boys' and girls' differential patterns of achievement on multiple-choice and constructed-response items.

Cognitive Styles

The Nature of Cognitive Styles

The absence of gender-related differences in cognitive skills insofar as science achievement is concerned does not exclude the possibility of differences between girls' and boys' cognitive styles. Cognitive styles differ from cognitive skills in that they are concerned with the form rather than the content of cognitive activity (Messick, 1984; Witkin, C. Moore, Goodenough & Cox, 1977). As such, cognitive styles relate to individual preferences for perceiving, organising and thinking about the world (Messick, 1984; Witkin et al., 1977).

Cognitive styles such as field articulation (that is, field-dependence and field-independence), impulsivity and reflectivity, together with converging and diverging styles underscore particular ways of processing information (Messick, 1984; Saracho, 1989; Witkin et al., 1977). Witkin, Oltman, Raskin and Karp (1971, cited by Lu & Suen, 1995) identify the importance of the field articulation construct for achievement tests. Field-independent individuals are supposed to be more able to isolate the salient

features of problems, ignore task-irrelevant contexts and restructure problems, whereas field-dependent individuals are supposed to be more able to view questions globally rather than analytically (Witkin et al., 1977). The hypothesis that field-independent students are advantaged by multiple-choice response formats because they have greater analytical skills (R. Murphy, 1982; P. Murphy, 1988, 1991) has some support in the literature (for example, Burkhalter & Schaer, 1984-85; Crow & Piper, 1986; Okebukola, 1986) although this advantage may not exist when multiple-choice questions are written according to psychometric guidelines (Armstrong, 1993). Moreover, there is also some evidence to the contrary, that field-independent students perform relatively better than field-dependent students on constructed-response questions rather than multiple-choice questions (Lu & Suen, 1995).

A Comparison of Boys' and Girls' Cognitive Styles

Most of the field articulation research has employed the *Embedded-Figures-Test* and the *Rod-and-Frame Test*. These tasks are regarded as measures of global analytic ability (Witkin et al., 1977) because they test the ability to ignore distracting cues and disembed elements from their background (Gipps & P. Murphy, 1994; Halpern, 1992). In this context, highly differentiated individuals are able to separate themselves from their environment and they are able to separate items in their environment from each other (Halpern, 1992). These individuals, unlike their field-dependent peers, make greater use of mediating processes such as analysing and restructuring, and they are less distracted by the most obvious cues (J. Davis, 1991).

Field articulation studies report that children are more field-dependent than adults and that females are more field-dependent than males (Demick, 1991; Witkin et al., 1977). However, these differences, which are reported to emerge in adolescence (Maccoby & Jacklin, 1974; Witkin et al., 1977) are not always observed. Burstein, Bank and Jarvik (1980) report from 37 studies that were reviewed elsewhere, that 11 studies showed males to be more field-independent, 23 showed no gender differences, 2 showed mixed findings and that 1 study showed females are more field-independent than males. Additionally, Hyde (1981) reported practically important gender differences (median $d = +0.51$, $\omega^2 = .025$) from her meta-analysis that correspond to

the situation where 63% of the boys and 37% of the girls achieve scores on field articulation tests above the mean (Rosenthal & Rubin, 1982a).

Although the literature reports practically important, gender-related differences in field articulation, there is some question about the extent to which measures of field-independence and field-dependence require the use of cognitive skills (J. Davis, 1991; Messick, 1984; Wapner & Demick, 1991). J. Sherman (1967) asserts that the *Rod-and-Frame* and the *Embedded-Figures-Tests* have strong spatial components and that gender differences in field articulation are artifacts of gender-related differences in spatial skills. This view is supported by developmental data, which show that the development of gender-related differences in field articulation is paralleled by the development of gender-related differences in spatial skills (Maccoby & Jacklin, 1974). Additional support for this view is provided by Hyde, Geringer and Yen (1975) who report that gender differences in field articulation disappear when spatial differences are controlled. It is important to note that these spatial skills differ from the spatial skills that underscore science achievement (Fennema & J. Sherman, 1977; Linn & Petersen, 1986). Hence, there is no evidence that either sex is superior at analytical tasks and restructuring problems, the cognitive styles that are presumed to underlie superior achievement on multiple-choice items.

Implications of Research in Cognitive Styles

The gender-related differences in field articulation, the most important of the cognitive styles insofar as achievement on multiple-choice items is concerned, appear to be artifacts of spatial skills that are unrelated to science achievement. Accordingly, there is little evidence that boys and girls differ according to their analytical skills. Hence, further examination of these cognitive styles is unlikely to contribute toward a better understanding of girls' and boys' patterns of achievement according to test item-response formats.

Boys' and Girls' Affective Characteristics

The Nature of Girls' and Boys' Affective Characteristics

The affective domain includes a number of closely related constructs, such as attitudes, beliefs, values, interests and motivation (Simpson, Koballa, Oliver & Crawley, 1994). Attitudes are of considerable interest to science educators because attitudes are associated with students' behaviour (Shrigley, Koballa & Simpson, 1988) and there is some thought that attitudes influence academic achievement (Bloom, 1976; Schibeci, 1984). Simpson et al. (1994) assert, in this vein, that

Those who teach science have come to a universal observation: Students' behavior is influenced by the values they hold, the motivation they possess, the beliefs they bring from home to the classroom, and the myriad attitudes they have formulated about school, science, and life in general. The key to success in education often depends on how a student feels toward home, self, and school. (p.211).

However, the conceptualisation of attitudes and their relationship with behaviour is problematic (Koballa, Crawley & Shrigley, 1988; Simpson et al., 1994) for reasons that are discussed later in this section. This review follows the approach of Ajzen (1988) and it conceptualises attitudes as predispositions that vary in their direction and strength toward some object, person or idea. As such, attitudes are derived from students' perceptions and their beliefs, and they underscore affective responses and behavioural intentions (Ajzen, 1988; Fishbein & Ajzen, 1975; Shrigley et al., 1988).

The literature is unanimous in its reports that boys and girls undergo different socialisation experiences, and that they learn behaviours and dispositions according to socially- and culturally-determined expectations (Halpern, 1992; P. Murphy, 1988). Psycho-social gender-differences, that is learned differences between girls and boys in their non-cognitive personality and social behaviours, have been documented for anxiety, confidence, aggression, empathy, conformity, self-esteem and interest in science (Linn & Hyde, 1989; Linn & Petersen, 1985b). Typically, boys have been characterised as aggressive, ebullient and confident, whereas girls have been characterised as submissive, intuitive, timid, diffident and nurturing (Easlea, 1986; Kelly, 1985; P. Murphy, 1988). These gender differences are situational (Linn & Hyde, 1989) and they are underpinned and maintained by differences in the learned

psychological constructs by which boys and girls orient themselves to their external world (Halpern, 1992; Leder, 1985). Research, insofar as science achievement is concerned, has assumed that girls and boys develop affective characteristics that differentially impinge on their achievement (Kahle & Meece, 1994), and that it is impossible to separate the influence of students' affective responses from their test answers (Gipps & P. Murphy, 1994). It is also important to note in this context that school science is perceived by boys and girls as a male domain (Easley, 1986; Kelly, 1985; Sjoberg & Imsen, 1988).

Attitudes Toward Science and Science Achievement

Much of the literature describing attitudes toward science lacks clarity because attitude measures typically include such diverse affects as enjoyment, interest, anxiety, motivation, self-concept, perceived usefulness of science and science careers.

Consequently, it is sometimes unclear how researchers have operationalised attitudes (Rennie & Punch, 1991) and how some reviewers (for example, Fleming & M. Malone, 1983; Steinkamp & Maehr, 1983) have integrated research on students' attitudes toward science. Moreover, the direction of the causal pathway is not straightforward. Whereas Keeves and Kotte (1992) report that the causal path from attitudes to achievement is preferable empirically and logically, Rennie and Punch (1991) report that affect is more strongly related to previous achievement. It appears, however, that the relationship between attitudes toward science and science achievement is bi-directional (Steinkamp & Maehr, 1983; Willson, 1983).

Much of the science attitude research has assumed that attitudes are highly related to achievement (Fraser, 1982; Simpson et al., 1994). However, the literature generally reports 'hardly impressive' (Willson, 1983, p.848), or at the best modest (Simpson & Oliver, 1990) correlations between science attitudes and science achievement. For example, correlations in the order of +0.16 have been reported by Willson (1983), and Steinkamp and Maehr (1983). However, relatively higher correlations are reported when unidimensional measures of science attitude are correlated with achievement (Willson, 1983). As a case in point, Kremer and Walberg (1981) report a correlation between motivation and science achievement of

+0.37. Recently, Weinburgh (1995) reported from her meta-analysis mean correlations between attitudes and science achievement of +0.50 for boys and +0.55 for girls. However, the interpretation of her findings is difficult because it is unclear how attitudes were conceptualised in the studies that she combined. Of greatest interest to this review is the approach that divides science-related attitudes into self-, school- and home-related variables (for example, Simpson & Oliver, 1990; Simpson & Troost, 1982). Simpson and Oliver (1990) report that self-related science affects are the strongest predictors of science achievement, and that family and school influences are heavily mediated through them. Hence an understanding of gender differences in self-related affects, and how they operate for the various test item-response formats, may help explain girls' and boys' differing patterns of achievement on test item-response formats.

A Comparison of Boys' and Girls' Self-Related Affective Characteristics

Self-Concept

Self-concept is a store of self-perceptions (Shavelson, Hubner & Stanton, 1976) that provide answers to questions such as; Who am I?; What do I believe?; and What do I value? (Sinclair, 1991). These self-perceptions are derived from students' interactions with significant others, their self-attributions and their social experiences (Byrne & Shavelson, 1987). On the other hand, self-esteem involves personal judgements of self-worth (Coopersmith, 1967) and it is associated with the extent that students feel that they can live up to other people's expectations (Sinclair, 1991). However, distinctions between self-concept and self-esteem are often blurred in practice (Poiner, 1995).

The literature (for example, Maccoby & Jacklin, 1974; Marsh, J. Parker & Barnes, 1985) reports no convincing evidence for 'presumptions of universally low self esteem among [school-age] girls' (Poiner, 1995, p.30). Instead, gender-related differences in self-concept are associated with gender-related stereotypes (L. Jackson, Hodge & Ingram, 1994; Renshaw, 1990). For example, the lower self-concepts of girls in science and mathematics and their higher self-concepts for English have been well documented (Byrne & Shavelson, 1987; Eccles Parsons, Adler & Kaczala, 1982;

Marsh et al., 1985).

Self-Confidence

Self-confidence describes expectancies for achieving particular outcomes and self-appraisals of one's capabilities for executing particular courses of action (Bandura, 1977, 1986, 1997; L.G. Jones & L.P. Jones, 1989; Lenney, 1977; Licht, Stader & Swenson, 1989; M. Lundeberg et al., 1994; Maccoby & Jacklin, 1974). The distinction between these two aspects of self-confidence is necessary because academic achievement may be influenced by the likelihood that behaviours will lead to desired outcomes as well as by students' confidences in their capabilities (Schunk, 1989, 1990). The attributions, that is the reasons that students give for their successes and failures, are also important in this context.

Attribution theory has empirical support in its explanations for achievement (Wilder & K. Powell, 1989). Different types of causal attributions lead to differing achievement expectancies (Licht & Dweck, 1983) and self-fulfilling prophecies for success or failure (Beyer, 1990; Eccles Parsons, Adler & Meece, 1984; Whitley, McHugh & Frieze, 1986). Whitley and his associates (1986) report from their meta-analysis that men and women make similar types of attributions and that there are small differences in the extent of mens' and womens' attributions. Although other reviewers (for example, Licht, 1987; Licht & Dweck, 1983; Licht et al., 1989) report that females and males have different attributional patterns, it appears that these differences are unlikely to completely explain gender-related differences in academic achievement (Leder, 1985; Whitley et al., 1986; Wilder & K. Powell, 1989).

Overall, one pattern is quite clear from the literature. Boys, tend to overestimate their likely achievements in science and a wide range of curricular areas relative to girls (Beyer, 1990; Fennema & J. Sherman, 1977; Hyde, Fennema, Ryan et al., 1990; L.G. Jones & L.P. Jones, 1989; Kelly, 1988; Lenney, 1977; Linn & Hyde, 1989; Licht & Dweck, 1983; M. Lundeberg et al., 1994; Vollmer, 1986), even when their prior experience is controlled (Furst, Tenebaum & Weingarten, 1985; Maccoby & Jacklin, 1974; Mednick & Thomas, 1993). There is also some evidence that gender-related differences in confidence are moderated by situational factors and that they are related to the sex-typing of domains (Beyer, 1990; Hyde, Fennema, Ryan

et al., 1990, L.G. Jones & L.P. Jones, 1989; Lenney, 1977; M. Lundeberg et al., 1994). For example, girls display less confidence than boys in their science abilities, and their lower confidences persist even when they achieve as well as boys (Fennema & J. Sherman, 1977; Kelly, 1988; Linn & Hyde, 1989; Licht & Dweck, 1983). Hence, gender-related differences in students' confidences may help explain girls' and boys' patterns of achievement on multiple-choice and constructed-response test formats.

Anxiety

Anxiety is a key psychological construct that is related to theories of learning and motivation (Bandura, 1991, 1993; Carver & M.F. Scheier, 1991; Covington, 1992; Tobias, 1985). Moreover, stress is related to anxiety and is antecedent to it. Whereas stress is associated with the appraisal of personally important situations that threaten either loss or one's sense of well-being, anxiety is evoked when the requisite coping strategies are lacking (Bandura, 1991; Becker, 1982; I. Sarason, 1991; Schwarzer, van der Ploeg & Spielberger, 1982). As such, anxiety is a complex and unpleasant reaction to stressful situations, irrespective of whether these situations are real or imagined (Becker, 1982; I. Sarason, 1984; Sieber, 1980). It is also important to note in this context that various situation-specific forms of anxiety have been identified. For example, distinct types of anxious responses are associated with solving science and mathematics problems, working with dangerous substances, performing in front of others and taking tests (Czerniak & Chiarelott, 1984; Wynstra & Cummings, 1993). Others (for example, Covington, 1992; Tobias, 1985) have described the operation of different types of anxiety at pivotal points in the educative process.

There is ample evidence from laboratory studies and class tests to show the debilitating influence of anxiety in evaluative settings. Highly test-anxious students do not achieve as well as their peers on tests, and they have lower levels of self-confidence and depressed expectancies for success (Czerniak & Chiarelott, 1984; Westerback & Primavera, 1992). Seipp (1991), reports from her recent meta-analysis an effect size of 0.43 for differences in academic achievement between high and low test-anxious students. This effect size is practically significant and it corresponds to a

situation where 39% of the highly test-anxious students and 61% of the students from low anxiety group report scores above the mean (Seipp, 1991). Additionally, Hembree's (1988) meta-analysis of the test anxiety research literature shows the existence of practically important gender differences that disadvantage junior high school girls ($d = -0.43$) as well as senior high school and college girls ($d = -0.27$).

There is also some evidence that anxiety is related to examination formats. Shaha (1982) reports that matching test formats elicited less test anxiety than multiple-choice response formats among high school students, Weare (1984) that adult students expressed a preference for multiple-choice questions over essay questions and more anxiety on essay response formats, and Crocker and Schmitt (1987) that college students expressed more test anxiety for parallel forms of open-ended questions than corresponding multiple-choice items. Additionally, Zoller and Ben-Chaim (1988) report that students express less test anxiety when they are given examination-formats that match their preferences, and that the administration of these preferred examination formats is accompanied by increases in some students' grades. A few studies have investigated gender-related differences in test anxiety and test item-response formats. In one such study, P. Murphy (1981, cited by Gipps & P. Murphy, 1994) reported that the 15-year old boys in her study expressed particularly more anxiety about constructed-response questions than the 15-year old girls in her sample. Hence, gender-related differences in test anxiety may help explain boys' and girls' patterns of achievement on multiple-choice and constructed-response formats.

Implications of Research in Affective Characteristics

Gender-related differences have been identified for various facets of self-concept, confidence and achievement-related anxiety. These affects are contingent on situational factors, and as such, they may operate differentially for girls and boys on multiple-choice and constructed-response test formats. Consequently, these affective factors may help better understand boys' and girls' patterns of achievement according to test item-response formats.

GENDER, AFFECTIVE CHARACTERISTICS AND ACHIEVEMENT

Self-Concept and Academic Achievement

The Nature of Self-Concept

The early self-concept literature (for example, Coopersmith, 1967; Rogers, 1951) assumed that self-concept is a unidimensional construct. For example, Coopersmith (1967) theorised that individual components of self-concept cannot be isolated from each other because the construct is too heavily dominated by a general factor. However, more recent conceptualisations (for example, Harter, 1982; Marsh, 1987; Shavelson et al., 1976) posit multidimensional models. For example, Harter (1982) proposed that self-concept is composed of a general self-concept factor and highly specific factors that are independent of each other. Additionally, hierarchical models (for example, Shavelson et al., 1976) locate general self-concept at the apex of a hierarchy and perceptions of behaviour at the base. According to these models, general self-concept is differentiated into inferences about the self in academic and non-academic settings, and then into more specific inferences about the self in domains such as English, Mathematics, Peers and Physical appearance. Hierarchical models (for example, Marsh, 1987; Shavelson et al., 1976) are favoured over unidimensional models because they account for the higher correlations between behaviours and specific measures of corresponding self-concepts than with measures of general self-concept (Marsh et al., 1985; Marsh & Shavelson, 1985).

Self-Concept and Academic Achievement

Gorrell (1990) observes that the causal direction between self-concept and academic achievement has been contested between those who advance skills models (where the causal direction is from behaviour to self-concept) and those who argue for enhancement models (where the causal direction is from self-concept to behaviour). Whereas evidence can be marshalled to support either of these viewpoints, the relationship is more complex than either one or the other of these

views alone (Wigfield & Karpathian, 1991). Although self-concept is both a producer and a product of experience, the literature appears to favour the influence of enhancement effects over effects stemming from behaviour (Gorrell, 1990; Wigfield & Karpathian, 1991).

One criticism of early self-concept research is that it addressed substantive problems before issues of definition and measurement were resolved (Shavelson et al., 1976; M.A. Scheier & Kraut, 1979). Consequently, average correlations between general self-concept and academic achievement are small and positive at the best. For example, Hansford and Hattie (1982) report a mean correlation in the order of +0.18 between general self-concept and academic achievement, as do Mboya (1989) and Lyon (1993). However, higher correlations between academic self-concept and academic achievement are reported in the literature (for example, Lyon, 1993; Mboya, 1989). For example, Mboya (1989) reported a correlation of +0.52 between academic self-concept and academic achievement. It is quite clear that global measures of self-concept obscure the meaning of students' specific inferences about themselves, and that more meaningful interpretations can be made from specific measures of self-concept. In this vein, Bandura (1977, 1986) contends that self-concept research has provided little understanding of human functioning in specific domains. Instead, he proposes that research should focus on students' beliefs about their capabilities for completing individual tasks (Bandura 1977, 1984, 1986).

Implications of Self-Concept Theory and Research

It is quite clear that global measures of self-beliefs do not capture the relationship between self-concept and achievement as well as measures of self-concept that are specifically related to academic achievement in particular domains. Research relating to girls' and boys' self-related cognitions on multiple-choice and constructed-response formats requires measures of students' self-related cognitions that are sensitive to differences between multiple-choice and constructed-response test formats, a degree of specificity that cannot be elucidated by measures of self-concept.

Self-Efficacy and Academic Achievement

Self-Efficacy Beliefs and Social Cognitive Theory

Social cognitive theory (Bandura, 1986) highlights the importance of the social origins of cognition and behaviour, together with the causal influence that thought processes have on motivation, affect and behaviour. Bandura asserts that human nature is defined by capabilities for symbolising, forethought, vicarious learning, self-regulation and self-reflection. Of these capabilities, none is argued to be more pivotal than symbolic capabilities, which are used to create internal models of experiences, visualise scenarios, evaluate courses of action and to predict outcomes. Bandura (1986) asserts that these symbolic activities are antecedent to much of human behaviour.

Social cognitive theory views persons neither as automatons who mechanically respond to external stimuli, nor as autonomous agents (Bandura, 1986, 1989). Instead, it holds that a complete understanding of human behaviour is underscored by a knowledge of three sources of influence, namely, personal factors (such as, cognition, affection and emotion), behaviour and environmental influences (Bandura, 1986; Maddux, 1995b). Of the cognitions that influence behaviour, none is more important than perceptions of one's capabilities to exercise control over situations (Bandura, 1989). Bandura (1986) labels these judgements as efficacy beliefs, or perceptions of self-efficacy.

Individuals are believed to develop self-efficacy beliefs by inferential processes. Four sources of information have been described by Bandura (1977, 1986, 1997); mastery experiences, vicarious experiences, verbal persuasion and indices of physiological reactions. Other theorists (for example, Maddux, 1995b, S. Williams, 1995) add two more categories; namely, imaginal experiences and emotional reactions that are implicit to social cognitive theory and subsumed within Bandura's categories. Information from these six sources, together with personal and situational factors are examined, weighed up and integrated by self-reflective thought (Bandura, 1982, 1989; Schunk, 1989). Overall, mastery experiences are regarded as the most influential source of efficacy information (Bandura, 1986, 1997). In general, successes are

believed to enhance beliefs of self-efficacy and failures to lower them (Bandura, 1977; Schunk, 1989). Although distal experiences have less influence than proximal experiences on judgements of self-efficacy, the key factor is not temporal proximity, but rather the absence or presence of potent intervening experiences (Bandura, 1986).

The Nature of Self-Efficacy Beliefs

Bandura (1986, p.391) conceptualises efficacy as a 'generative capability in which cognitive, social and behavioral subskills must be organized into integrated courses of action to serve innumerable purposes.' Essentially, self-efficacy is a set of context-dependent beliefs about one's capability (Sherer, Maddux, Mercandante, Prentice-Dunn, Jacobs & Rogers, 1982) to exercise the control that is required to bring about these purposes (Bandura, 1989). Self-efficacy is unrelated to the actual exercise of control, the possession of component subskills or the likely outcomes and their value (Bandura, 1986; Schunk, 1989, 1990). For example, judgements of self-efficacy relate to capabilities such as driving through congested city traffic or on mountain roads, rather than to subskills such as turning on the ignition or changing lanes (Bandura, 1984, 1986). This understanding of self-efficacy is underscored by the conceptual difference between the possession of skills and perceptions that these skills can be used effectively in diverse contexts. Moreover, Bandura (1977, 1984, 1986) highlights the distinction between acts and outcomes, and between self-efficacy and outcome expectancies. Outcome expectancies are conceptually distinct from efficacy expectations (see Figure 2.1) because they relate to the expectation that once an act has been employed that it will lead to a particular outcome (Bandura, 1977, 1986, 1993).

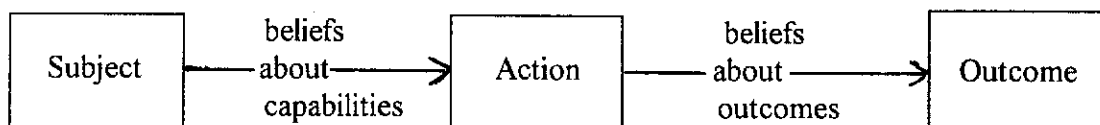


Figure 2.1

Relationship Between Self-Efficacy Beliefs and Outcome Expectancy

Although beliefs about capabilities and beliefs about outcomes are related, the distinction between the two is necessary because extraneous factors may cause outcomes to be related only loosely to a person's level of performance (Bandura, 1984, 1986, 1989). However, when outcomes are determined by competency levels, such as for cognitive tasks, self-efficacy judgements account for most of the variance in expected outcomes (Bandura, 1982, 1991, 1995).

Three dimensions of self-efficacy are implicit in its conceptualisation. The strength of self-efficacy beliefs is related to the amount of confidence that control can be exercised in specific situations, whereas the level of self-efficacy relates to the degree of difficulty that situations pose (Bandura, 1977, 1986, 1989). For example, a higher level of self-efficacy is required for an ex-smoker to refrain from smoking in presence of smokers than in the company of non-smokers (Maddux, 1995b). Generality, the third dimension of self-efficacy, relates to the extent to which people judge themselves as efficacious across a range of domains (Bandura, 1986). Generality is dependent on the extent to which behaviours and situations across domains share similar features and the extent to which they require similar skills and functioning (Bandura, 1990, cited by Maddux, 1995b).

Although Bandura (1986) asserts that judgements of self-efficacy are task-specific and that they relate to generative capabilities, decontextualised instruments have been used by researchers to measure 'generalised' self-efficacy (Randhawa et al., 1993). Examples include measures of general self-efficacy (Sherer et al., 1982), academic self-efficacy (Benson, 1989) and science self-efficacy (Smist, 1993). These generalised measures are multifaceted and they sample tasks or behaviours from the area of interest. Although the generalised self-efficacy construct provides a less precise measure of the efficacy-achievement relationship (K. Stanley & M. Murphy, 1997) it elicits additional information about student's self-evaluations of their skills (M. Berry & West, 1993; Maddux, 1995a). In this context, the importance of generalised self-efficacy is advanced by Maddux (1995a), who argues that Andre Agassi's tennis coach should be more interested in Agassi's general self-efficacy for serving with speed and accuracy and returning serves, rather than his self-efficacy for winning the United States Open tennis tournament. Hence, generalised self-efficacy provides information about component skills that may help understand individual

judgements of self-efficacy and interpret performances.

Perceptions of self-efficacy bear an apparent conceptual similarity to other cognitions about one's capabilities. For example, self-efficacy is only distinguishable from perceived competence and self-confidence according to the specificity of beliefs (Schunk, 1990). However, if self-concept was to be measured at lower levels of the self-concept hierarchy, as Marsh, Walker and Debus (1991) have suggested is possible, then self-concept and self-efficacy would still represent different constructs. Self-concept is a multi-dimensional construct that includes notions of self-confidence, self-worth, stability and the structure of self-beliefs (Pajares & M. Miller, 1994; Schunk, 1991) whereas self-efficacy is a unidimensional construct (Randhawa, Beamer & I. Lundeberg, 1993). Hence, different meanings are attached to each of these constructs. For example, the self-concept question, 'Are you a good math student?', taps different cognitive and affective processes than the self-efficacy question, 'Can you solve this specific problem?' (Pajares & M. Miller, 1994, p.194). Moreover, a person can have high perceptions of their self-efficacy and yet a low self-concept because their efficacy beliefs relate to capabilities in areas that hold little value or meaning for the person (Bandura, 1984; Maddux, 1995b). Nevertheless, the literature has reported a very strong relationship between generalised self-efficacy and self-esteem (K. Stanley & M. Murphy, 1997).

Perceptions of Self-Efficacy and Academic Achievement

Self-efficacy is postulated to have a major causal influence on behaviour. Bandura (1984, p.246) asserts that highly efficacious people have beliefs that 'produce their future rather than simply foretell it' by influencing how they think, feel, motivate themselves and behave (Bandura, 1977, 1986, 1993). For example, Bandura (1993) argues that guides to performances are shaped through the construction of anticipatory scenarios. Highly efficacious people envisage their successes and they set themselves challenges and intensify their efforts when their performances fall short, persevere in the face of failure, approach threatening tasks non-anxiously and make causal attributions that support a success orientation (Bandura, 1984). Moreover, Bandura (1997) asserts in this context that human functioning is best served by

unnecessarily optimistic self-appraisals of one's own efficacy. On the other hand, people who have low self-efficacies envisage their failure and they are characterised by avoidance behaviours, lower aspirations, apprehension and achievement-related anxiety, together with low motivation and causal attributions for luck and failure (Bandura, 1984, 1993; Benson, Bandalos & Hutchinson, 1994; Pajares & M. Johnson, 1994).

Accordingly, perceptions of self-efficacy are hypothesised to influence much of the choice and direction of students' achievement behaviours provided that students value the outcomes, have the requisite skills and maintain positive outcome expectancies (Bandura, 1977; Schunk, 1989, 1991). Additionally, in test-like situations, self-efficacy is related to persistence, better management of time and a lesser likelihood for the premature closure of problems (Bandura, 1997; Bouffard-Bouchard, Parent & Larivee, 1991; Multon et al., 1991). Moreover, self-efficacy is a better predictor of academic achievement than past performances and the possession of specific knowledge and skills (Bandura, 1977, 1982, 1984; Bouffard-Bouchard et al., 1991; Multon et al., 1991; Siegel, Galassi & Ware, 1985; Schunk, 1991). Hence, low achievement stems, not only from a lack of skills, but also from a failure to use skills because of the mediating influence that perceptions of self-efficacy have on achievement (Bandura, 1984).

The research literature reports positive correlations between measures of self-efficacy and academic achievement. J. Williams (1994), for example, reports that efficacy expectations accounted for 35% of the variance in high school students' mathematics scores. Additionally, there is evidence that students' judgements of their self-efficacy are more predictive of their achievement in academic settings than other affective measures such as self-esteem (Mone, D.D. Baker & Jeffries, 1995) and subject-specific measures of self-concept, anxiety and perceived usefulness (Pajares & M. Miller, 1994). A recent meta-analysis (Multon et al., 1991) reports an average correlation of +0.38 between self-efficacy beliefs and academic performance, with self-efficacy beliefs accounting for 14 % of the variance in academic achievement. Examination of possible moderators in this relationship showed that correlations with achievement were larger for high school students ($r = .41$) than for elementary school students ($r = 0.21$) and for self-efficacy measures that were proximal and specific to

tasks. For example, higher correlations of +0.52 were reported for measures relating to basic scholastic skills whereas lower correlations of +0.36 were reported for school grades that were less clearly matched to the form, content and timing of the self-efficacy measures.

Implications of Self-efficacy Theory and Research

Although self-related cognitions are associated with achievement, it is quite apparent that the extent of this relationship is obscured by measures that are distal and of a general dispositional nature. However, sizeable correlations greater than those reported in the science attitude and the self-concept literature are reported when students' inferences about their specific capabilities to complete designated tasks are correlated with their achievement. Hence, measures of boys' and girls' self-efficacies for answering science questions set in multiple-choice and constructed-response formats may help better understand their differing patterns of achievement on these item response formats.

Test Anxiety and Academic Achievement

Arousal Interpretations of Test Anxiety

Anxiety as a Learned Drive

Early research and theory relating anxiety and learning began with the work of S. Sarason and his associates. These researchers gave a behaviourist interpretation to anxiety, conceptualising it as a learned psychological drive. Mandler and S. Sarason (1952, p.166) theorised the existence of self-centred anxiety responses in evaluative settings that are 'manifested as feelings of inadequacy, helplessness, heightened somatic reaction, anticipations of punishment or loss of status and esteem, and implicit attempts at leaving the test situation.' These task-irrelevant responses were regarded as either dysfunctional or facilitative according to whether they competed with task-relevant drives, or whether they increased task-relevant efforts so that anxiety is reduced. Test anxiety was thought to be facilitative in easy situations and

debilitative in challenging contexts (Alpert & Haber, 1960; Covington, 1992; Tobias, 1985).

Early test anxiety research tested for the presence of the cognitive, physiological and behavioural responses that were theorised by Mandler and S. Sarason (1952). An abundant body of evidence reporting the presence of task-irrelevant cognitions (for example, Deffenbacher, 1978; Hammermaster, 1989; Morris, M. Davis & Hutchings, 1981; I. Sarason, 1960, 1984) and avoidance behaviours (Galassi, Frierson & Sharer, 1981; Geen, 1987) emerged for highly test-anxious students together with little evidence for the presence of physiological reactions (Holroyd, Westbrook, M. Wolf & Badhorn, 1978). Essentially the literature agrees with Mandler and S. Sarason's (1952) descriptions for anxious responses except that heightened perceptions of autonomic activity are reported rather than actual increases in autonomic activity (Holroyd & Appel, 1980; Holroyd et al., 1978).

The research literature reporting the presence and effects of facilitating anxiety is characterised by theoretical dilemmas and equivocal findings. For example, Sieber, O'Neil and Tobias (1977) assert that the term 'facilitating anxiety' is a misnomer, and they argue that facilitating anxiety is more closely related to motivational constructs than dread, worry and other negative feelings that characterise anxiety. Moreover, it appears that arguments about the presence or absence of facilitating anxiety are underpinned by researchers' theoretical perspectives. For example, M. Eysenck (1982) reported general support in the literature for the facilitative influence of anxiety in straightforward situations whereas Tobias (1985) reviewed the same body of literature and he reported otherwise. Moreover, Covington (1992) recently asserted from his self-worth perspective that anxiety does not facilitate the production of quality work whereas Bandura (1997) reports from his social cognitive perspective that anxiety can have a facilitative effect.

Anxiety as Emotional Arousal

A central issue underscoring early conceptualisations of test anxiety related to whether anxiety is a stable personality characteristic or a situational response to some perceived threat. Spielberger and his associates clarified and operationalised the distinction between these two aspects of anxiety. According to Spielberger (1966),

state anxiety is a transitory emotional condition that is characterised by subjective feelings of tension, nervousness and worry, and by heightened autonomic nervous system activity. On the other hand, trait anxiety was conceptualised as a dispositional personality variable that refers to an individual's proneness to anxiety during stressful situations.

Spielberger, Anton and Bedall (1976) theorised that highly test-anxious students respond to testing with elevated levels of state test anxiety. According to this conceptualisation, high levels of autonomic activity are the primary cause of diminished achievement because they trigger the worrisome thoughts that have a deleterious influence on achievement. However, there is little evidence for a relationship between emotional and physiological arousal and diminished academic achievement (Doctor & Altman, 1969; Morris & Fulmer, 1976; Spiegler, Morris & Liebert, 1968), and furthermore, there is some evidence that high and low test-anxious students display similar patterns of emotional arousal (Holroyd et al., 1978). Additionally, it appears that the symptoms of test anxiety and diminished achievement on tests are a function of how autonomic activity is perceived and appraised, rather than autonomic arousal *per se* (Holroyd & Appel, 1980).

Cognitive Conceptualisations for Test Anxiety

Empirical foundations

The development of factor analysis enabled Liebert and Morris (1967) to identify a cognitive factor and an emotional factor on Mandler and S. Sarason's *Test Anxiety Questionnaire* (1952). The cognitive component (that is, thoughts about the consequences of failure and doubts about one's ability) was denoted as worry and the emotional component (that is, subjective perceptions of autonomic arousal and affective reactions) was denoted as emotionality (Liebert & Morris, 1967; Morris et al., 1981; Morris & Liebert, 1969). It is important to note in this context that emotionality is distinct from emotional arousal.

The research evidence is abundantly clear that worry and emotionality co-vary during test-taking situations (for example, Doctor & Altman, 1969; Liebert & Morris, 1967; Morris & Fulmer, 1976; Morris & Liebert, 1969; Spiegler et al., 1968) and that

they are elicited by different cues. Emotionality has been shown to vary as a function of temporal proximity to examinations and it is best conceptualised as a reflexive response (Morris et al., 1981). For example, emotionality has been reported to increase several days before examinations (Morris et al., 1981; Spiegler et al., 1968) and then decrease during examinations (Doctor & Altman 1969, Kim & Rocklin, 1994; Spiegler et al., 1968). However, worrisome thoughts are aroused well before examinations and they persist throughout examinations (Doctor & Altman, 1969, Kim & Rocklin, 1994; Spiegler et al., 1968). Additionally, worry scores are reported to increase with ego-involving instructions, and with the difficulty and perceived importance of examinations (Becker, 1982; Deffenbacher, 1980; Hembree, 1988). Moreover, the frequency of worrisome thoughts do not change during examinations unless they are accompanied by changes in achievement expectancies (Morris & Fulmer, 1976).

The shift toward cognitive explanations to help explicate diminished academic achievement was based on empirical findings, which show that worry is inversely related to achievement and that emotionality is inconsistently related to achievement (for example, Doctor & Altman, 1969; Liebert & Morris, 1969; Morris & Liebert, 1970). For example, Morris and Liebert (1970) report that worry was inversely related to academic achievement on class examinations for high school and university students, and that emotionality was weakly correlated with achievement for their university sample, although they reported the same strength correlations for their high school sample. However, Morris and Liebert (1970) reported the absence of an association between emotionality and achievement when the effects of worry were taken into account. At times, more complex and conflicting relationships have been detected; that emotionality is not significantly related to achievement except for students who report low worry scores (Doctor & Altman, 1969; Morris & Liebert, 1969), and to the contrary, that emotionality is significantly related to the achievement of students who report high worry scores (Deffenbacher, 1977). Overall, the empirical evidence points to the importance of cognitive conceptualisations for the interaction between test anxiety and academic achievement.

Cognitive-Attentional Models for Test Anxiety and Academic Achievement.

Cognitive-attentional models of test anxiety (for example, I. Sarason, 1980, 1984; Wine, 1971, 1980) assume that personality can be understood from an information-processing perspective (I. Sarason, B. Sarason & Pierce, 1991). These models posit that test-anxious students divide their attention between task-relevant cognitions and cognitive-attentional variables such as worry, self-criticism and task-irrelevant thoughts, and that this divided attention is the cause of their diminished academic achievement. It is important to note in this context that there is a large body of empirical evidence which reports that highly test-anxious students rate themselves, the task and their abilities more negatively than their less test-anxious peers (Blankstein, Flett, Boases & Toner, 1991; Deffenbacher, 1980; Galassi et al., 1981; Ganzer, 1968; Hembree, 1988; M. Many & W. Many, 1975; I. Sarason, 1960). Consequently, it is theorised that highly test-anxious students operate in a 'dual task' mode (Hammermaster, 1989) and their task-irrelevant cognitions interfere with the recall and processing of their prior learning (I. Sarason, 1980; Tobias, 1985). Therefore, cognitive-attentional models do not conceptualise any facilitative effects for test anxiety (Kim & Rocklin, 1994).

The divided attention hypothesis provides an information-processing explanation consistent with the large body of research evidence that achievement is inversely related to worry and generally unrelated to emotionality. Moreover, additional support is provided by reports that students' diminished achievements are most pronounced when their cognitive resources are strained (Deffenbacher, 1978; Hembree, 1988; Paulman & Kennelly, 1984; I. Sarason, 1980, 1984), and by reports that the achievements of highly test-anxious students improve when the information-processing demands of tasks decrease (Tobias, 1985).

However, cognitive-attentional formulations of test anxiety do not explain the extent of the variation in students' achievement scores (Birenbaum & Nasser, 1994; Blankstein et al., 1991; Bruch, 1981; Bruch, Juster & Kaflowitz, 1983; R. Smith, Arnkoff & Wright, 1990). Moreover, there is some evidence that depressed test performances are not associated with the presence of negative thoughts *per se* (Blankstein et al., 1991; Bruch et al., 1983; Bruch, Kaflowitz & Kuethe, 1986; Galassi et al., 1981). Blankstein et al. (1991), for example, assert that high and low test-

anxious students appear to be differentiated from each other by the extent of their negative self-cognitions and their failure to engage in task-facilitative thoughts, rather than by their diminished achievement. Finally, there is little evidence that cognitive treatments for test anxiety are accompanied by concomitant increases in academic achievement unless they are combined with other treatment methods (Hembree, 1988; Tryon, 1980).

Self-Regulatory Models

Although the empirical evidence supports a cognitive conceptualisation for test anxiety, cognitive-attentional frameworks provide a limited understanding of the relationship between test anxiety and academic achievement. However, self-regulatory models of behaviour (for example, Bandura, 1991, 1997; Carver & M.F. Scheier, 1991) conceptualise alternative cognitive frameworks between academic achievement and test anxiety in terms of human agency.

According to these frameworks, cognitive variables such as perceptions of self-efficacy, outcome expectancies and goal-related mechanisms underscore how people influence their thoughts, motivations and behaviours (Bandura, 1989, 1991, 1997; Carver & M.F. Scheier, 1991). Although strong perceptions of self-efficacy are related to efficient cognitive processing (Maddux & Lewis, 1995), the focus of the self-regulatory models lies with human agency. These models posit mechanisms whereby people establish goals that are used as reference points to monitor their progress and adjust their actions so that they conform with their goals. In contrast to cognitive-attentional conceptualisations, self-focussed attention does not necessarily result in dysfunctional performances (Carver & M.F. Scheier, 1991). Instead, the latter authors argue that one must be self-focussed before being task-focussed, and Bandura (1991, 1997) asserts that the difference between facilitation and dysfunction is dependent on the meanings that are attached to self-cognitions. For example, anxiety is produced when self-regulatory processes are at odds with a perceived inefficacy to control situational variables and apprehensive cognitions (Bandura, 1989, 1991; Carver & M.F. Scheier, 1991). Accordingly, Carver and M.F. Scheier (1991) theorise that anxiety has a causal influence on performance due to the interruption of self-regulatory activities. However, Bandura (1991, 1997) asserts that anxiety and

diminished achievement are co-effects. Notwithstanding these points of difference, both perspectives conceptualise diminished achievement in terms of the withdrawal of highly anxious people from threatening situations.

Research findings that attribute students' diminished academic achievement to the subjective meanings they attach to their negative cognitions are consistent with self-regulatory frameworks. In one such case, Bruch et al. (1986) reported that 35% of the variance in test scores was associated with variability in the subjective meanings that students attach to their covert thoughts. Moreover, those students who expressed strong beliefs in their competency reported a facilitative effect for their negative cognitions. A few studies (for example, Galassi et al., 1981; Prins, Groot & Hanewald, 1994; Zatz & Chassin, 1985) have also considered the role of on-task and coping cognitions. Whereas Galassi et al. (1981) and Blankstein et al. (1991) report that the proportion of positive cognitions among highly test anxious college students was inversely related to levels of test anxiety, other studies (Prins et al., 1994; Zatz & Chassin, 1985) report that highly test-anxious school students have a higher incidence of on-task and coping thoughts in addition to a greater number of off-task thoughts and negative self-evaluations. Moreover, both Prins et al. (1994), and Zatz and Chassin (1985) report that on-task cognitions are unrelated to task performance and that coping cognitions are negatively related to task performance. Findings from these studies that are consistent with Kendall's (1984) assertion that personal gains through self-encouragement are due to the absence of negative thinking rather than the power of positive thinking.

Skills Deficit Models

Skills deficit models for test anxiety posit that test anxiety is caused by an awareness of either inadequate preparation for tests or low test-taking skills, and that, as a consequence, students are likely to achieve poorly (Tobias, 1985). Hence, test anxiety and diminished achievement are conceptualised by these models as co-effects.

Skills deficit conceptualisations are supported by correlational data that associate high levels of test anxiety with deficient study skills and test-taking skills. For example, test-anxious students procrastinate on academic tasks and they are more

likely to have ineffective study habits (Birenbaum & Nasser, 1994; Culler & Holahan, 1980; Hembree, 1988; P. Kalechstein, Hocevar, Zimmer & M. Kalechstein, 1989; Wittmaier, 1972). Moreover, Culler and Holahan (1980) report from their study no support for the common stereotype of the highly test-anxious student who knows the subject matter but ‘freeze[s] up at test time’ (p.18). Instead, high and low test-anxious students are reported to differ in their knowledge of test-taking strategies (Birenbaum & Nasser, 1994; Bruch, 1981). Additionally, Paulman and Kennelly (1984) report that grade point average is unrelated to anxiety after controlling for the extent of students’ skills. However, other studies (for example, Birenbaum & Nasser, 1994; Everson, Millsap & Browne, 1989; Hunsley, 1985; Paulman & Kennelly, 1984; R. Smith et al., 1990) report that cognitive conceptualisations of test anxiety are implicated in the relationship between test anxiety and diminished achievement as well as inadequate preparation and deficient skills. For example, Everson et al. (1989) report that test anxiety and skills deficits each contributed toward diminished achievement on reading comprehension and mathematics examinations for university students. Moreover, reports that training in study skills has little effect on either test anxiety or academic achievement (Hembree, 1988) and that such training is most useful when it is combined with other treatments for reducing test anxiety (Hembree, 1988; Tryon, 1980) are consistent with these findings.

Implications of Test Anxiety Theory and Research

Overall, the evidence suggests that cognitive-attentional, self-regulatory and skills deficit variables are associated with diminished academic achievement. Additional evidence for this viewpoint is provided by R. Smith et al. (1990) who report that each of these conceptualisations makes a unique contribution to understanding the academic achievement of college students. However, skills deficit variables should equally apply to both multiple-choice and constructed-response test formats, provided that each of these test formats measures similar constructs and there are no differences between girls and boys in their test-taking skills. It is important to note in this context, that this review has not reported any gender differences in test-taking skills or in the skills that are required to answer multiple-

choice and constructed-response test formats. Furthermore, gender-related differences in students' perceptions of their skills, irrespective of whether these skills exist, are subsumed by the self-efficacy construct. Hence, cognitive-attentional and self-regulatory models of test anxiety provide a theoretical base that may help to better understand boys' and girls' patterns of achievement on multiple-choice and constructed-response items.

A THEORETICAL AND METHODOLOGICAL FRAMEWORK FOR INVESTIGATING BOYS' AND GIRLS' FORMAT-RELATED SCIENCE ACHIEVEMENTS

Modelling Girls' and Boys' Science Achievement According to Test Item-Response Formats

Although research has investigated the relationships between perceptions of self-efficacy and academic achievement, and test anxiety and achievement, no research appears to have investigated boys' and girls' academic achievements on multiple-choice and constructed-response formats for these variables. Hence, an original contribution to the research literature can be made in this domain. Additionally, it is important to recognise that the extant test anxiety and self-efficacy literature, on which the following conceptualisations are based, has limited generalisability because many of its findings are derived from laboratory experiments rather than classroom settings (Gross, 1990; Hunsley, 1985; Schunk, 1991, 1995). Furthermore, relatively few studies have been conducted among high-school age pupils. For example, J. Williams (1991) reports that less than 18 % of the 562 test anxiety studies integrated by Hembree (1988) were derived from high school samples. Instead, most test anxiety research has involved small groups of college and university students in North America.

Notwithstanding these inherent limitations, it is possible to conceptualise a model to promote a better understanding of high school girls' and boys' science achievements on multiple-choice and constructed-response formats in terms of their self-efficacy and their worrisome cognitions. The literature indicates that cognitive-

attentional and self-regulatory conceptualisations of test anxiety, together with boys' and girls' perceptions of self-efficacy are most likely to be implicated in gender-related differences in science achievement according to test item-response formats.

The model (see Figure 2.2) shows that perceptions of self-efficacy for answering individual test questions are dependent on situational variables such as the content and difficulty of individual test items, and also on girls' and boys' generalised self-efficacies. In this context, generalised self-efficacy refers to students' confidences for answering multiple-choice and constructed-response tests.

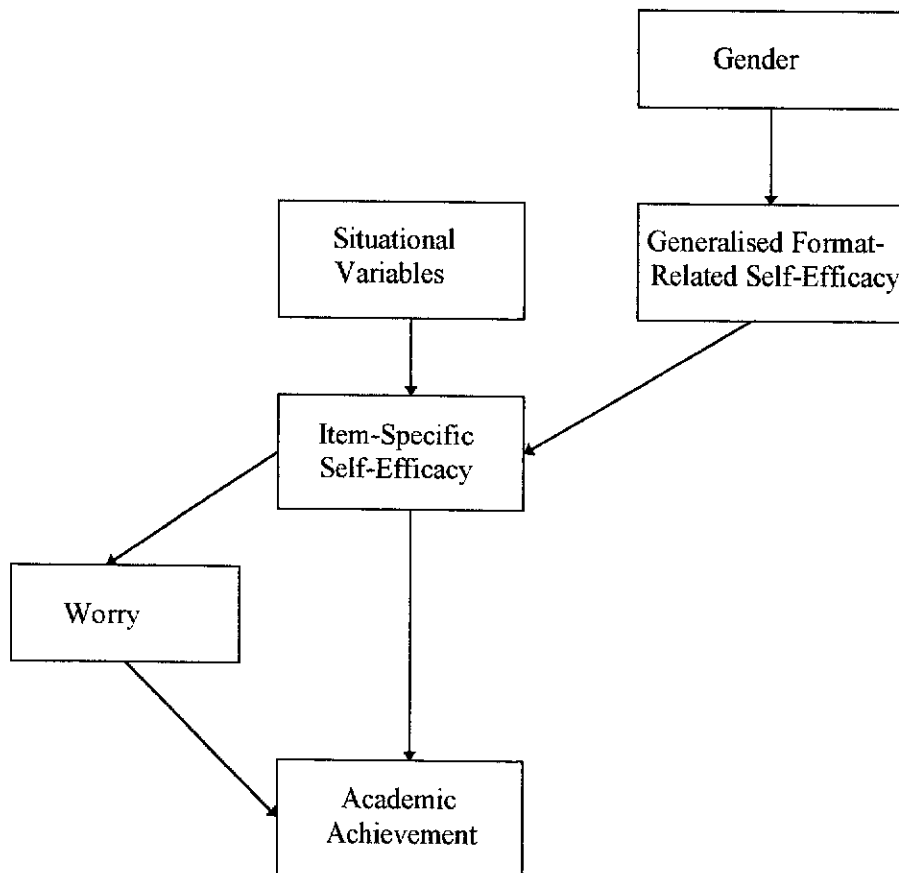


Figure 2.2

Model for Boys' and Girls' Achievements on Equivalent Multiple-Choice and Constructed-Response Test Items

Perceptions of self-efficacy for answering individual test questions are more specific than these generalised beliefs of self-efficacy and they involve cognitive appraisals of the situational variables at the level of the individual test item. Additionally, social cognitive theory asserts that perceptions of self-efficacy exert a causal influence on

achievement and that worrisome thoughts co-vary with achievement. On the other hand, cognitive-attentional conceptualisations of test anxiety posit that perceptions of inefficacy are antecedent to worrisome cognitions, and that worrisome cognitions interfere with students' performances on tests. Because there appears to be empirical evidence for self-regulatory and cognitive-attentional conceptualisations, it is proposed in this context that both processes act in concert with each other. As noted previously, skills deficit models of test anxiety do not appear to be implicated in this model because multiple-choice and constructed-response items appear to measure the same constructs, and also because there is little evidence for any gender-related differences in test-taking behaviours such as the omission of difficult items and guessing tendencies.

According to this model, gender has an indirect influence on girls' and boys' self-efficacies for answering individual test items, and this influence is mediated through boys' and girls' generalised self-efficacies for answering multiple-choice and constructed-response items. Although gender could have a direct influence on girls' and boys' item-specific self-efficacies by interacting with situational variables (such as the context of individual items), this model does not show this relationship because it is developed for equivalent multiple-choice and constructed-response test items.

Consequently, high school boys' and girls' patterns of achievement on multiple-choice and constructed-response formats may be illuminated by an examination of their relative science achievements on multiple-choice and constructed-response test questions, together with gender-related differences in perceptions of self-efficacy, generalised self-efficacy and worrisome cognitions. Such research should, as identified earlier, go beyond the process of identifying the extent of any such differences to investigate the subjective meanings that girls and boys attach to these cognitions to explicate their relationship with achievement.

Methodological Considerations

Critique of Self-Efficacy and Test Anxiety Research Methodologies

To date, an overwhelming majority of self-efficacy and test anxiety research has employed quantitative research techniques, which have been reliant on self-report measures of self-efficacy and test anxiety. Schunk (1991, 1995) argues that many self-efficacy research paradigms have failed to take into account the complexity of classroom activities, and he identifies the richness of alternative types of data. The same criticism applies for much of the test anxiety research and it parallels the wider debate between the relative merits of the qualitative and quantitative research traditions (for example, Gage, 1989). Essentially, critics of the quantitative research methodologies point to the complexity of human affairs, and they argue that quantitative analyses unduly simplify human behaviour and that they devalue participants' meaning-perspectives (Gage, 1989). For example, most test anxiety measures report the relative presence of test anxiety but neither its content nor its meaning for individual students (Wigfield & Eccles, 1989; Wine, 1980).

Self-Efficacy

Although the self-efficacy research literature invariably acknowledges its origin to the ideas of Bandura, much of the self-efficacy research is based on 'poorly defined constructs' (Pajares & M. Miller, 1994, p.194) that are at odds with his conceptualisation of self-efficacy (Maddux, 1995b; Owen & Froman, 1988; Pajares & M. Miller, 1994). For example, self-efficacy measures have confounded beliefs about outcomes with self-efficacy and sought global assessments of subskills rather than specific measures of generative capabilities (Langenfeld & Pajares, 1993; Pajares & M. Miller, 1994; Vispoel & Chen, 1990). Furthermore, some self-efficacy measures have little conceptual similarity with criterial measures (Langenfeld & Pajares, 1993; Pajares & M. Miller, 1994). For example, confidence to succeed has been matched with academic performance (for example, Benson, 1989) and capabilities to complete mathematics courses with problem solving (for example, Randhawa et al., 1993).

Pajares and M. Miller (1994) assert that the mismeasurement of self-efficacy is a recurring theme in the literature and that it leads to confounded relationships, ambiguous findings and uninterpretable results.

Bandura (1977, 1986) argues that the most precise measure of the relationship between judgements of self-efficacy and criterial measures requires the detailed assessment of the level, strength and generality of self-efficacy beliefs. In traditional operationalisations of self-efficacy (for example, Marsh et al., 1991), individuals are presented with graduated self-efficacy scales for a series of tasks within the domain under investigation. These tasks are arranged in order of increasing difficulty or stressfulness. The level of self-efficacy is measured by asking respondents to identify the tasks they can do, and the strength of their self-efficacy by asking them to estimate the degree of certainty with which they can execute the necessary behaviours. However, traditional operationalisations of self-efficacy are problematic for cognitive domains (Bandura, 1986; Schunk, 1991) because the operations that are required to solve tasks are not always apparent, which makes ordering these tasks difficult (Schunk, 1991). Moreover, Schunk (1991) asserts that there is no evidence to suggest that self-efficacy constitutes hierarchical scales in cognitive domains as it does in other domains of functioning.

A survey of the literature shows that most self-efficacy research within the cognitive domain has measured the strength of self-efficacy beliefs and ignored the level and generality of self-efficacy perceptions (Maddux, 1995b; Vispoel & Chen, 1990). This approach is justified because evidence from studies that have measured both the level and the strength of self-efficacy perceptions show that their correlations are sufficiently high (Owen & Froman, 1988; Vispoel & Chen, 1990) to make the measurement of one or the other of these dimensions redundant (Owen & Froman, 1988). Additionally, Vispoel and Chen (1990) report that the meaning of generality, the third dimension of self-efficacy, is ambiguous and that concerns about scope validity makes its measurement problematic.

Although the self-efficacy and the generalised self-efficacy constructs are related, they are conceptually distinct. Nevertheless, the collection of data for both of these constructs is relevant for the purpose of developing a better understanding of boys' and girls' patterns of science achievement according to test item-response

formats. In this context, the measurement of self-efficacy provides an index of the self-related cognitions that students hold about their capabilities for answering individual test items. The literature suggests that it is sufficient to measure the strength of these cognitions in a quantitative manner, similar to the traditional operationalisation for self-efficacy. Additionally, an investigation of girls' and boys' generalised self-efficacies for answering questions set in multiple-choice and constructed-response formats should facilitate interpretation of their self-efficacy judgements. Such data, which could seek students' meaning-perspectives about each of these response formats would make necessary the use of qualitative research techniques.

Worrisome Cognitions

Test anxiety research has generally used specially constructed instruments for measuring students' anxiety in test or test-like situations. These test anxiety scales provide a closer match between the anxiety that is experienced in these settings and academic achievement than measures of general anxiety. For example, Seipp (1991) reports from her meta-analysis a correlations of -0.23 between measures of test anxiety and achievement and a correlation of -0.16 between measures of general anxiety and achievement. However, early test anxiety instruments, such as the *Test Anxiety Questionnaire* (Mandler & S. Sarason, 1952), the *Test Anxiety Scale* (S. Sarason, Davidson, Lighthall & Waite, 1958) and the *Achievement Anxiety Test* (Alpert & Haber, 1960) are unsuitable for this research because they provide global measures of test anxiety, rather than separate scores for worry. Consequently, data obtained from these instruments have limited explanatory power because they confound worry with emotionality (Seipp, 1991).

Additionally, many of the early test anxiety measures (for example, the *Test Anxiety Questionnaire*, *Test Anxiety Scale*, *Achievement Anxiety Test*) conceptualise test anxiety as a dispositional personality trait, rather than as a situational response to tests. Notable exceptions include Liebert and Morris' (1967) *Worry-Emotionality Questionnaire*, which measures state test anxiety and Spielberger's (1980) *Test Anxiety Inventory*, which conceptualises test anxiety as a situation-specific personality

trait. Nevertheless, Seipps' meta-analysis shows, contrary to expectations, that the traditional test anxiety instruments all correlate similarly with achievement. Her results also show that self-reported anxiety is more pronounced when it is measured after tests than before tests.

For the present study, it was considered that the explication of boys' and girls' differences in achievement according to test item-response formats requires the administration of a scale that measures the extent of students' worrisome cognitions for multiple-choice and constructed-response questions. However, it was also recognised that these types of measures are limited because they reveal only the relative presence of worry without understanding the processes by which these data are generated (L. Wolf & J. Smith, 1995). Moreover, as identified earlier, the explanatory power of the worry construct is increased by a knowledge of the subjective meanings that students attach to their worrisome cognitions. This type of data, where students report their meaning-perspectives, was considered to be best yielded by qualitative procedures.

SUMMARY

Examination of the research literature shows that girls' and boys' relative achievements across a wide range of curricular areas are influenced by the choice of test formats that are used to assess their knowledge and understanding. The hypothesis most likely to explicate these differences posits the operation of all-pervasive construct-irrelevant factors. Thus, for this present study, a decision was made to collect qualitative and quantitative data relating to boys' and girls' self-efficacies and their worrisome cognitions, for the purpose of developing a better understanding of girls' and boys' patterns of achievement according to test formats. The research questions which framed the study and the methodology employed to provide answers to these questions are described in detail in the next chapter.

CHAPTER 3

RESEARCH METHODOLOGY

OVERVIEW

This chapter builds on the theoretical and methodological base provided in the previous chapter to describe the methodology that was employed to investigate boys' and girls' patterns of science achievement on multiple-choice and constructed-response formats. The first section of this chapter describes the research design (including the three questions developed to guide the research) and research setting; the second section describes the development and properties of the research instruments; and the final section describes the procedures that were used to collect the data.

RESEARCH DESIGN

Research Questions

The previous chapter culminated in the construction of a model that linked girls' and boys' perceptions of self-efficacy and their worrisome cognitions to their academic achievement. The literature review also highlighted the importance of the subjective meanings that boys and girls attach to their thoughts during tests. Hence, the initial questions guiding the research for the Year 9 students who attended one particular school were;

- Are there gender-related differences in the content and extent of students' self-efficacies for answering science achievement questions set in multiple-choice and constructed-response formats? And consequently: is there a gender-related interaction between girls' and boys' perceptions of self-efficacy on science achievement questions and the response format these questions require?
- Are there gender-related differences in the content and extent of students' worrisome cognitions for answering science achievement questions set in multiple-choice and constructed-response formats? And consequently: is there

a gender-related interaction between boys' and the girls' worrisome cognitions on science achievement questions and the response format these questions require?

- How are the girls' and the boys' perceptions of self-efficacy and worry related to their achievement on multiple-choice and constructed-response test formats?

Conceptual Issues

As noted in the previous chapter, the self-efficacy and test anxiety research literature is dominated by quantitative approaches that have utilised experimental and quasi-experimental research designs. Recent commentators (Schunk, 1991, 1995; Wigfield & Eccles, 1989; L. Wolf & J. Smith, 1995) have pointed to the limitations associated with these approaches and they have suggested the need for research that is informed by qualitative as well as quantitative data. Furthermore, Krockover and Shepardson (1995) have called for researchers to address the 'missing links' (p.223) in gender equity research, by going beyond the reporting of quantitative findings for isolated studies to the development of contextual understanding. However, considerable debate exists in the literature regarding the relationship between the quantitative and qualitative research traditions (for example, Firestone, 1987; Gage, 1989; Guba & Lincoln, 1989; Hammersley, 1998; Hitchcock & Hughes, 1989; Howe, 1988; Lincoln & Guba, 1985; Miles & Huberman, 1984; S. Miller, Nelson & M. Moore, 1998; Peshkin, 1993; Schrag, 1992; J. Smith, 1997; J. Smith & Heshusius, 1986), which are otherwise broadly denoted as the positivist and the interpretivist-constructivist paradigms (Lincoln & Guba, 1985).

An examination of the research methodology literature indicates two conflicting approaches to the debate — that of the 'conceptual purists' and the 'pragmatists.' Conceptual purists (for example, Guba & Lincoln, 1989; Hitchcock & Hughes, 1989; J. Smith & Heshusius, 1986) locate the debate at the paradigmatic level. According to their viewpoint, the positivist and interpretivist-constructivist research paradigms are incompatible because their ontological and epistemological foundations prescribe different types of research strategies. Hence, it is not the combination of qualitative and quantitative data that is problematic for these commentators, but rather the

combination of research strategies that are derived from differing assumptions about the nature of the world and the purpose, role and approach of the researcher (Guba & Lincoln, 1989; Hitchcock & Hughes, 1989; J. Smith & Heshusius, 1986).

Consequently, conceptual purists argue for an 'either/or' approach toward the research traditions. However, their critics argue that the positivism they criticise has been largely replaced by postpositivism (Howe, 1988; Reichardt & Rallis, 1994a), and furthermore, that conceptual purists engage in a form of reductionism by ignoring the common conceptual ground between postpositivist and interpretivist-constructivist approaches (Hammersley, 1998; Reichardt & Rallis, 1994a; M. Smith, 1994).

The distinguishing feature of pragmatic approaches is the triumph of the empirical over the conceptual (Howe, 1988), that is, the '*looking away from first things, principles, 'categories', supposed necessities; and of looking towards last things, fruits, consequences, facts*' (Gage, 1988, quoting James, [1907] 1955, p.47, italics in original). Similarly, Howe (1988) asserts that the worth of paradigms lies in 'how well they inform, *and are informed by*, research methods that are successfully employed' (p.10, italics in original). Consequently, pragmatists locate the debate between the research traditions at the technical level. Some commentators (for example, Gage, 1989; Howe, 1988) ameliorate paradigm differences by regarding the research traditions as no more than different disciplinary orientations to questions of interest and judgements about how these questions should be investigated. Therefore, they do not regard these different approaches as logically oppositional, and their postpositivist research designs combine methods shaped by assumptions from each of the research traditions (S. Miller et al., 1998). However, Guba and Lincoln (1989) contend that these types of mixed-method designs do no more than incorporate qualitative data into a positivist framework.

The stance adopted in this present research was to regard the research paradigms as incommensurable (in the sense that they cannot be blended together), but not logically oppositional. In this sense, S. Smith et al. (1998, p.408) poses the question (emphasis added):

What might happen if we collectively reframed the debates about paradigms, viewing them as positive, not negative, and allowing *generative dialogue* rather than battlefield metaphors to shape new visions for the future of our field?

Therefore, the question driving the research design was not which of the paradigms to

adopt, but rather how to utilise perspectives from each of the paradigms. In the context of this present study, Bernstein's (1991) notion of 'dialogical encounter' (p.337) was used to inform the research design.

Originally, Bernstein (1991) advanced his notion of dialogical encounter as a way of resisting tendencies toward fragmentation in the academic field of philosophy, which like education, is characterised by a plurality of traditions and orientations. He, that is, Bernstein (1991) proposes a dialectical approach whereby differing philosophical orientations are viewed as 'conversational partner[s]' (p.337) of the other. According to this scheme, which is consistent with the 'traditional scholarly norm of openness' (Donmoyer, 1996, p.20), any one of the research traditions would not be assimilated into the categories of the other, trivialised, treated with indifferent tolerance or dismissed outright (Donmoyer, 1996).

In educational contexts, Donmoyer (1996) reports using dialogical encounter as a way of developing the annual meeting program of the American Educational Research Association and reviewing manuscripts for publication. However, very few single research studies appear to have been designed around principles consistent with Bernstein's (1991) dialogical encounter. Instead, researchers have been cautioned against conducting a 'sophisticated quantitative study while doing an in-depth qualitative study simultaneously' (Bogdan & Biklen, 1992, p.39). Nevertheless, S. Miller et al. (1998) do not exclude the possibility that these complex approaches may constitute good research.

Finally, in practical terms, research designs that relegate qualitative data to the roles of illustrating and exemplifying findings uncovered by quantitative data analyses, or to 'making real' the particulars of the case under study were rejected in this present study. Instead, approaches from each of the research traditions were used, but at differing levels of analysis to address the questions guiding this study. Although data from the positivist paradigm and the interpretivist-constructivist paradigm are analysed in separate chapters, the findings from each of these research traditions are interwoven to produce the working hypotheses that emerge from this study.

The Research Context

Research Site

The research data were collected at the college where the researcher was employed to teach science and religious education. This school is a private, coeducational, P-12 college with a roll of about 1100 students and it is located in a rapidly growing urban area in South-East Queensland. The college is operated by a mainline Protestant denomination and it enjoys a high reputation for its academic standards and its support of students. The school is well-resourced, and in 1996 the tuition fees were moderate at \$2,800 per annum for students in the secondary department.

The college has a commitment to social justice, which is underscored by the Gospel values that it assumes. Issues relating to the relative participation of girls and boys in school activities and their academic achievements have been discussed at staff level and they have been of concern for the last eight years. The chief concerns, however, have related to boys' lack of participation in school activities, their reluctance to take on leadership roles and their relatively lower levels of academic achievement in the junior high school years. There have been no formal discussions at the staff level regarding the interplay between gender-related differences in achievement and methods of assessment.

The Students

Although the students who attend this college come from a wide range of socio-economic backgrounds, a large majority of the parents who send their children to this college enjoy higher than average household incomes. Anecdotal evidence suggests that the parents who send their children to this school generally hold high expectations for their achievement. The overwhelming majority of the students are of Anglo-Saxon origin. For example, only three students from the classes that participated in the research had non-European backgrounds and less than 2% of the students spoke languages other than English at home. Hence, the students attending this college are

not representative of the general population of Year 9 science students in the state of Queensland, where substantially higher proportions of the students have multi-cultural backgrounds and less privileged socio-economic backgrounds.

Research data were collected from 5 intact Year 9 (1996) science classes (73 boys and 81 girls). As part of the normal school policy and procedures, students had been assigned to these classes at the end of the previous year so that the classes would be comparable for their gender balance, and students with leadership potential, patterns of disruptive behaviours and special needs. Anecdotal evidence from teachers at the research site, together with the experience of the researcher, indicated that the Year 9 classes in 1996 were particularly difficult to interest, motivate and teach. The ages of the students, as reported on the *Survey About Science Tests*, ranged from 13 years and 10 months to 15 years and 8 months, and the median age was 14 years and 6 months.

The Instructional Context

Year 9 is the second year of secondary education in Queensland. Junior secondary curriculum documents are developed within schools in response to syllabi prescribed by the Department of Education in Queensland. Although non-government schools are not obliged to follow these guidelines, many do so in practice. Additionally, the quality of students' work is assessed and moderated within individual schools.

Science is a compulsory subject at the school until Year 11. Year 9 students study four science topics each year; a biology, chemistry, physics and an earth science topic, each for the duration of one school term. Pen-and-paper achievement tests that sample the unit objectives are given at the conclusion of each unit. These achievement tests are generally composed of multiple-choice and short-answer questions and they contribute two thirds of the marks toward the semester totals. Students have not been required in the past to construct extended-response answers, such as paragraph answers, during their science tests. Consequently, the research was limited to the use of multiple-choice and short-answer item-response formats. Additionally, the junior science work program prescribes that the marks in

achievement tests are equally divided between content outcomes and process outcomes.

Each Year 9 science teacher at the research site is given the responsibility for planning one of the science units and developing a series of common assessment tasks. Hence, there is considerable uniformity between the classes insofar as the content of their science instruction and their assessment are concerned. There were considerable differences, however, between the Year 9 science teachers in the ways they related to their classes and in the ways they implemented the common curriculum.

The Cooperating Teachers

A final year student-teacher on her practicum and four teachers, other than the researcher, were involved in teaching the Year 9 science classes while the research data were collected. These teachers ranged from a recent graduate with one semester's teaching experience to three teachers with more than 15 years' experience each. Two of the Year 9 science teachers were females and the other two were males. The researcher was well-known to these teachers, apart from the graduate. The student-teacher completed her practicum under the guidance of the researcher, over a period of six weeks during the third school term.

The cooperating teachers were supportive of the research program and several indicated an interest by initiating discussions about methodological issues and the emerging findings. Although these teachers were cooperative, they initially expressed some reservations about the use of two equivalent tests rather than a common test as they had done in the past. Additionally, the data collections were subjected to several limitations that were necessary to maintain positive working relationships with the cooperating teachers. These limitations accommodated these teachers' preferences for awarding half marks for partially correct constructed-response answers and the allocation of students to the research groups.

Overview of the Research Design

In operational terms, the purpose of this study was to contribute toward a better understanding of boys' and girls' science achievements on multiple-choice and short-answer constructed-response formats. As indicated earlier, a 'binocular' (Reichardt & Rallis, 1994b) view of the research questions was provided by a positivist research design and an interpretivist-constructivist research design. As described in detail later in this chapter, data from quasi-experimental and survey methodologies were collected for coarse-grained analyses that aggregated data and compared girls' and boys' format-related achievements, self-efficacies and worrisome cognitions. Additionally, interview data were collected in the interpretivist-constructivist design for fine-grained analyses that clarified and extended the theory relating to the model through individual students' meaning-perspectives. The researcher also maintained a journal during the course of this present study. Entries in this journal related to the research context, the implementation of the quasi-experimental data collections, as well as the ongoing and preliminary analyses of the interview data. Specific details of the procedures that were employed to enhance the *quality* of the study, that is, the validity and the reliability of the positivist data (see Gable & M. Wolf, 1993; Miles & Huberman, 1994), and evidentiary warrants for assertions in the interpretivist-constructivist design (see F. Erickson, 1986) are described throughout this chapter.

The quasi-experimental methodology involved the collection of data by the administration of test booklets containing specially designed achievement questions set in multiple-choice and constructed-response test formats, self-efficacy scales and a worry instrument under examination conditions. (The construction of these research instruments is described later in this chapter). These data were collected during end-of-unit achievement tests because the consequential nature of summative testing influences conative, affective and performance variables (I. Sarason, 1980; L. Wolf & J. Smith, 1995). Additionally, these data collections were replicated to determine the stability of findings over unit topics because there is some evidence that girls' and boys' relative achievements on science tests are influenced by unit contexts (Bransky & Qualter, 1993; G. Erickson & L. Erickson, 1984; Jovanovic et al., 1994).

Two possible confounding influences, namely, familiarity with the content and

the order of presentation for the test item-response formats were eliminated by the administration of four separate test booklets (see Figure 3.1) to equivalent groups of students. The problem of familiarity, whereby students' achievement, perceptions of self-efficacy and the extent of their worrisome cognitions could be influenced if they were to answer two sets of questions that differ only in their response format was addressed by the construction of two sets of achievement questions for each of the test item-response formats. According to this scheme, the multiple-choice questions in test booklets 1A and 1B were equivalent to the constructed-response questions in test booklets 2A and 2B, and the constructed-response questions in test booklets 1A and 1B were equivalent to the multiple-choice questions in test booklets 2A and 2B.

<u>Test booklet 1A</u> Multiple-Choice questions (Set 1) <i>Attitude Toward Tests Survey</i> Constructed-Response questions (Set 1)	<u>Test booklet 1B</u> Constructed-Response questions (Set 1) <i>Attitude Toward Tests Survey</i> Multiple-Choice questions (Set 1)
<u>Test booklet 2A</u> Multiple-Choice questions (Set 2) <i>Attitude Toward Tests Survey</i> Constructed-Response questions (Set 2)	<u>Test booklet 2B</u> Constructed-Response questions (Set 2) <i>Attitude Toward Tests Survey</i> Multiple-Choice questions (Set 2)

Figure 3.1
Comparison of the Test Booklets

All students responded to either one or the other of the equivalent questions. The possible confounding influence of the order of presentation for the item-response formats was overcome by requiring one half of the students to complete the multiple-choice questions first (Test booklets 1A and 2A) while the remainder of the students completed the constructed-response questions first (Test booklets 1B and 2B). The *Attitude Toward Tests Survey* was included in the test booklets only once, immediately after the first set of achievement questions because time constraints precluded the administration of the *Attitude Towards Tests Survey* for a second time.

Consequently, the extent of the boys' and the girls' worrisome cognitions was measured only for the first set of questions they completed on each of the achievement tests.

The survey methodology (*Survey About Science Tests*) sought qualitative data from all the students who participated in the quantitative data collections, partly to collect data relevant to the research questions, and partly to identify students who would be selected for the interviews. Additionally, the qualitative data yielded by this survey was collected to triangulate (Denzin, 1988; Mathison, 1988) with the other research data to increase the rigour of this study.

An interview methodology was employed to describe and develop an understanding of the subjective meanings that students attach to multiple-choice and constructed-response test items, their worrisome thoughts and their self-efficacies during testing. These interviews were conducted over two interviewing cycles according to a predetermined interview guide (Taylor & Bogdan, 1984) and situational cues. In the first interviewing cycle, students responded to a series of 'flexible and dynamic' (Taylor & Bogdan, 1984, p.77) open-ended questions that were derived from the research questions. These questions allowed the development of new questions for the following interviews. In this sense the researcher, rather than the interview questions, was the research tool (Miles & Huberman, 1994; Taylor & Bogdan, 1984). The second interviewing cycle continued after the preliminary analyses of the interviews from the first cycle were complete, and after revisions were made to the interviewing guide to accommodate the new perspectives that emerged during these analyses. Emerging hunches and provisional findings were recorded by the researcher in a journal throughout the interviewing. In all, 22 students were purposely selected (Merriam, 1988) for interviewing, on the basis of their responses on the test booklets and the *Survey About Science Tests*, according to their potential for yielding disconfirming data (F. Erickson, 1986). Additional details regarding the selection of these students are provided later in this chapter.

The students were generally quite cooperative in providing the research data, according to evidence from the interviews, examination of their responses on the research instruments and reports from the cooperating teachers. Only one student declined to be interviewed. Nevertheless, some students expressed their concerns that

the tests might not have been equivalent and that they might not have had enough time to complete the test booklets in the allocated time.

Research Plan

The data collections were planned to occur in three phases (see Table 3.1).

Table 3.1
Overview of the Research Plan

Timeline	Specific Tasks
Phase 1 – Instrument development (affective variables)	
March – May 1996	Development of self-efficacy scale Pilot and refinement the of self-efficacy scale
May – June 1996	Development of the <i>Attitude Toward Tests Survey</i> (Worry) Pilot and refinement of the <i>Attitude Toward Tests Survey</i>
Phase 2 – Development and administration of test booklets	
May –June 1996	Development of first achievement test (<i>Earth Science</i>) Administration of the <i>Earth Science</i> test booklets
August–September 1996	Development of second achievement test (<i>Organisms and Food</i>) Administration of the <i>Organisms and Food</i> test booklets
Phase 3 – Collection of the qualitative data	
October 1996	Development of the <i>Survey about Science Tests</i> Pilot and refinement of the <i>Survey About Science Tests</i>
October – December 1996	First Interviewing cycle & ongoing analyses Preliminary analysis of interview data Second Interviewing cycle

The first phase focused on the development of the instruments that were used to measure the extent of the girls' and the boys' perceptions of self-efficacy and the extent of their worrisome cognitions for multiple-choice and constructed-response test questions. The focus of the second phase involved the construction and administration of the test booklets to collect the quasi-experimental data relevant to the research questions. The final phase focussed on the collection of qualitative data through surveying all the students and the purposeful selection and interviewing of students who had participated in the quasi-experimental data collections. Additionally, the researcher maintained his journal during the second and third phases of the research plan. Generally, entries were made following the administration of each of the test booklets and continuously while the interviews were being conducted.

RESEARCH INSTRUMENTS

Overview of the Research Instruments

A summary of the research instruments that were employed in the positivist design and the types of data they collected is shown in Table 3.2 (page 77). The relationship of the research instruments to the research questions is outlined in the following sections, together with background information on the development, use, validity and reliability of each of these instruments.

Self-Efficacy Scale

Relationship of the Instrument to the Research Questions

Self-efficacy was conceptualised in the previous chapter as students' perceptions of their capabilities to execute particular courses of action. Hence, the self-efficacy scale was developed as an item-specific measure of the strength of boys' and girls' perceptions of their capabilities for correctly answering individual achievement questions set in multiple-choice and constructed-response formats. Constructed-response and multiple-choice self-efficacies were operationalised as the

mean of each student's item-specific self-efficacies for each of these test formats when the self-efficacy scale is completed before each achievement question according to the direction, 'Rate your confidence that you can correctly answer this question.'

Table 3.2
Overview of the Research Instruments

Research instrument	Variables measured	Response-format
Self-efficacy scale	Strength of self-efficacy perceptions	5-point Likert-type scale
<i>Attitude Toward Tests Survey</i>	Extent of worrisome cognitions	4-point Likert-type scale
Achievement Test 1 (<i>Earth Science</i>)	Extent of multiple-choice achievement	Multiple-choice (4 choices)
	Extent of constructed-response achievement	Short-answer
Achievement Test 2 (<i>Organisms and Food</i>)	Extent of multiple-choice achievement	Multiple-choice (4 choices)
	Extent of constructed-response achievement	Short-answer
<i>Survey About Science Tests</i>	Perceptions about self-efficacy, worry, multiple-choice and constructed-response questions.	Open-ended questions

Description and Development of the Self-Efficacy Scale

Vispoel and Chen (1990) report from their review that the strength of self-efficacy perceptions is measured mainly by Likert-type confidence scales and certainty scales. Bandura's original operationalisation of self-efficacy (for example, Bandura, 1982) required students to rate their certainty that they could complete particular tasks on 11-point certainty scales ranging from 0 to 1.0. Although such scales have been commonly used with college and university students in the United States, mathematics teachers at the research site advised that some of the Year 9 students

would not understand the anchoring points on these types of scales. Consequently, a Likert-type confidence scale, anchored by verbal descriptors, was chosen to measure self-efficacy. This type of scale has been commonly used to measure item-specific confidences for a wide range of age groups (for example, Fullarton, 1993; Glenberg & Epstein, 1987; L.G. Jones & L.P. Jones, 1989; M. Lundeberg et al., 1994; Maki & S. Berry, 1984; Pressley, Levin, Ghatala & Ahmad, 1987). It must be noted, however, that these types of scales collect ordinal rather than interval data, and that the interpretation of these data assumes that each of the scale units represent equal-sized differences in the construct that is measured (Wiersma, 1969). However, it must also be recognised that almost all the self-efficacy research has employed Likert-type scales (Vispoel & Chen, 1990) and that there appeared to be no other suitable alternatives for measuring self-efficacy.

Most research that has measured item-specific confidences has utilised either 4-point (for example, Pressley et al., 1987), 5-point (for example, Fullarton, 1993; M. Lundeberg et al., 1994) or 6-point Likert-type scales (for example, Glenberg & Epstein, 1987) to maximise the validity and reliability of the inferences that can be drawn from the data (Gable & M. Wolf, 1993). The vast majority of these self-efficacy scales have very good internal consistencies (Cronbach's alpha) of above 0.80 (Vispoel & Chen, 1990). A 5-point Likert-type response scale was chosen for this study to maximise the variability among students' responses without burdening them with needlessly fine discriminations. It was thought by this researcher that students might answer the self-efficacy scale carelessly or that they might omit answering this scale during tests if rating their self-efficacy was too onerous. Although there is some debate about the use of scales that provide a neutral point by having an odd number of responses, Gable and M. Wolf (1993) point to the paucity of definitive answers to this question. Similarly, there are no definite answers as to whether the labelling of response categories should be end-defined or all category-defined (Gable & M. Wolf, 1993). This research utilised an end-defined scale, anchored by the words 'very little' and 'quite a lot', to measure the strength of students' self-efficacies for answering achievement questions.

Although item-specific confidences have been measured in the literature before and after the completion of cognitive tasks, the conceptualisation of self-efficacy

requires that self-efficacy judgements are made before tasks are attempted. Moreover, there is some limited evidence to suggest that students' item-specific confidences differ depending on whether they rate their confidence before or after completing cognitive tasks (Glenberg & Epstein, 1987). Consequently, a series of instructions was developed to ensure, as far as possible, that students knew how to rate their self-efficacy (see Appendix A).

Pilot study of the self-efficacy scale

Procedures for the Pilot Study

A pilot study was conducted to evaluate the construction, administration and scoring of the self-efficacy scale. One intact class of 31 Year 9 science students from the research site participated in the pilot study during their end-of-term 1 achievement test. The students in this class were representative of the population of students who would be participating in this study in terms of their ages, and their socio-economic and cultural backgrounds.

The self-efficacy scale was printed next to each multiple-choice question and each constructed-response question on the achievement test (see Appendix A), which was comprised of 17 multiple-choice and 9 constructed-response questions. A single self-efficacy scale was provided for each of the constructed-response questions because the individual parts of these questions were similar to each other.

The test was administered to the class in their scheduled science lesson by the researcher. The researcher instructed the students regarding how they should answer the self-efficacy scale by reading aloud the instructions that were written during the development phase while the students read their copy of the instructions. Additional emphasis was given to the instruction that students were to answer the self-efficacy scale before solving the question. An *ad hoc* statement, that students would have sufficient time to complete the test, was made because the researcher was concerned that the students might not complete the self-efficacy scale if they thought otherwise. No further instructions were given and no assistance was provided to the students while they answered the achievement test and the self-efficacy scales. Feedback about the use of the self-efficacy scale was sought from the students in the following science

lesson, two days later, by asking them to write their responses to a series of questions (see Appendix A) about the self-efficacy scale onto the back of their test papers after they had finished viewing their results.

The self-efficacy scales were hand-scored by assigning each of the responses a number so that high scores represented high levels of self-efficacy. For example, the response 'A' was scored as 1 and the response 'E' was scored as 5. Omitted or ambiguous items were assigned the mean self-efficacy rounded to the nearest 0.5, except that students who failed to make responses for two or more constructed-response questions or more than three multiple-choice questions were deleted from the reliability analyses. Single missing responses for 8 students and 2 missing responses for 1 student were prorated according to these procedures. Moreover, 7 responses were rounded to the nearest 0.5 for several students who used the self-efficacy scale as a continuum. Data for all 31 students were analysed.

Results From the Pilot Study

No overt negative reactions to the self-efficacy scale were observed from the students either during the test or in the following lesson. The students appeared to be willing participants, and several students were quite interested to compare estimates of their self-efficacy with their achievement.

Analysis of the data yielded a barely adequate Cronbach alpha of .67 for multiple-choice self-efficacy (17 items) and an acceptable value of .86 for the constructed-response self-efficacy (9 items) in comparison to .80, the benchmark that Gable and M. Wolf (1993) specify for affective instruments.

There was almost unanimous agreement among the students that they understood how to use the scale and approximately three quarters of the students indicated that the instructions were clear. It is important to note in this context that no students asked any questions about the self-efficacy scale during the test. However, some students were observed to answer the achievement questions before they rated their self-efficacy. Of the 27 students who responded to Question 4 in the debriefing (Appendix A, 'How often did you answer the confidence scale before you solved the question?'), 11 students indicated that they always answered the self-efficacy scale first and 12 students indicated that they followed this procedure most of

the time. Additionally, about half of the students reported that they occasionally forgot to answer the self-efficacy scale first and several students commented that the order of answering the self-efficacy scale made no difference to them. Students' responses to Question 5 (Appendix A, 'What did you mean about your confidence when you circled the letter 'D'?') indicated that many students misunderstood the instructions for answering the self-efficacy scale, in that they reported their confidence about the correctness of their answers rather than their capabilities for answering questions correctly.

Although a third of the class thought that no improvements to the administration of the self-efficacy scale were necessary, the remainder of the class suggested a variety of improvements. Of these improvements, three suggestions were noteworthy because they had a direct bearing on this research. It was apparent that some students confused confidence in their capabilities for answering questions with their predicted achievement. Eric, for example, equated a 'D' on the self-efficacy scale with an achievement grade of 'B.' It was in this sense that one fifth of the students suggested that the order of the letters should be reversed so that their self-efficacy ratings correspond to the letter grades by which their end-of-semester achievement is reported. Otherwise, several students suggested the use of numbers, rather than letters on the self-efficacy scale 'because people might think that it refers to their grade.' Finally, Luke commented that 'everyone doesn't judge the same.' This comment was supported by the variety in students' answers to Question 5 (Appendix A, 'What did you mean about your confidence when you circled the letter 'D'?'). Whereas some students interpreted 'D' to mean 'slightly confident' or 'a little bit of confidence,' other students interpreted it to mean 'just about sure' or 'fairly confident, but not sure.'

Refinements to the Self-Efficacy Scale

Students' reports from the pilot study suggested several modifications to the design and administration of the self-efficacy scales. First, the numbers 1 to 5 replaced the letters 'A' to 'E' to avoid confusion with the letter grades that are awarded for science achievement at the end of each semester. Second, reports from the students suggested that they needed some guidance to distinguish between

confidence in their abilities and confidence about the correctness of their answers. Hence, the instructions for answering the self-efficacy scale were modified to highlight the order for answering the self-efficacy scale and the achievement items. Finally, an example corresponding to a confidence rating of '4' was provided to counter students' uncertainties about the meaning of this position on the self-efficacy scale. The revised instructions for answering the self-efficacy scale, together with the revised self-efficacy scale are located in Appendix A.

Although the internal consistency of the multiple-choice self-efficacies was barely satisfactory, higher values could be expected for this research study when the achievement tests are administered to the larger target population of students. Nevertheless, it was decided to increase the number of the multiple-choice items on the achievement tests so that the internal consistency of the multiple-choice self-efficacy data would be enhanced.

The Attitude Toward Tests Survey

Relation of the Instrument to the Research Questions

The *Attitude Toward Tests Survey* was designed to measure the extent of the girls' and the boys' worrisome cognitions for answering particular sets of multiple-choice and constructed-response test items. Worry was conceptualised in the previous chapter as situation-specific cognitive concern about the likelihood and consequences of failure during tests. An examination of the test anxiety research literature uncovered several instruments consistent with this conceptualisation. However, the *Worry-Emotionality Questionnaire* (Liebert & Morris, 1967) and its revisions (for example; Morris et al., 1981) do not have the same rigorous history of instrument development and validation as does Spielberger's (1980) *Test Anxiety Inventory* (Conetta & Schick, 1983; DeVito, 1988). Moreover, the *Test Anxiety Inventory* has been employed to successfully distinguish between different types of test anxieties for the same group of students (Benson, 1989). Hence, worry was operationalised as the sum of students' scores on the worry items from Spielberger's *Test Anxiety Inventory* when students are directed to respond to these items according

to how they feel about the achievement questions they have just completed. As described in more detail later in this section, Spielberger's (1980) worry items were embedded in the *Attitude Towards Tests Survey*.

Description of the *Test Anxiety Inventory* (Worry Scale)

Spielberger's (1980) *Test Anxiety Inventory* is a self-report scale that measures the extent of specific symptoms of anxiety during tests. It consists of 20 items that require students to rate how frequently they experience particular symptoms of anxiety on 4-point Likert-type scales, which range from (1) almost never, through (2) sometimes and (3) often, to (4) almost always. The *Test Anxiety Inventory* provides a measure of omnibus test anxiety and separate measures for worry and emotionality. The worry and emotionality scales are each comprised of 8 items and the remaining 4 items load onto each of the factors that define the worry and the emotionality scales. The minimum score on each scale is 8 and the maximum score is 32, where high scores represent high levels of either worry or emotionality.

The *Test Anxiety Inventory* was developed to measure test anxiety in high school and college students. However, its language is also appropriate for junior high school-age students (DeVito, 1988) and it has been used successfully at this level (Spielberger, 1980). Spielberger (1980) presents normative data in his test manual for samples of students from the United States that show high levels of internal consistency (Cronbach's alpha) of at least 0.83 on the worry scale. Moreover, Spielberger (1980) reports median item-remainder correlations between 0.58 and 0.72 as an additional measure of the *Test Anxiety Inventory's* internal consistency.

Evidence for the validity of the *Test Anxiety Inventory* has been obtained from factor analytic studies and correlational analyses. Factor analytic studies from samples of college students in the United States support the conceptual distinction between the worry and the emotionality components of the *Test Anxiety Inventory*, and they also indicate that the factor structure is invariant across gender (Everson, Millsap & Rodriguez, 1991; Hedl, 1982; Spielberger, 1980; Ware, Galassi & Dew, 1990). In this context, Everson et al. (1991) report generally higher levels of test anxiety for females from their confirmatory factor analysis and gender differences for

item reliabilities and correlations between the factors. However, Ware et al. (1990) report from their confirmatory factor analysis that gender differences in correlations between the factors are not statistically significant. Correlations for each of the test anxiety scales with measures of academic achievement show that the worry scale is correlated more highly with academic achievement than the emotionality scale, which provides some evidence for the discriminant concurrent validity of the *Test Anxiety Inventory* (Hedl, 1982; Spielberger, 1980).

Development of the *Attitude Towards Tests Survey*

The worry items on the *Test Anxiety Inventory* (see Appendix B) are written such that high scores relate to high levels of worry. Because these items all correspond to negative thoughts about testing they could cause a type of a response bias whereby students respond in the same way to similar items that are next to each other (Gable & M. Wolf, 1993). Therefore the items from Spielberger's worry scale were embedded within a series of positive statements about testing to form a new instrument, the *Attitude Toward Tests Survey* (see Appendix B). These positive statements were adapted from a checklist of positive thoughts that were catalogued by Galassi et al. (1981) from a study of students' positive and negative thoughts during tests. Although embedding the worry items into the *Attitude Toward Tests Survey* was less economical with testing time, it avoided the difficulty of reworking Spielberger's items to reflect both positive and negative thoughts about testing. Moreover, it is important to note that the negation and reverse scoring of items is problematic because this type of alteration does not necessarily preserve an instrument's underlying factor structure (Gable & M. Wolf, 1993).

Pilot Study

Procedures for the Pilot Study

As indicated earlier, Spielberger's worry scale was initially validated in the context of the standard administration of the *Test Anxiety Inventory* to undergraduate and high school-aged students in the United States of America. No data are available

regarding its administration to high school students in Australia or for conditions other than its standard administration. Hence a pilot study of the *Attitude Towards Tests Survey* was conducted at the research site to determine the validity of the interpretations that could be drawn from the worry items.

The *Attitude Toward Tests Survey* was piloted with a total of 197 students from four intact Year 8 classes and three intact Year 10 classes at the research site. The students in these classes were representative of the socioeconomic and cultural backgrounds of the students who participated in the research and their ages were similar. None of the students who participated in this pilot study was involved in any other aspect of the research.

The *Attitude Toward Tests Survey* was administered to students in each of the participating classes at the same time after they completed the Australian Schools Science Competition (ASSC) organised by the Educational Testing Centre at the University of New South Wales. Questions in the ASSC test students' ability to reason in a range of scientific contexts. These tests were administered by the students' home class teachers after their 'form' period. Although the questions on the tests differed for the Year 8 and the Year 10 students, the tests were administered under the same conditions, and the format, number and types of questions were identical.

The *Attitude Toward Tests Survey* was distributed to the students with the test booklets and the answer sheets for the Australian Schools Science Competition. Students were instructed to complete the survey once they had finished answering the test. The supervising teachers were instructed to emphasise the instruction on the survey form, that students answer the *Attitude Toward Tests Survey* for the questions they had just completed so that the *Attitude Toward Tests Survey* would provide a measure of state worry. The supervising teachers were also instructed not to explain the meaning of any of the items on the *Attitude Towards Tests Survey* so that no additional sources of variation were introduced into the worry data.

A template was constructed to hand-score Spielberger's worry items from the *Attitude Toward Tests Survey*, according to the standard procedures specified in the test manual (Spielberger, 1980). The score for each worry item corresponded to the number that was circled. According to Spielberger's (1980) instructions, students

who failed to respond to two or more items on the worry scale (specifically, two Year 10 students and four Year 8 students, in this case) were deleted from the analyses so that the validity of the worry scale would not be compromised. Additionally, another student was deleted because he provided yes and no answers to the questions rather than circling the appropriate responses. Single missing responses for 26 students and 16 single ambiguous responses were prorated by determining the mean of the worry items that the students responded to, and then rounding this number to the nearest value.

Results From the Pilot Study

Hand examination of the *Attitude Towards Tests Survey* indicated no particular patterns of distribution for the ambiguous responses. However, item 11, 'I seem to defeat myself while working on important tests' yielded 10 of the 26 omitted responses. A large majority of these omissions were made by Year 8 students. The absence of particular types of repeating response patterns on the survey forms provided some evidence that the pilot study was treated seriously by the participating students.

Data for all the 191 students were analysed using the Statistical Package for the Social Sciences. Means, standard deviations and item to rest-of-scale correlations were calculated for each of the worry items (see Appendix B, Table B.1). The mean scores for these items ranged from 2.08 for item 9 to 1.59 for item 5. Consequently, none of the items needed to be deleted because their means approached the extremities of the response scale. Additionally, all the worry items demonstrated sufficient discriminatory power (Gable & M. Wolf, 1993), yielding moderate to large standard deviations between 0.72 for item 11 and 1.00 for item 13. Additionally, the item to rest-of-scale correlations were all above the threshold of .30 specified by Gable and M. Wolf (1993), ranging from .36 for item 1 to .60 for item 9. The item analyses were repeated separately for the boys and the girls (see Appendix B, Table B.2) and all the items were judged to be sufficiently discriminating and correlated with the rest of the scale for both the girls and the boys. The internal consistency of the worry scale (Cronbach's alpha) was calculated at .78 (.76 for the boys and .81 for the girls), which indicates that the girls' and the boys' responses on worry scale were

sufficiently reliable.

A principal components analysis was conducted on the data for all 191 cases (see Appendix B, Tables B.3 & B.4). The communality of all the items was set to 1.00 and a single factor with an eigenvalue greater than or equal to one, that is, Kaiser's criterion (Ferguson & Takane, 1989) was extracted. All the worry items had substantial loadings, ranging from .50 for item 1 to .73 for item 9 onto this factor, which provides evidence for the unidimensionality of the items. Hence it was considered that the items on the worry scale could be used to discriminate between similar groups of students according to the extent of their worrisome cognitions during tests.

The data from the pilot study were then subjected to further principal components analyses (see Appendix B, Tables B.5 - B.7) to establish if the factor structure of the worry scale is invariant across gender. The numbers of the boys and the girls in these analyses exceeded the minimum value of cases to item ratio of 10:1 recommended by Gable and M. Wolf (1993). The principal components analysis for the girls yielded a single factor solution. Loadings onto this factor ranged from .44 for item 1 to .75 for factor 9, indicating that the items can be used as a single scale for distinguishing between girls who are high in worry and low in worry. However, two factors were extracted for the boys. Varimax rotation of the axes, where all factors are normalised before rotation, all vectors are of unit length and all variables receive the same weighting, that is, Kaiser's normalisation (Ferguson & Takane, 1989) supported the two factor solution. Items 5, 7, 9 and 13 appeared to load onto one factor and items 1, 3, 11 and 15 onto the other. However, another indicator of the unidimensionality of a scale is to compare the magnitude of the first eigenvalue to the second eigenvalue. For the boys, this ratio was 2.66 and for the girls it was 3.60. Although the scale appeared to function differently for the girls and the boys, the ratios of the first and second eigenvalues suggested that the scale could be regarded as unidimensional, although much less so for boys than for girls.

Refinements to the *Attitude Towards Tests Survey*

The item analyses indicated that item 1 had the lowest item to rest-of-scale correlations. A minor change to the wording of this item was made to make it more

applicable to the students' context. The reworked item substituted the word 'subject' for 'course' so that it read 'Thinking about my marks in a subject interferes with my work on tests.' Additionally, the third item was made more relevant to the students' context by changing the wording to 'my mind freezes up on important tests.' The 11th item, 'I seem to defeat myself while working on important tests' was retained unaltered, despite the number of omitted responses because almost all the omitted responses were attributable to the Year 8 students who were younger than target population. The revised version of the *Attitude Towards Tests Survey* is located in Appendix B. Because changes were made to the worry items, an additional principal components analysis was planned for the worry scores from the first of the quasi-experimental data collections. (The results for these analyses are reported in Chapter 4).

Achievement Tests

Relationship of the Achievement Tests to the Research Questions

End-of-unit achievement tests were developed for two consecutive Year 9 Science topics, namely, *Earth Science* and *Organisms and Food*. These achievement tests provided measures of girls' and boys' achievement on multiple-choice and constructed-response items, and they also provided the items that students responded to when they reported the strength of their self-efficacies and the extent of their worrisome cognitions. Multiple-choice questions were operationalised as those questions which required students to select the correct response to a statement from three distractors. Constructed-response questions were operationalised as those questions which required students to generate a written response of at the most two sentences. Multiple-choice achievement and constructed-response achievement were operationalised as the sum of each student's scores for each of the item-response formats on each of the achievement tests.

Development of the Achievement Tests

The development of the achievement tests was guided by tables of specifications (Gronland, 1993) for each of the tests. Content outcomes (that is, level 1 of Bloom's (1956) taxonomy of educational objectives) and process outcomes (that is, levels 2 to 6 of Bloom's (1956) taxonomy of educational objectives) were equally weighted. The content of the tests was weighted according to the relative amount of time that was spent on each of the content areas during instruction. Additionally, questions on the unit tests sampled the learning objectives for the content and the process outcome according to their weighting in the table of specifications. The choice of response format for the achievement questions was dictated by the research questions and the research design. The experience of the researcher and the cooperating teachers at the research site suggested that the maximum number of questions that could be completed by all the students during the programmed time of 48 minutes was 20 multiple-choice questions, each worth 1 mark, and 10 constructed-response questions each worth 2 marks. A greater or lesser number of test items could compromise the reliability of the data because reliability increases with the number of test items and it starts to decrease when students do not have sufficient time to complete questions in the allocated time (Mehrens & Lehmann, 1984).

Achievement questions were obtained from books 2 and 3 of the *Science Item Bank* (Australian Council for Educational Research, 1978) and a collection of end-of-unit test papers at the research site that had been developed and administered over the previous years. The questions were modified, where it was possible, to cast them into multiple-choice and constructed-response formats so that each constructed-response question was paralleled by two equivalent multiple-choice questions. It is important to note in this context that these parallel items tested the same content or process outcomes, that they were couched in the same contexts and that they referred to the same figural information. Questions that could not be adequately cast into both response formats were deleted. Additionally, some achievement questions needed to be written where no suitable questions were available. The quality of the test questions for each of the response formats was then compared with checklists provided by standard test construction manuals (Adkins, 1974; Gronland, 1993) and

modifications were made in line with their suggestions. The correct responses to the multiple-choice questions were randomly distributed among the distractors according to the method described by Gronland (1993). This process involved randomly opening a book to a page, looking at the final digit of the page number on the right-hand side of the book and using that digit to determine the location of the correct answer among the distractors. Finally, the achievement items were arranged to produce two equivalent tests for each unit topic, and a series of instructions for answering the achievement tests was devised.

The content validity of the achievement tests was enhanced by utilising tables of specifications and by seeking advice from the cooperating teachers. The cooperating teachers were given the first draft of each test paper and they were asked to comment on the content, difficulty and equivalence of the tests, as well as on the appropriateness of the language. The cooperating teachers wrote their comments and modifications onto their draft copies and the achievement tests took into account their suggestions. The achievement tests were not piloted and no formal tests of readability were completed. However, it is important to note that a substantial portion of the questions on the achievement tests were developed in previous years at the research site and modified with successive administrations, and as noted earlier, already validated items from the *Science Item Bank* (Australian Council for Educational Research, 1978) were also used.

A marking schedule was constructed for each test using the science textbook as a guide. Each multiple-choice question was scored either correct (1 mark) or incorrect (0 marks). Each constructed-response question, which corresponded to two multiple-choice questions was allocated 2 marks. Partially correct constructed-response answers were rewarded with half marks at the request of the cooperating teachers. Thus the scores for constructed-response questions ranged from 0 through $\frac{1}{2}$, 1 and $1\frac{1}{2}$ to 2 marks. The cooperating teachers commented on the correctness and the appropriateness of the marking schedule and modifications were negotiated in response to their suggestions. The relevant marking schedules are included with copies of test booklets 1A and 2A in Appendices C (*Earth Science* test) and D (*Organisms and Food* test).

Description of the Achievement Tests

Two achievement tests, Papers 1 and 2, containing equivalent multiple-choice and constructed-response items were prepared for each of the unit topics (see Appendices C & D). Each test contained 20 multiple-choice questions and 10 constructed-response questions. The multiple-choice items on one test were equivalent to the constructed-response items on the other, and vice-versa. An example of equivalent achievement items from the *Earth Science* test is shown below. Question 5 from paper 1 is equivalent to questions 9 and 10 from paper 2.

Question 5 (Paper 1)

Describe the way cold air moves in a cold front and the conditions that are required for the formation of rain clouds.

Question 9 (Paper 2)

When a cold front passes, the cold air moves ...

- A. past the warmer air.
- B. underneath the warmer air.
- C. through the warmer air.
- D. over the top of the warmer air.

Question 10 (Paper 2)

A cold front produces rain clouds when the ...

- A. cold air contains water vapour.
- B. warm air is less dense than the cold air.
- C. relative humidity of the warm air is high.
- D. cold air is moving faster than the warm air.

The achievement items were ordered in the test booklets, first according to their response formats, and second according to types of outcomes. The items measuring content outcomes preceded the items measuring process outcomes. A space was provided beneath the constructed-response items for the students to make their responses and it was intended that students would answer the multiple-choice items by circling the letter corresponding to their choice.

The Survey About Science Tests

Relationship to the Research Questions

The *Survey About Science Tests* was designed to collect qualitative data about the content and functioning of the boys' and the girls' generalised self-efficacies and their worrisome cognitions during tests. This type of data was required to test for the presence of gender-related differences in students' format-related self-efficacies and their format-related worrisome cognitions, and also to identify individual students for the interviews.

Development of the Survey About Science Tests

The first step involved the derivation of survey questions from the research questions. Other questions relating to demographic and school-related variables, which could help interpret the research data, were included at this stage. The initial questions were reviewed by another researcher who also works in the field of gender and assessment. Following the advice of this researcher, the original list of 26 questions was reduced to a list of 12 questions to make the survey less onerous to complete. Additionally, the language framing the questions was simplified to make it more appropriate for junior high school students.

Open-ended written response formats and tabular response formats were chosen to avoid imposing a series of *a priori* categories onto the respondents (Hitchcock & Hughes, 1989). Sufficient space was provided for each question so that students could write several responses into the tables or to allow for short explanations, according to the requirements of each question. Moreover, space was left at the end of the survey form for students to furnish additional comments about science tests. Finally, a series of instructions was developed for the pilot survey, which requested students to answer the questions as completely as they could. These instructions also assured students that there were no right or wrong answers to the questions and that the confidentiality of their responses would be respected. Additionally, space was provided on the front of the *Survey About Science Tests* for students to write their

name, sex, class and age.

Pilot Study of the *Survey About Science Tests*

Procedures for the Pilot Study

The purpose of the pilot study was to increase the validity and reliability of the *Survey About Science Tests*. The pilot draft of the *Survey About Science Tests* (see Appendix E) was administered by the researcher to his Year 10 class at the research site during one of their time-tabled religious education lessons. The purpose of the survey was explained to the class and the students were instructed to answer the survey questions according to the directions on the front cover. Additionally, the students were instructed to write their responses at the end of the survey form to the following questions, which were written onto the whiteboard;

Which questions...

- didn't make sense?
- need to be reworded?
- couldn't be answered (why)?

The students were allowed to continue with their programmed work when they finished answering the survey form, and the survey forms were collected when it appeared that all the students had finished answering them. None of the students who participated in this pilot study was involved in any other aspect of the research.

Results From the Pilot Study and Changes to the *Survey About Science Tests*

Almost all the students completed the survey within fifteen minutes with some students finishing in less than five minutes. It appeared that most students treated the survey seriously although several of the girls wrote that they thought that the questions were pointless. Most students wrote only one response to each of the questions that required them to respond in the tabular format, and many of their answers to the other questions were briefer than anticipated. Hence, the survey form was modified by placing numbers into each of the tables to indicate that students could make multiple responses, and the instruction on the front cover, 'as completely as you can', was underlined. Additionally, there was a disproportionately high

number of omitted responses for Question 1. The students mentioned in their written reports for this question that they ‘can’t remember any of this’ or that they ‘just do the test,’ and there was no evidence that any of the students failed to understand this question. Nevertheless, Question 1 was retained unchanged because a large majority of the students still provided responses to this question.

Additionally, the students made positive comments about the wording of the questions. However, several students commented that the survey required too much writing and that it was a ‘tad too long.’ Moreover, one fifth of the students reported that some of the questions were repetitious. Consequently, Questions 4 and 8 were deleted from the survey. Although several students thought that Question 2 was repetitious, this question was retained because it is conceptually different from the other questions on the survey. The revised version of the *Survey About Science Tests* is located in Appendix E.

PROCEDURES FOR THE COLLECTION OF THE DATA

General Aspects of the Data Collections

A letter seeking formal permission to conduct the research was written to the Principal of the school at the start of the fourth school term in 1995, although verbal permission had been granted in the previous year. The Principal was provided with a copy of the research proposal and his attention was drawn to the ethical guidelines underpinning the research. Permission was granted to conduct the research, and the Principal informed the Head of the Secondary Department and the Head of the Science Department of his decision. Consequently, the researcher was allocated a Year 9 Science class in 1996 according to his requests.

The research followed a series of ethical guidelines that protected students and ensured that the findings are as non-biased and as accurate as possible. A ‘utilitarian’ (Miles & Huberman, 1994, p.289) stance was taken toward the recruitment of participants, field work and reporting. Permission to conduct the research was sought from the Principal of the school because the research required personal data from the students that could be sensitive to them. Once the Principal had given his ‘informed

consent' (Miles & Huberman, 1994), a letter was written to the parents of the Year 9 students advising them of the research. Parents were invited, in this letter, to discuss any of their concerns with the researcher. Additionally, the researcher spoke to each Year 9 class to inform them about the research in general terms and to answer their questions. Students' participation in the interviews was voluntary, and their right not to answer questions was respected. Furthermore, the data collections proceeded on the understanding that all research data would be kept confidential and that pseudonyms would be used to conceal the identity of the students and the school. Additionally, because the quantitative data collections required minor modifications to the usual testing procedures, the achievement tests were developed in consultation with the cooperating teachers and Head of the Science department.

A summary of the data collections is shown below in Table 3.3, which indicates the timing of the data collections, the stage of the research and the types of data that were collected.

Table 3.3:
Overview of the Data Collections

Timeline	Data Collected
June 1996 (First Collection of the Quasi-Experimental Data)	Achievement, item-specific self-efficacy and worry for multiple-choice and constructed-response questions.
September 1996 (Second Collection of the Quasi-Experimental Data)	Achievement, item-specific self-efficacy and worry for multiple-choice and constructed-response questions.
October 1996 (Collection of the Survey Data)	Perceptions about achievement, multiple-choice and constructed-response questions, generalised self-efficacy and worrisome cognitions.
October-December 1996 (Collection of the Interview Data)	Meaning-perspectives about multiple-choice and constructed-response questions, generalised self-efficacy and worry

The Quasi-Experimental Data Collections

Quantitative data relating to the extent of boys' and girls' achievement, worry and item-specific perceptions of self-efficacy were collected from all Year 9 students at the research site on two separate occasions. The five classes of students were divided into four research groups. Classes 9A, 9B, 9D and 9E formed Research Groups 1, 2, 3 and 4 respectively. Students from the 9C class, who were taught by the researcher, were allocated among the research groups to ensure that the groups were balanced for the numbers of girls and boys of the same ability, as measured by their prior science achievement. It is also important to note in this context that overall, there were no practically important differences in mean intelligence as measured by the *ACER Intermediate Test G* (Australian Council for Educational Research, 1982) between the boys and the girls for whom these data were available (see Appendix F, Table F.1). On the other hand, small ($d = -0.23$) gender differences in mean intelligence quotient scores were detected favouring the girls for the combined Research Groups 1 and 2 whereas gender differences in mean intelligence quotient scores for Research Groups 3 and 4 were negligible ($d = +0.13$). However, these data were available for only 129 of the 154 students who were involved in this research.

A total of 147 Year 9 students (82 girls and 65 boys) were involved in the first of the quasi-experimental data collections and 147 Year 9 students (66 boys and 81 girls) were involved in the second of the quasi-experimental data collections. Two students were absent from their science classes for the first test and five students were absent for the second test. Additionally, three students joined the Year 9 Science classes at the school between the times of the quasi-experimental data collections.

Four other teachers, as well as the researcher, were involved in the quasi-experimental data collections on each occasion. Because of sicknesses and commitments to other school activities, only two of the cooperating science teachers supervised the first of the quantitative data collections, and only one of these teachers supervised the second of the quantitative data collections. Nevertheless, the supervising teachers were trained in the administration of the research instruments by the researcher, who spoke individually to each teacher. The attention of the

supervising teachers was drawn to the instructions printed onto the front of the test paper, and how the self-efficacy scale and the *Attitude Towards Tests Survey* were to be answered. The instructions that the students must complete the self-efficacy scale before answering the achievement questions, and that the achievement questions must be answered in the order of presentation were emphasised to the supervising teachers. Additionally, the researcher had visited each of the classes several days before the first of the quasi-experimental data collections to reassure the students that they would have sufficient time to complete all the tasks in the test booklets.

The quasi-experimental data collections took place at the end of the second and third school terms. All of the classes completed the test booklets at a common time under standard test conditions. Students were expected to do their work without talking or reference to their books. Research Groups 1, 2, 3 and 4 were allocated test booklets 1A, 1B, 2A and 2B respectively in the first of the quasi-experimental data collections, and test booklets 2B, 2A, 1B and 1A in the second of the quasi-experimental data collections. A series of instructions were printed on the front of each test booklet for answering the test booklets, and the modified instructions regarding how students should answer the self-efficacy scale (see Appendix A) were distributed to each student. The supervising teachers were requested to read to their classes the instructions for the self-efficacy scale and the instructions on the front cover of the test booklet. Additionally, the instruction that the *Attitude Towards Tests Survey* was to be answered for the questions that students had just finished answering was highlighted. Classes were allowed 45 minutes to complete the test booklets. The supervising teachers were instructed to answer procedural questions about the research instruments but to refrain from answering questions relating to the content of the achievement tests.

The first of the quantitative data collections proceeded as planned for all the classes except that the teacher supervising the 9A students, who were all in Research Group 1, was approximately ten minutes late to class. However, inspection of the test scripts revealed that all the students finished the test. The second of the quasi-experimental data collections proceeded according to plan except that the teacher supervising the 9A class reported that some of the students turned to the centre of their test booklets so that they could complete the multiple-choice questions first.

However, this teacher redirected the students to answer the questions in the order that they were presented.

The scoring of the self-efficacy scales and the *Attitude Towards Tests Surveys* were completed by the researcher for all the classes according to the procedures described earlier in this chapter. Additionally, the researcher checked that the self-efficacy scales and the *Attitude Towards Tests Surveys* were correctly scored for all the papers after completing the initial scoring of these instruments. The achievement questions for each of the tests were marked according to the marking schemes agreed to by all the teachers. Each of the cooperating science teachers hand-scored the multiple-choice achievement questions for their classes for the first of the quasi-experimental data collections, and the researcher hand-scored the multiple-choice questions for the second of the quasi-experimental data collections. The scoring of the multiple-choice achievement questions was checked in each case by sorting the test scripts into their class groups and then into alphabetical order. Six scripts were selected from each class group, by counting from the top of each pile the scripts corresponding to a series of numbers drawn randomly from a box. The multiple-choice questions on these papers were then re-scored. In each case there was 100% agreement with the initial scoring of these papers. The cooperating teachers scored the constructed-response questions for each test according to the marking schedules. However, because discrepancies were noted between the cooperating teachers in the way they scored these questions, the researcher re-scored all the constructed-response achievement questions analytically, question-by-question.

Collection of the Survey Data

The *Survey About Science Tests* was administered by the researcher and each of the cooperating science teachers to their Year 9 science classes at the start of the fourth school term, immediately after the students finished viewing their scored copy of the *Organisms and Food* test paper. The cooperating teachers were instructed to read the instructions on the front of the *Survey About Science Tests* to their classes and then to allow 15 minutes for the students to complete the survey. The teachers were also instructed that they could clarify procedural matters relating to the

administration of the survey to students, but that they could not explain the meaning of any of the words or the questions on the survey. Students were not permitted to talk while they completed the survey. A total of 145 students (81 girls and 64 boys) from the five Year 9 classes that participated in the quasi-experimental data collections completed the *Survey About Science Tests*, and nine students were absent from their science classes on that day. Procedures for coding and aggregating the qualitative data from the *Survey About Science Tests* are described with the analyses of these data in the next chapter.

The Interviews

General Aspects of the Interviews

Interviews were conducted by the researcher with 22 students (10 boys and 12 girls) to determine the meaning-perspectives they held about multiple-choice and constructed-response questions, their generalised self-efficacies and their worrisome thoughts during tests. For reasons explained later in this chapter, these students were purposely selected so that, as a group, they displayed the full range of variation among the students with respect to the extent and influence of their self-efficacies. Hence, girls and boys with high, low and intermediate self-efficacies, as identified by the quasi-experimental data collections, were selected from groups of students who reported on the *Survey About Science Tests* the full variety of perspectives about the influence (or lack of influence) of their positive and negative thoughts during tests. These interviews were guided by perspectives described within standard works on qualitative research methods (for example, F. Erickson, 1986; Hitchcock & Hughes, 1995; Merriam, 1988; Miles & Huberman, 1994; Taylor & Bogdan, 1984). Twelve students (6 boys and 6 girls) participated in the first interviewing cycle. Nine of these students were re-interviewed together with an additional ten students in the second interviewing cycle. These interviews were conducted over a 6-week period after the preliminary analyses of the test booklets and the *Survey About Science Tests* were completed. It is possible that the students who were interviewed may have discussed the content of these interviews with their friends who were interviewed later, thereby

influencing the responses of these students.

Contact was made with individual students several days before the interviews to invite their participation and to negotiate interview times. Only one student declined to be interviewed. Most of these students were interviewed during their lunchbreak by the researcher. However, six of the students were withdrawn from their English lessons with the permission of their English teacher, and six of the follow-up interviews were conducted during the students' social education lessons, with the permission of their social education teacher. All the interviews were conducted in a science laboratory where there was minimal disruption from staff and other students. The interviews ranged in length from approximately 15 to 20 minutes and they were audio-taped, with the students' permission. Transcripts of the interviews were produced by the researcher no more than two days after each interview. As mentioned earlier, and described in more detail later in this chapter, the researcher maintained a journal during the interviews.

The questions during the first and the second interviewing cycles centred on four areas, namely; the research techniques associated with the test booklets, test formats, self-efficacy and worrisome thoughts. The first interviewing cycle was exploratory in its nature. The preliminary interviewing guide, which was used in these interviews, consisted of a dynamic series of questions that were shaped by the research questions, concerns about the rigour of the quasi-experimental data collections and insights from the ongoing analyses of the interviews. The second interviewing cycle was guided by the revised interviewing guide (see Appendix G), which accommodated revisions suggested by the preliminary analyses of the data from the first interviewing cycle. The general sequence of questions in the interviewing guide was maintained throughout the interviews, although the wording of the questions varied so that the interviews represented 'conversation[s] with a purpose' (Merriam, 1988, p.71) rather than orally administered surveys. Moreover, other questions arose in the interviews in response to students' answers to previous questions. However, none of the interviews from the second interviewing cycle were formally analysed until the data collections were complete, due to time constraints imposed by the researcher's teaching role. Nevertheless, nine of the students who participated in the first interviewing cycle were re-interviewed at the end of the

second interviewing cycle about insights that emerged during the first interviewing cycle.

Procedures for Enhancing the Quality of the Interview Data

F. Erickson (1986) describes five criteria that must be addressed to ensure researcher bias is minimised and that evidentiary warrants exist for inferences drawn from qualitative data, namely; (i) the amount of evidence, (ii) variety in the kinds of evidence, (iii) the interpretive status of evidence, (iv) the search for disconfirming evidence, and (v) the analysis of discrepant cases. The first four of these criteria are discussed in this chapter and the final criterion in Chapter 5.

F. Erickson's first criterion, which relates to the amount of evidence, was addressed by continuing the interviewing process until the categories were saturated (Lincoln & Guba, 1985), that is, until no significant new insights were added by successive interviews. In total, 22 students were interviewed. This number of students was more than sufficient considering the paucity of the new insights provided by the last half dozen students.

F. Erickson's second criterion, which relates to variety in the kinds of evidence, was satisfied by the use of multiple research methods. The employment of multiple research methods, that is, methodological triangulation, involved the collection of documentary data from the *Survey About Science Tests* and the test booklets, as well as data from the interview transcripts. Moreover, interview accounts were 'cross checked' (Taylor & Bogdan, 1984) with students, throughout the interviews, against the documentary evidence to assist with the formulation of plausible explanatory frameworks (Mathison, 1988). However, the fine-grained analyses were derived almost entirely from the interviews because these interviews subsumed data from the *Survey About Science Tests*. One limitation of this study, insofar as interpretive research designs are concerned, is the failure to collect data through participant observation. This researcher judged that observations of students working during tests were unlikely to yield fruitful data for the purposes of this study.

F. Erickson's third criterion, which relates to the interpretive status of the evidence, was satisfied partly by the researcher's prolonged engagement at the

research site, where he was employed as a permanent member of the teaching staff. Data were collected from the students over a period of 7 months. In the first interviewing cycle, sufficient time was available between the interviews for the preliminary analyses and refocussing the questions in preparation for the next interview. Additionally, a period of intensive analysis and refocussing took place at the conclusion of the first cycle, but before the commencement of the second interviewing cycle. However, as described previously, none of the interviews from the second interviewing cycle was formally analysed until the end of the interviews.

F. Erickson's third criterion was also addressed by ethical procedures to preserve the best interests of the students and ensure that the findings are as non-biased as possible. Students were invited to participate in the interviews, which were entirely voluntary and without reward. They were reminded at the start of the interviews about the purpose of the interview (namely, to seek their views about testing), and they were assured of their anonymity and the confidentiality of their responses. Moreover, students were not forced to answer questions they could not answer and the researcher adopted a non-judgemental stance to encourage their honest responses. There were no reasons to suggest that any of the students were unwilling or unreliable informants. Additionally, the veracity of the researcher's constructions were checked while he was in the field. 'Member checks' (Guba & Lincoln, 1989, p.238), during the interviews, required students to comment on the researcher's interpretations and the plausibility of emerging trends. Further, discussions were held with one of the cooperating teachers to check the plausibility of the researcher's interpretations and for the existence of alternative frameworks. Finally, the researcher made entries into a journal, which is described in detail later in this chapter, that enabled him to monitor his 'progressive subjectivity' (Guba & Lincoln, 1989, p.238) during the collection and analysis of the interview data.

F. Erickson's fourth criterion, which relates to the adequacy of the disconfirming evidence, was satisfied by the purposeful selection of students for the interviews to achieve maximum variation among the students regarding their conceptual importance to this study. Therefore, the selection of these students was not dictated by concerns about whether they were representative of the students at the research site. The procedures for the selection of these students were described

earlier in this chapter.

Journal

The researcher also maintained a journal while the data were collected. This journal served several purposes. First, it provided a description of the research context and a log of the procedures that were followed for the data collections. It also included the researcher's commentary about the implementation of the data collections, together with anecdotal reports from the cooperating teachers when they were provided. Generally, these types of entries were made after the administration of each test booklet and the *Survey About Science Tests*. Records from this journal were consulted to describe, in this present chapter, the research context and the procedures for collecting the research data.

Second, this journal was used, in accordance with F. Erickson's (1986) third criterion, to enhance the interpretive status of the data. These types of entries recorded the researcher's insights from the preliminary analyses of the interviews, at the end of each interview from the first cycle and also at the end of the first interviewing cycle. In one sense the journal was used in the first interviewing cycle as an aid toward the development of the final interviewing guide, and it chronicled the development of the new interview questions and the deletion of earlier questions. Additionally, the journal was kept to monitor the researcher's progressive subjectivity through records of his *a priori* constructions and his emerging findings. These latter records were kept to ensure that this researcher did not afford excess privilege to his biases during the final analyses of the interview data and they were consulted when the interpretive commentary framing each of the assertions was written in Chapter 5.

SUMMARY

This chapter has built on the theoretical and methodological base that was provided in the previous chapter by describing the methodology that was employed to develop a better understanding of girls' and boys' patterns of achievement according to test item-response formats. The research design involved the use of positivist and

interpretivist-constructivist methodologies side-by-side to provide a binocular view of this issue. The research instruments and the data collection procedures were also described with reference to the research context. The next two chapters present the results — Chapter 4 details the coarse-grained analyses for all the boys and girls involved in the positivist research design, and Chapter 5 details the fine-grained analyses for the students who were interviewed according to the interpretivist-constructivist research design.

CHAPTER 4

ANALYSIS OF THE DATA ... I

OVERVIEW

This chapter presents the ‘coarse-grained’ analyses of the positivist data that were collected for the 154 girls and boys who participated in this study. The coarse-grained analyses aggregate and compare data for all these boys and girls to identify gender-related differences that may help to better understand their achievements on multiple-choice and constructed-response test formats.

The first section of this chapter provides a summary of the procedures that were used to analyse the data and details on the functioning of the research instruments; the second section presents analyses relating to the content and extent of girls’ and boys’ perceptions of self-efficacy and their worrisome cognitions during tests; and the final section of this chapter compares boys’ and girls’ achievements on multiple-choice and constructed-response formats, and it tests the conceptualisations that underpin this study.

CONTEXT FOR THE ANALYSES

Analytical Procedures

Participants in the Coarse-Grained Data Analyses

Quantitative data from the test booklets and qualitative data from the *Survey About Science Tests* were analysed to inform answers to each of the research questions that were developed in Chapter 2, for the 154 girls and boys who participated in this study. Table 4.1 (page 106) shows the numbers of students from each of the research groups who were included in the data analyses.

Table 4.1
Numbers of Students Included in the Data Analyses

Students	Test 1	Test 2	Survey	Interviews
	<u>n</u>	<u>n</u>	<u>n</u>	<u>n</u>
Research Group 1				
Boys	15 (1)	12 (5)	16	2
Girls	18 (1)	17	20	2
Research Group 2				
Boys	14 (2)	17	15	2
Girls	22	20 (2)	20	6
Research Group 3				
Boys	11 (4)	11 (5)	17	5
Girls	18 (2)	18 (2)	20	1
Research Group 4				
Boys	14 (2)	13 (3)	15	1
Girls	18 (3)	17 (2)	22	3
Total Students				
Boys	54 (9)	53 (13)	63	10
Girls	76 (6)	72 (6)	82	12

Note. Numbers in parentheses show the numbers of students deleted from the analyses because they provided either spurious or incomplete data. 154 Year 9 students (73 boys and 81 girls) were involved in at least one of the data collections.

Procedures for Analysing Data From the Test Booklets

Procedures for scoring the self-efficacy scale, the worry items, and the multiple-choice and constructed-response achievement questions from the test booklets were described in the previous chapter. Additionally, responses on the self-efficacy scales and the *Attitude Towards Tests Survey* were scanned for particular types of repeating patterns to identify any students who might have answered these items spuriously. Quantitative data from the students who were suspected to have answered items spuriously were deleted from the analyses, together with data from the students who failed to meet the scoring criteria described in Chapter 3. The numbers of students

who were deleted from the analyses are also shown in Table 4.1.

The quantitative data were aggregated for girls and boys within Research Groups 1 and 2, as were the data for the boys and girls in Research Groups 3 and 4 to counterbalance the order that the multiple-choice and constructed-response items were answered. Tests to determine the statistical significance of gender-related differences were not conducted because they were deemed to be inappropriate for this study. It is important to note that tests of statistical significance are only employed when inferences are made about populations from representative samples (Ferguson & Takane, 1989; W. Jackson, 1995). In this context, the students who participated in this study were not representative of the wider population of Year 9 science students, and the differences between the girls and the boys were absolute differences (W. Jackson, 1995) because all the Year 9 science students at the research site were involved in this study. Therefore, average effect sizes were calculated across the research groups to measure the extent of these absolute differences and to decide whether these differences are likely to be practically significant. This procedure was justified because each of the research groups contained similar numbers of boys as they also did for the numbers of girls. Guidelines for the interpretation of effect sizes were provided in Chapter 2.

Procedures for Analysing Data From the *Survey About Science Tests*

Students' responses to questions on the *Survey About Science Tests* were broken into individual 'chunks' (Miles & Huberman, 1994, p.56), each consisting of single units of information. These units of data were analysed at the phrase level, question-by-question, according to procedures described in standard research methodology textbooks (Hitchcock & Hughes, 1995; Merriam, 1988; Miles & Huberman, 1994; Taylor & Bogdan, 1984).

According to these procedures, the units of data were grouped according to their similarity with each other, and a descriptive code was written for each of these groups. Units of data that could not be grouped with each other were retained, but not coded. Groups of data that appeared to be related to each other were combined into categories and a code-name was derived for each category. These categories,

which reflected the purposes of the research, were developed exhaustively so that they were mutually exclusive of each other and independent of each other (Merriam, 1988, p.136). It is important to note in this context that these groups and categories were not defined *a priori* by the researcher. Instead, they were allowed to emerge from the data analyses. Although these response categories were not identical to the theoretically derived categories for the data from the test booklets, it was possible, as shown later in this chapter, to develop links between these categories and the research questions.

The proportions of girls and boys responding to each question on the *Survey About Science Tests* were calculated to assist in the reporting and interpretation of students' responses. Moreover, the average numbers of responses for the boys and girls were determined for questions where they could make more than one response. The grouping and categorisation of all the units of data were check-coded (Miles & Huberman, 1994, p.64) by the researcher against operational definitions for each of the groups and the categories, six weeks after the data were coded. The reliability of the initial coding was calculated according to the formula described by Miles and Huberman (1994, p.64), where

$$\text{Reliability} = \frac{\text{number of agreements}}{\text{number of agreements} + \text{number of disagreements}}$$

The percentages of girls and boys responding to each of the categories were then compared to identify the presence of gender-related differences that may help to better understand their responses to tests and to test items that require different item-response formats.

The Quality of the Research Data

Internal Consistency of the Quasi-Experimental Data

The internal consistency (Cronbach's alpha) of the self-efficacy data were very high, ranging from .88 to .95 for multiple-choice self-efficacy and from .85 to .93 for constructed-response self-efficacy (see Table 4.2). For the worry data, the internal consistencies ranged from .81 to .87 for multiple-choice worry, and from .77 to .89 for constructed-response worry (see Table 4.2). Given the benchmark of .80 that

Gable and M. Wolf (1993) prescribe for affective instruments, data from both the self-efficacy scales and the worry scales appear to be highly reliable for all groups of students on each of the test booklets. Hence, the self-efficacy and worry data were judged to have sufficient internal consistency for analytical purposes.

The internal consistencies of the achievement scores were considerably less than the internal consistencies for the self-efficacy and worry data (see Table 4.2).

Table 4.2

Internal Consistency (Cronbach Alpha) of the Quasi-Experimental Data

Group	n	Self-Efficacy		Worry ^a		Achievement	
		MC	CR	MC	CR	MC	CR
<i>Earth Science test</i>							
1	33	.95	.93	.81	-	.62	.81
2	36	.94	.87	-	.77	.56	.81
1 & 2	69	.94	.90	-	-	.58	.82
3	29	.92	.87	.87	-	.52	.70
4	32	.94	.85	-	.86	.56	.80
3&4	61	.93	.86	-	-	.57	.75
<i>Organisms and Food test</i>							
1	29	.94	.88	-	.88	.70	.89
2	37	.93	.93	.81	-	.53	.72
1&2	66	.93	.91	-	-	.60	.83
3	29	.94	.88	-	.89	.59	.89
4	30	.88	.85	.91	-	.69	.83
3&4	59	.92	.86	-	-	.64	.86

Note. MC = Multiple-Choice; CR = Constructed-Response. ^aWorry data were collected only for the questions on the first part of each achievement test. Dashes in the table show cells where no data were collected.

For the multiple-choice achievement scores, the internal consistencies ranged from .52 to .70, and for the constructed-response achievement scores they ranged between .70 and .89. The multiple-choice achievement data were less reliable than the benchmark

of .70 that Gable and M. Wolf (1993) specify for achievement tests, whereas the achievement data for the constructed-response items met this criterion. One reason that might explain these lower reliabilities lies in the fact that each of the achievement tests was written to test a whole term's work, and consequently, each test required a relatively wide knowledge of science content. For example, the content of the *Earth Science* test included items on plate tectonics, volcanoes, earthquakes, rocks and minerals, and weather. Therefore, the achievement tests that were constructed for this research were likely to produce data with lower internal consistencies than achievement tests for shorter and more specifically focussed units of work. Moreover, the lower reliabilities of the multiple-choice items than the constructed-response items are most likely an artefact of the test formats. It is important to recall that scoring the multiple-choice items involved allocating either one or no marks to each item, whereas half marks were allowed for partially correct constructed-response items. Given that this allocation of marks is an artefact of the test formats, the achievement data were judged to be suitable for analytical procedures.

The Survey Data

The *Survey About Science Tests* did not require students to respond to researcher-defined constructs such as 'self-efficacy' and 'worry'. Instead, inferences about these constructs were developed by linking categories of students' responses to survey questions about their positive and negative thoughts, confidence and aspects they regard as easy and hard about test items to the research questions. One consequence of this approach was that the overlapping self-efficacy and worry constructs were confounded with each other. Therefore, in these analyses, format-related effects associated with boys' and girls' self-perceptions of low efficacy are grouped with the format-related effects that were associated with high levels of worry, and format-related effects that were associated with low worry are grouped with format-related effects that were associated with high self-efficacy.

Nevertheless, the check-recheck reliabilities for the coding of survey data into the categories ranged between .90 and .98 (Table 4.3, page 111). These reliabilities

show a very high level of agreement with the initial coding of students' responses after 6 weeks, and the suitability of these data for analytical purposes.

Table 4.3
Check-Recheck Reliabilities for the Survey Data

Question	Reliability
1a Negative thoughts about tests	.90
1b Positive thoughts about tests	.90
5 Characteristics of low confidence	.96
6 Students' responses to low confidence in tests	.93
8a Easy aspects of multiple-choice items	.98
8b Easy aspects of written answer items	.97
9a Difficult aspects of multiple-choice items	.95
9b Difficult aspects of written answer items	.92
10 Determinants of future achievement	.96

Factor Structure of Worry Items

Analysis of the worry data yielded by the pilot study indicated that, although items on the worry scale can be considered as unidimensional, they were much less so for boys than for girls. Consequently, a principal components analysis was conducted on the worry data yielded by the first of the quasi-experimental data collections to check its underlying factor structure. Interpretation of these analyses took into account several limitations of the data. First, the worry data were obtained from four different research groups who answered the worry scale in response to similar (but nevertheless, different) achievement items. Second, worry data for only 141 students were analysed. Although the ratio of cases to items for girls was in line with the benchmark of 10:1 advised by Gable and M. Wolf (1993), the ratio of cases to items for boys was only 7.6:1. However, Gable and M. Wolf (1993) note that some researchers in the field of instrument development have worked with minimum ratios of cases to items as low as 3:1. Nevertheless, neither of these limitations were judged to unduly threaten the validity of the worry data.

The principal components analysis was conducted first on data for all 141 cases.

The communality of all the items was set to 1.00 and a single factor with an eigenvalue greater than or equal to one, that is, Kaiser's criterion (Ferguson & Takane, 1989) was extracted (see Appendix H, Table H.1). All the worry items had substantial loadings, ranging from .51 for item 5 to .81 for item 3 onto this factor (Appendix H, Table H.2) indicating that the worry scale is unidimensional and that it can be used to discriminate between students according to the extent of their worrisome cognitions during tests.

The worry data were then subjected to further principal components analyses to establish if the factor structure was invariant across gender. Both of these principal components analyses yielded two factor solutions (Appendix H, Tables H.3 & H.4). Varimax rotation of the axes, where all factors are normalised prior to rotation, all vectors are of unit length and all variables receive the same weighting, that is, Kaiser's normalisation (Ferguson & Takane, 1989) supported separate two factor solutions for the girls and the boys (Appendix H, Table H.5). For the girls, items 1, 7 and 15 had substantial loadings onto one factor, items 5 and 11 had substantial loadings onto the other factor, and items 3, 9 and 13 loaded onto both factors. For the boys, items 1, 3, 9, 11, 13 and 15 had substantial loadings onto one factor, item 5 loaded onto the other factor, and item 7 loaded onto neither factor. Although the worry scale appeared to have functioned differently for the boys and the girls, the ratios of the lengths of the first two eigenvectors, 3.93 for the girls and 2.93 for the boys, indicated that the worry scale could still be regarded as unidimensional for the purposes of this study.

The Test Booklets

As explained in Chapter 3, the test booklets differed from the science achievement tests that are generally administered at the research site in three ways; (i) Students were asked to complete a self-efficacy scale for each of the achievement items, (ii) Students were required to complete the *Survey About Science Tests* halfway through the achievement tests, and (iii) Students were required to answer the multiple-choice and constructed-response achievement items in a particular order. Students were asked to comment on each of these changes during the interviews

because these departures from the standard test procedures could influence the reliability and validity of the quantitative data that were collected. Additionally, data were also sought to ascertain whether these research procedures functioned as they were intended.

As indicated earlier, a total of 12 girls and 10 boys from the research groups were interviewed (see Table 4.1). Evidence from these interviews suggests that overall, the above three changes do not unduly threaten the reliability and the validity of this study. Even though all but two of the students reported that rating their confidence and completing the *Attitude Towards Tests Survey* made them more aware of their thoughts and feelings during the achievement tests, only two students thought that these changes might have actually influenced their achievement. Additionally, four students reported that they were concerned about the time that was required to complete the extra tasks and three students thought that reversing the order of the item-response formats might have influenced their achievement.

The students who were interviewed were also asked about the way they answered the test booklets. A large majority of the students followed the instructions for answering the *Attitude Towards Tests Survey* as they were given. However, about half the students rated their self-efficacy before answering the achievement items and the other half rated their self-efficacy after answering the achievement items. Hence, boys' and girls' self-efficacies are confounded with their achievement expectancies. Although there is some evidence that suggests the order of answering confidence scales influences students' ratings of their confidences (Glenberg & Epstein, 1987), Bandura (1997) intimates that achievement expectancies for cognitive tasks are closely related to self-efficacy. Hence, differences in the order that the self-efficacy scales were answered were judged to pose an acceptable level of threat to the reliability and validity of this study.

THE CONTENT AND EXTENT OF GIRLS' AND BOYS' SELF-EFFICACIES AND THEIR WORRISOME COGNITIONS DURING TESTS

Research Questions Addressed

This section addresses the first two research questions for all 154 Year 9 boys and girls at the research site, that is,

- Are there gender-related differences in the content and extent of students' self-efficacies for answering science achievement questions set in multiple-choice and constructed-response formats? And consequently: is there a gender-related interaction between girls' and boys' perceptions of self-efficacy on science achievement questions and the response format these questions require?
- Are there gender-related differences in the content and extent of students' worrisome cognitions for answering science achievement questions set in multiple-choice and constructed-response formats? And consequently: is there a gender-related interaction between boys' and the girls' worrisome cognitions on science achievement questions and the response format these questions require?

Qualitative data from the *Survey About Science Tests* were used to compare the content of the girls' and the boys' positive and negative thoughts during tests, and quantitative data from the test booklets were used to determine the extent of the boys' and the girls' perceptions of self-efficacy and their worrisome cognitions during tests. Evidence for interactions between gender and test-item response formats were sought in the test booklets from girls' and boys' ratings of their self-efficacy and worry.

The Content of Boys' and Girls' Negative and Positive Thoughts During Tests

Introduction

The first question on the *Survey About Science Tests* asked students to list the negative and positive thoughts they experience during tests. It did not require them to

distinguish between multiple-choice and constructed-response questions, or to respond to researcher-defined categories such as 'self-efficacy' or 'worry'. As such, students' responses to this question, which are shown in Table 4.4, provide a context for the interpretation of their self-efficacies and their worrisome thoughts.

Additionally, analyses of girls' and boys' positive and negative thoughts, together with their responses to Questions 7 ('Write out some of the things you find easy about multiple-choice and written answer questions') and 8 ('Write out some of the things you find difficult about multiple-choice and written answer questions'), which are shown in Tables 4.5 and 4.6 (pages 116 & 117), enable the determination of format-related effects for the boys and the girls.

Table 4.4
Types of Thoughts Experienced by Girls and Boys During Tests

Content of Thoughts	Percentage of Students	
	Girls	Boys
Negative thoughts		
Worry	79	70
• Worry about results	54	56
• Worry about progress	49	33
• Worry about consequences	27	6
Test-directed thoughts	32	24
Self-directed thoughts	23	11
School-directed thoughts	13	22
Positive thoughts		
Favourable consequences	17	37
• Favourable achievement	17	29
• Future rewards	5	11
Positive self-talk	37	11
Self-directed thoughts	28	19
Favourable progress	22	24
Unclassified thoughts		
Escape	26	25

Note: Response categories are ranked from the most frequent to the least frequent.

Table 4.5
Features of Test Items Judged as Easy

Features of test items	Percentage of Students	
	Girls	Boys
Multiple-Choice Items		
Answers are provided	37	44
Able to use test-wise strategies	45	29
• Able to guess	35	17
• Other test-wise strategies	13	11
Have a better chance	33	30
Able to recognise answers	18	16
Don't require writing	13	14
Less time consuming	6	14
Constructed-Response Items		
Chance to demonstrate learning	17	17
Can express ideas	17	11
Easier to understand	17	8
Able to use test-wise strategies	10	6

Note: Response categories are ranked from the most frequent to the least frequent.

Table 4.6
Features of Test Items Judged as Difficult

Features of test items	Percentage of Students	
	Girls	Boys
Multiple-Choice items		
Distinguishing between distractors	41	33
Wording of items	15	3
Constructed-Response items		
More rigorous	44	19
Writing responses	35	25
Difficulty understanding	15	10
Time consuming	7	14

Note: Response categories are ranked from the most frequent to the least frequent.

The Content of Girls' and Boys' Negative Thoughts During Tests

The proportions of boys and girls responding to questions about their negative thoughts during tests, and the mean frequencies for their responses to these questions are shown in Appendix I (Table I.1). Almost all the girls and boys listed examples of the types of negative thoughts they experience during tests. However, the response characteristics for both the boys and the girls were greater for constructed-response questions than multiple-choice questions. These data suggest that the students regarded constructed-response items as more difficult than multiple-choice items. Overall, the mean frequencies of the girls' and the boys' responses for survey Questions 8a and 8b were similar, though marginally higher for the girls than the boys. Therefore, interpretation of gender-related differences for the boys' and the girls' negative thoughts on multiple-choice and constructed-response tests must take into account the girls' relatively higher response rates for Question 8 (Appendix I, Table I.1).

Five categories of negative thoughts emerged from analyses of the girls' and the boys' responses to Question 1, namely; worry, test-directed thoughts, self-directed thoughts, school-directed thoughts and thoughts about escape. Thoughts about escape also figured in students' responses to Question 1 as positive thoughts about

tests. However, the content of these thoughts was similar to the content of the negative thoughts that were categorised as escape. The proportions of the boys and the girls responding to each category and selected groups within these categories are shown as percentages in Table 4.4. Interpretation of these data is not problematic because the girls' and the boys' response characteristics were similar (Appendix I, Table I.1).

Thoughts expressing concern about receiving unfavourable results (for example, 'I'm going to fail'), the consequences associated with an unfavourable achievement ('What are my parents going to say if I fail') and concerns about progress during tests ('I can't do this', 'I don't know any of this', 'I don't have enough time', and 'I can't remember the stuff I studied') were grouped together because of their conceptual similarity with each other. Moreover, the categorisation of these responses is consistent with the way that worry was conceptualised in Chapter 2. Additionally, worrisome thoughts that related to progress, such as 'I can't remember the stuff I studied,' involved expressions of generalised low self-efficacy. Under this classificatory scheme, low self-efficacy appeared as one of the dimensions of worry. On the other hand, worrisome thoughts that were associated with achievement appeared to be more closely related to students' achievement and outcome expectancies than their generalised self-efficacies.

A high proportion of boys (70%) and a slightly higher proportion of the girls (79%) reported the presence of worrisome thoughts during tests. Although similar proportions of the girls and the boys expressed concerns about their results, a much greater proportion of the girls than the boys expressed their concerns about the consequences of receiving an unfavourable achievement (27% in comparison to 6%) and about their progress during tests (49% in comparison to 33%). These data indicate that a higher proportion of the girls than the boys perceive generalised thoughts of low self-efficacy during tests.

Self deprecatory thoughts (for example, 'I am really stupid'), thoughts relating to self-blame ('I should have studied more') and self-doubt ('Have I studied enough?') were categorised as self-related thoughts because students directed these thoughts toward themselves. Comparison of the boys' and the girls' responses show that a higher proportion of the girls (23%) reported the presence of negative self-

related thoughts than the boys (11%). These gender-related differences were derived from the higher proportion of the girls (15%) than the boys (5%) who expressed thoughts of self-blame and regret. Otherwise, similar proportions of the girls and the boys produced responses that were coded as self-deprecatory thoughts and self-doubt, which represented generalised expressions of their low self-efficacy.

Groups of negative thoughts coded as difficulty (for example, 'This test is so hard'), content validity ('We haven't been taught this'), confusion ('Questions don't make sense') and time ('There isn't enough time') were categorised as test-directed thoughts. A higher proportion of the girls (32%) than the boys (24%) expressed these specific types of negative thoughts. On the other hand, a higher proportion of the boys (22%) than the girls (13%) reported negative thoughts that were directed more generally at their schooling (for example, 'Science is so difficult', 'I hate tests', 'I hate science tests').

Overall, it is important to note that the boys and the girls reported similar types of negative thoughts as each other during tests. Although worrisome thoughts were by far the most prominent of these negative thoughts, there was a relatively greater tendency for the girls to direct their negative thoughts at themselves and for the boys to direct their negative thoughts outside themselves toward schooling and tests.

The preceding analyses show that a large majority of the students (80% of the girls and 87% of the boys) gave responses relating to the difficulty of tests. These thoughts were expressed, either as negative thoughts that were directed toward tests or as worrisome thoughts. Additionally, Question 8 from the *Survey About Science Tests* required students to report aspects of multiple-choice and constructed-response questions they regarded to be difficult. Therefore, format-related differences between the boys and the girls for their self-efficacies and their worrisome cognitions can be inferred from their responses to survey Question 8.

Two categories of negative responses emerged when the girls and the boys were asked to describe aspects of multiple-choice questions they regarded to be difficult (Table 4.6). Nearly half of the students reported that distinguishing between the multiple-choice distractors (for example, 'They sometimes confuse me because they are so alike', 'When you think 'none of these answers sound wright (sic) or all of those answers sound possible', and 'can't find one that you think is right') is

problematic. A much smaller proportion of the students reported difficulties associated with the wording of multiple-choice questions ('It's hard to get a good picture of what they are talking about', 'Don't know what some words mean').

Four categories of negative responses emerged when the boys and the girls were asked to describe the difficulties they associated with constructed-response questions (Table 4.6). One third of the students expressed the view that constructed-response questions are more rigorous than multiple-choice questions either because these types of questions require more knowledge (for example, 'They are harder because you have to think of everything', and 'You have to study more as the answer is not readily available as in multiple-choice') or because they (that is, the students) cannot guess constructed-response answers ('[You] can't guess with much success', 'They don't give any clues like multiple choice do'). One third of the students also identified the construction of written responses as problematic. Negative thoughts grouped into this category related to the expression of students' ideas ('[You] have to make up your own words', 'Sometimes you know it but you do not know how to write it out on paper'), the construction of long responses ('writing so much', 'writing really long answers') and sundry aspects relating to writing, spelling and grammar ('have to write out the answer', 'spell things wrong'). Additionally, smaller proportions of the students provided responses relating to their difficulty understanding constructed-response questions ('Questions are hard to understand', 'The question is not clear') and time ('They take a lot of time').

The Content of Girls' and Boys' Positive Thoughts During Tests

The proportions of the boys and the girls responding to questions about their positive thoughts during tests and the mean frequencies for their responses to these questions are shown in Appendix I (Table I.2). A large majority of the girls and the boys listed examples of positive thoughts they experience during tests. Fewer students reported aspects of constructed-response questions they regarded to be easy in comparison with multiple-choice questions. This pattern was consistent with the boys' and the girls' responses to Question 6 on the *Survey About Science Tests* ('Which type of questions do you generally feel more confident answering; multiple-

choice or written answer questions?'), where 77% of the students reported that they were more confident answering multiple-choice questions than constructed-response questions. Overall, the mean frequencies of the girls' and the boys' responses for each of the questions were similar. However, a much lower proportion of the boys than the girls reported positive thoughts about constructed-response items, which needs to be taken into account when interpretations are made of these data.

Five categories of positive thoughts emerged from analyses of the girls' and the boys' responses to Question 1, namely; favourable consequences, positive self-talk, self-directed thoughts, thoughts relating to favourable progress on the test and thoughts about escape. The proportions of the boys and the girls responding to each of these categories and selected groups within these categories are shown as percentages in Table 4.4. Interpretation of these data is not problematic because the girls' and the boys' response characteristics were similar (Appendix I, Table I.2).

About one quarter of the students produced responses to each category of positive thoughts. Groups of thoughts relating to favourable achievement (for example, 'I'm going to get an A or maybe a B') and future rewards ('My mum will be happy if I go well') were categorised as favourable consequences. Thoughts classified within this category represented positive achievement and outcome expectancies. The category of self-talk subsumed students' self-messages that were coded as encouragement and reassurance (for example, 'I've studied so I should no (sic) the answer'), exhortation ('Hurry up, I need time') and hope ('I hope I pass.'). Positive thoughts that the students expressed about themselves and their capabilities (for example, 'I am so smart', 'I can do this') were categorised as self-related thoughts. Although this last group of thoughts could be regarded as expressions of students' generalised self-efficacy, this categorisation is uncertain because these types of thoughts may represent no more than wishful thinking. The remaining category of positive thoughts related to students' perceptions about their progress during tests. Thoughts subsumed within this category related to the relative difficulty of tests (for example, 'This is easy'), doing well ('I'm going quite well'), answering questions correctly ('I know I got this one right') and time ('I'm going to finish this test'). This category of responses appeared to be related more to the boys' and the girls' achievement expectancies than their generalised self-efficacies.

A much higher proportion of the boys (37%) than the girls (17%) reported thoughts about favourable consequences, that is, positive achievement and outcome expectancies during tests. Higher proportions of the boys than the girls reported favourable consequences associated with their achievement (17% of the girls and 29% of the boys) and future rewards (11% of the boys and 5% of the girls). A much higher proportion of the girls (37%) than the boys (11%) expressed thoughts that were categorised as self-talk. This pattern of gender differences was consistent across all three groups of responses categorised as self-talk, namely, encouragement, exhortation and hope. Similarly, a higher proportion of the girls (28%) than the boys (19%) reported positive thoughts about themselves and their capabilities during tests, that is, thoughts relating to their generalised high self-efficacy. Nevertheless, similar proportions of the boys (24%) and the girls (22%) reported positive achievement expectancies relating to their progress during tests.

Approximately half of the students (47% of the girls and 60% of the boys) reported thoughts that were identified as expressions of their positive achievement and outcome expectancies, and to a lesser extent their generalised self-efficacies. Additionally, Question 7 from the *Survey About Science Tests* required students to report aspects of multiple-choice and constructed-response questions they regard to be easy. Hence, format-related differences between the boys and the girls for their expectancies and their self-efficacies can be inferred from their responses to Question 7.

Six categories of responses for multiple-choice questions (Table 4.5) emerged from students' answers to Question 7 on the *Survey About Science Tests* ('Write out some of the things you find easy about multiple-choice and written answer questions'). Nearly half of the students reported that multiple-choice questions are easier than constructed-response questions because the answer is provided for them (for example, 'The answer is there', 'You have the write (sic) answer in front of you') and that all that is required is to select the answer ('You simply have to pick an answer, not think one up'). A little more than a third of the students reported that multiple-choice questions are easier than constructed-response questions because they can use test-wise strategies such as guessing (for example, 'Guess if you don't know answer'), the elimination of responses ('If you are unsure you can cross out the

definitely wrong, & then have more chance of getting it right') and using clues within the question ('You have a clue about the answer'). Additionally, one third of the students also reported that multiple-choice questions are relatively easy because they have a better chance of getting them correct (for example, 'You get more of a chance of getting it right', 'get a 25% chance'). Only a small proportion of the students produced responses to the remaining three categories, that multiple-choice questions are easier because it is possible to recognise the answer (for example, 'You know the answer when you see it', 'It helps you to remember seeing the answer'), that writing is not required ('You don't have to write much', 'You don't have to worry about how you word the answer') and that they are less time consuming ('there (sic) shorter', 'They are quicker to do').

A little more than half of the students produced responses for Question 7 on the *Survey About Science Tests* ('Write out some of the things you find easy about multiple-choice and written answer questions') that related to constructed-response items. These responses were grouped into four categories (Table 4.5). Both the girls and the boys reported that constructed-response questions are relatively easy because they provide an opportunity for students to demonstrate their learning (for example, 'You can explain an answer further than a one word answer', 'You have a chance to explain in more depth'). Additionally, students reported that constructed-response questions are easier than multiple-choice questions because they can express their ideas by constructing their own explanations (for example, 'You can explain things your own way') and using their own words ('You can put it into your own words'). A small proportion of the students indicated that constructed-response questions are easy to understand (for example, 'Don't mix me up', 'They are sometimes worded easier than multiple-choice questions') and answer ('You can memorize (sic) definitions from the text book', 'straight forward'). Only a very small proportion of students reported that constructed-response questions are relatively easy because they can use test-wise strategies (for example, 'You can use your communication skills', 'Get 1/2 marks if you sort of know').

A Comparison of Format-Related Effects for Boys and Girls

Generally, higher proportions of the girls than the boys provided responses to the survey questions about multiple-choice and constructed-response questions (Appendix I). Therefore, the relative importance of the response categories for the boys and the girls must be considered rather than the absolute proportions of the girls and the boys responding to each category. As indicated earlier, interpretation of these data must recognise that low self-efficacy and worry are confounded in these analyses, as are self-efficacy and expectancy. However, empirical distinctions between self-efficacy and expectancies, irrespective of whether they are achievement or outcome expectancies, are inconsequential when outcomes are controlled by the quality of performances (Bandura, 1997) as they generally are for achievement tests. Moreover, Bandura (1997) asserts that anxiety and worrisome thoughts arise from perceptions of low self-efficacy. Therefore, inferences about the content of boys' and girls' generalised self-efficacies for constructed-response and multiple-choice questions, together with the content of their worrisome cognitions can be made from the data in Tables 4.5 and 4.6.

No gender-differences were apparent for the content of the girls' and the boys' generalised low self-efficacies (and high worrisome thoughts) for multiple-choice items (see Table 4.7).

Table 4.7

Relative Importance of the Major Response Categories for Boys' and Girls' Thoughts About Multiple-Choice Items.

	Boys		Girls
Low self-efficacy (and high worry)			
Distractors	(33)	Distractors	(41)
Wording of items	(3)	Wording of items	(15)
High self-efficacy (and low worry)			
Answers provided	(44)	Test-wise strategies	(45)
Have a better chance	(30)	Answers provided	(37)
Test-wise strategies	(29)	Have a better chance	(33)

Note. Response categories are ranked from the most frequent to the least frequent. Numbers in parentheses show the percentage of students responding to each category.

Both the boys and the girls clearly reported more difficulty distinguishing between multiple-choice distractors than understanding the wording of these types of items. However, the content of the girls' and the boys' elevated self-efficacies (and their perception of low worry) for multiple-choice questions differed. Although the boys and girls reported higher self-efficacies for multiple-choice questions because the answers are provided and because they have a better chance on these types of questions, the use of test-wise strategies was relatively more important for the girls than the boys.

Small gender differences were detected for the content of the boys' and the girls' low self efficacies (and high worrisome thoughts) on constructed-response questions (see Table 4.8).

Table 4.8

Relative Importance of the Major Response Categories for Girls' and Boys' Thoughts About Constructed-Response Items.

Boys		Girls	
Low self-efficacy (and high worry)			
Writing responses	(25)	More rigorous	(44)
More rigorous	(19)	Writing responses	(35)
Time consuming	(14)	Understanding	(15)
Understanding	(10)	Time consuming	(7)
High self-efficacy (and low worry)			
Demonstrate learning	(17)	Demonstrate learning	(17)
Can express ideas	(11)	Can express ideas	(17)
Easier to understand	(8)	Easier to understand	(17)

Note. Response categories are ranked from the most frequent to the least frequent. Numbers in parentheses show the percentage of students responding to each category.

The girls expressed relatively more worry than the boys about the greater rigour of constructed response questions whereas the construction of written responses was relatively more problematic for the boys. Additionally, the boys expressed relatively more concern about the time that is required to complete constructed-response questions whereas the girls reported that understanding constructed-response questions was relatively more problematic for them. On the other hand, the boys reported that the chance to demonstrate their learning was their chief source of high

self-efficacy for constructed-response questions. This factor appeared to be relatively less important for the girls' high self-efficacies.

The Extent of Girls' and Boys' Self-Efficacies for Multiple-Choice and Constructed-Response Test Items

Data Sources

Data for these analyses were obtained from the test booklets that were completed by all students for the *Earth Science* and the *Organisms and Food* achievement tests. Summary statistics for the extent of the boys' and the girls' self-efficacies by test format, achievement test and combinations of research groups are shown in Table 4.9. These summary statistics were obtained by aggregating raw scores for students in each of the research groups that answered the same multiple-choice and constructed-response questions. This procedure was essential because it counterbalanced the order that the multiple-choice and constructed-response questions were answered, and it was justified because each of the research groups contained similar numbers of girls and boys.

General Findings

The mean self-efficacies for the boys and the girls on science achievement items show that in every instance that the boys reported higher mean self-efficacies than the girls (Table 4.9 on page 127). This pattern of higher self-efficacies for the boys was detected for both the multiple-choice and the constructed-response test formats. Effect sizes for these differences in self-efficacy ranged from differences that were so small as to have no practical significance ($d = +0.08$) to large differences that were likely to be practically significant ($d = +0.82$). Moreover, the students generally expressed higher self-efficacies for multiple-choice items than the parallel constructed-response items (Appendix J, Table J.1). However, it is important to note in this context that these latter comparisons were drawn for items testing the same content but with different groups of girls and boys. These higher self-efficacies for multiple-

choice items than constructed-response items are consistent with the boys' and the girls' responses to Question 6 on the *Survey About Science Tests* ('Which type of questions do you generally feel more confident answering: multiple-choice or written answer questions?'), where 77% of the students reported that they were more confident answering multiple-choice questions and only 10% of the students expressed more confidence for answering constructed-response items.

Table 4.9
Item-Specific Self-Efficacies for Science Achievement Items

Students	n	MC items			CR items		
		<u>M</u>	<u>SD</u>	<u>d</u>	<u>M</u>	<u>SD</u>	<u>d</u>
<i>Earth Science test</i>							
Groups 1&2							
Boys	29	3.16	0.75		2.98	0.79	
Girls	40	3.10	0.89	+0.08	2.86	0.89	+0.15
Groups 3&4							
Boys	25	3.38	0.70		3.23	0.74	
Girls	36	2.84	0.64	+0.76	2.74	0.63	+0.70
<i>Organisms and Food test</i>							
Groups 1&2							
Boys	29	3.21	0.70		2.75	0.87	
Girls	37	2.93	0.79	+0.37	2.49	0.79	+0.32
Groups 3&4							
Boys	24	3.09	0.57		3.09	0.69	
Girls	35	2.58	0.68	+0.75	2.50	0.64	+0.82

Note. Self-efficacy judgements were made on 5-point scales (1 = very little confidence, 5 = quite a lot of confidence). Positive effect sizes indicate higher self-efficacy scores for boys than girls. MC = Multiple-Choice; CR = Constructed Response; d = effect size.

Specific Comparisons of Girls' and Boys' Item-Specific Self-Efficacies

Gender-related differences between the boys' and the girls' item-specific self-efficacies were much larger for the students in Research Groups 3 and 4 than for the students in Research Groups 1 and 2. Although this pattern was evident on both

achievement tests, it was less pronounced for the *Organisms and Food* test than for the *Earth Science* test. For the students in Research Groups 1 and 2, the mean effect sizes for differences between the girls' and the boys' self-efficacies were small for the multiple-choice ($d = +0.23$) and the constructed-response items ($d = +0.24$).

However, the mean effect sizes for differences between the boys' and the girls' self-efficacies for students in Research Groups 3 and 4 approached large sizes for multiple-choice ($d = +0.76$) and constructed-response items ($d = +0.76$).

When effect sizes for differences between the girls' and the boys' item-specific self-efficacies were averaged across the research groups, they showed that gender-related differences in self-efficacy were larger for the *Organisms and Food* test than the *Earth Science* test, for both the multiple-choice and constructed-response formats. For the *Earth Science* test, the mean effect sizes for differences between the boys' and the girls' self-efficacies approached moderate sizes for the multiple-choice ($d = +0.42$) and the constructed-response ($d = +0.43$) formats. However, the mean effect sizes for differences between the girls' and the boys' self-efficacies on the *Organisms and Food* test were moderate for both the multiple-choice ($d = +0.56$) and the constructed-response ($d = +0.57$) formats. Therefore, the boys expressed relatively more self-efficacy than the girls for the *Organisms and Food* topic than they did for the *Earth Science* topic.

Overall, the mean effect sizes for differences between the boys' and the girls' item-specific self-efficacies for multiple-choice and constructed-response formats, obtained by averaging effect sizes across research groups and achievement tests, were moderately sized for multiple-choice ($d = +0.49$) and constructed-response ($d = +0.50$) formats. These gender-related differences between the girls' and the boys' self-efficacies for answering science achievement items were likely to be practically significant.

Interactions Between Gender, Item-Specific Self-Efficacy and Response Formats

The previous analyses have shown that the boys and the girls reported lower item-specific self-efficacies for constructed-response formats than for multiple-choice formats, and that the boys reported higher self-efficacies than the girls for both

multiple-choice and constructed-response questions. These patterns existed for combinations of tests (Figure 4.1), research groups (Figure 4.2), and combinations of tests and research groups.

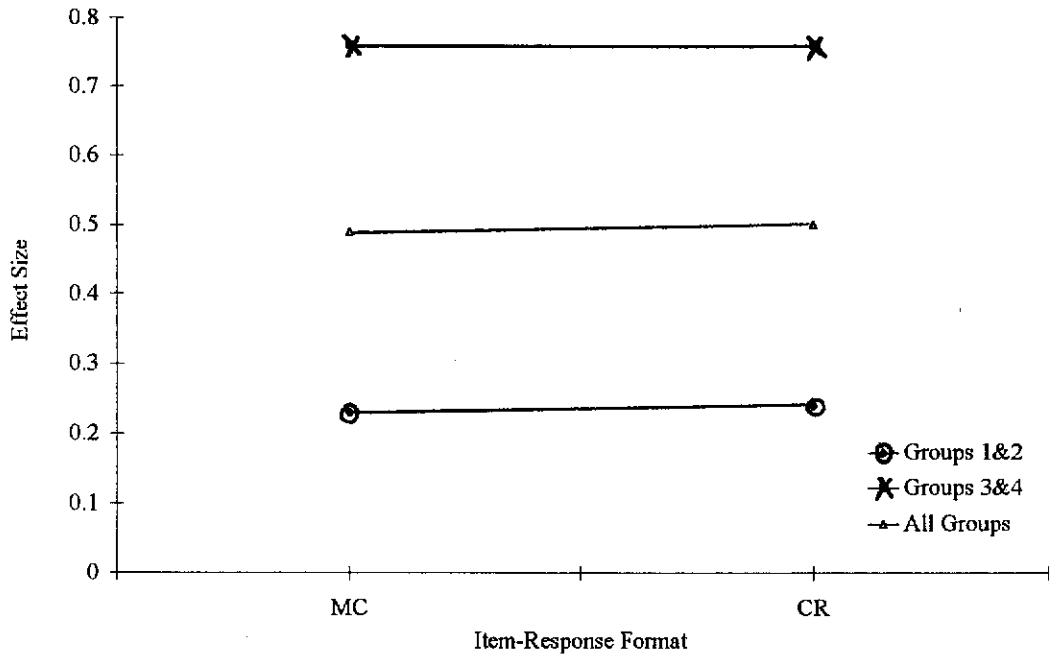


Figure 4.1
Differences Between Girls' and Boys' Self-Efficacies for Multiple-Choice and Constructed-Response Questions by Research Group

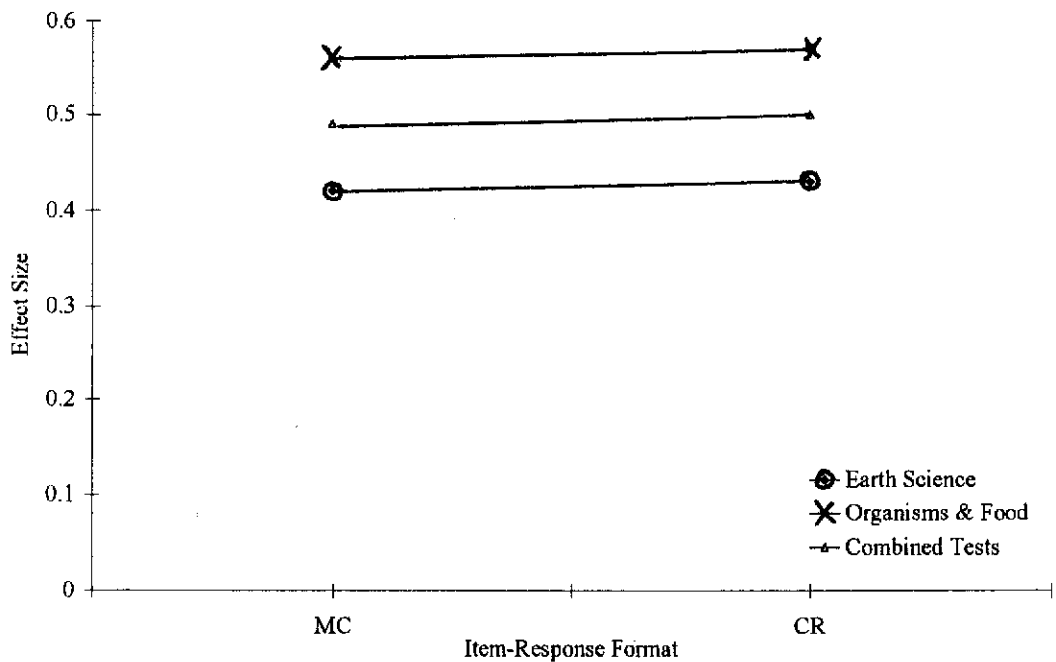


Figure 4.2
Differences Between Boys' and Girls' Self-Efficacies for Multiple-Choice and Constructed-Response Questions by Unit Topic

Importantly, Figures 4.1 and 4.2 show the absence of an interaction between gender and test item-response formats for students' self-efficacies on science achievement items.

The Extent of Boys' and Girls' Worrisome Cognitions for Multiple-Choice and Constructed-Response Test Items

Data Sources

Data for these analyses were obtained from the test booklets that were completed by all the students for the *Earth Science* and the *Organisms and Food* achievement tests. Summary statistics for the extent of the girls' and the boys' worrisome cognitions by test format, achievement test and research group are shown in Tables 4.10 and 4.11. However, it is important to recall that the worry data were obtained from students for only one of the item-response formats on each of the achievement tests, unlike the achievement and item-specific self-efficacy data.

General Findings

The girls reported higher worry scores than the boys in all but one of the cases (Tables 4.9 & 4.10, pages 131 & 132). Effect sizes for these differences ranged from differences that were so small that they were unlikely to be practically significant ($d = +0.05$) to moderately large differences that were likely to be practically significant ($d = -0.73$). Additionally, students generally expressed relatively higher levels of worry for constructed-response items than for the parallel multiple-choice items (Appendix J, Table J.2). However, it is important to note in this context that these latter comparisons were drawn for items that tested the same content but with different groups of boys and girls.

Table 4.10
Extent of Boys' and Girls' Worrying Cognitions for Answering *Earth Science* Questions

Students	n	MC items			CR items		
		<u>M</u>	<u>SD</u>	<u>d</u>	<u>M</u>	<u>SD</u>	<u>d</u>
Group 1							
Boys	15	16.40	4.67		-	-	
Girls	18	17.41	5.20	-0.21	-	-	-
Group 2							
Boys	14	-	-		14.93	3.05	
Girls	22	-	-	-	18.36	5.09	-0.73
Group 3							
Boys	11	14.91	4.39		-	-	
Girls	18	18.11	5.36	-0.62	-	-	-
Group 4							
Boys	14	-	-		15.78	5.75	
Girls	18	-	-	-	19.22	5.50	-0.59

Note. Worry data were collected only for the questions on the first part of each achievement test; dashes indicate cells where no data were collected. Minimum score = 8; Maximum score = 32. High scores correspond to high levels of worry. Positive effect sizes indicate higher worry scores for boys than girls. MC = Multiple-Choice; CR = Constructed Response; d = effect size.

Table 4.11
Extent of Girls' and Boys' Worrisome Cognitions for Answering *Organisms and Food* Questions

Students	n	MC items			CR items		
		M	SD	d	M	SD	d
Group 1							
Boys	12	-	-		17.92	6.05	
Girls	17	-	-	-	17.62	6.15	+0.05
Group 2							
Boys	17	14.03	4.82		-	-	
Girls	20	15.70	4.65	-0.35	-	-	-
Group 3							
Boys	11	-	-		15.55	4.99	
Girls	18	-	-	-	16.69	6.31	-0.20
Group 4							
Boys	13	16.69	7.08		-	-	
Girls	17	17.21	6.37	-0.08	-	-	-

Note. Worry data were collected only for the questions on the first part of each achievement test. Minimum score = 8; Maximum score = 32. High scores correspond to high levels of worry. Positive effect sizes indicate higher worry scores for boys than girls. MC = Multiple-Choice; CR = Constructed Response; d = effect size. Dashes indicate cells where no data were collected.

Specific Comparisons of the Extent of Girls' and Boys' Worrisome Cognitions

Effect sizes for differences between the extent of the boys' and the girls' worrisome cognitions, averaged across the *Earth Science* and the *Organisms and Food* tests, were larger for Research Groups 3 and 4 than for Research Groups 1 and 2. For the students in Research Groups 1 and 2, the mean effect sizes were small for the multiple-choice ($d = -0.28$) and the constructed-response ($d = -0.34$) items, whereas for students in Research Groups 3 and 4 they were small to moderately sized for the multiple-choice ($d = -0.35$) and the constructed-response items ($d = -0.40$).

When the effect sizes for differences between the extent of the girls' and the boys' worrisome cognitions were averaged across the research groups, the gender-

related differences in worry were smaller for the *Organisms and Food* test than the *Earth Science* test. For the *Earth Science* test, the mean effect sizes approached a moderate size for the multiple-choice items ($d = -0.42$) and they were moderately sized for the constructed-response ($d = -0.66$) items. However, the mean effect sizes for differences between the extent of the boys' and the girls' worrisome cognitions on the *Organisms and Food* test were small for the multiple-choice items ($d = -0.22$) and negligible for the constructed-response ($d = -0.08$) items.

Overall, the mean effect sizes for differences between the extent of the girls' and the boys' worrisome cognitions for answering multiple-choice and constructed-response items, obtained by averaging effect sizes across research groups and achievement tests, were small to moderately sized for multiple-choice ($d = -0.32$) and constructed-response ($d = -0.37$) items. These gender differences, where the girls reported higher worry scores than the boys, were likely to be practically significant.

Interactions Between Gender, Worry and Item-Response Formats

The preceding analyses for gender-related differences in the extent of the boys' and the girls' worrisome cognitions are displayed in Figures 4.3 and 4.4 (page 134). These figures show the relationship between gender and test formats when the worry data were combined by tests (Figure 4.3) and research groups (Figure 4.4). It is important to recall in this context that only one set of worry data was collected in each of the test booklets, and that these data related to the item format that the students had just completed answering. Therefore, direct comparisons with Figures 4.1 and 4.2, which display gender-related differences in item-specific self-efficacy are limited.

Figure 4.4 shows that the direction of the interaction between gender and test format was not consistent across the tests. On the other hand, the direction of the interaction was consistent when the worry data were analysed by research groups (Figure 4.3). The overall effect sizes for differences between the extent of the girls' and the boys' worrisome cognitions were slightly higher for the constructed-response items than the multiple-choice items. In other words, the boys expressed relatively less worry than the girls for constructed-response items. (Alternatively, the girls

expressed relatively less worry than the boys for multiple-choice items). However, the extent of this interaction was quite small and it is not likely that it had any practical importance.

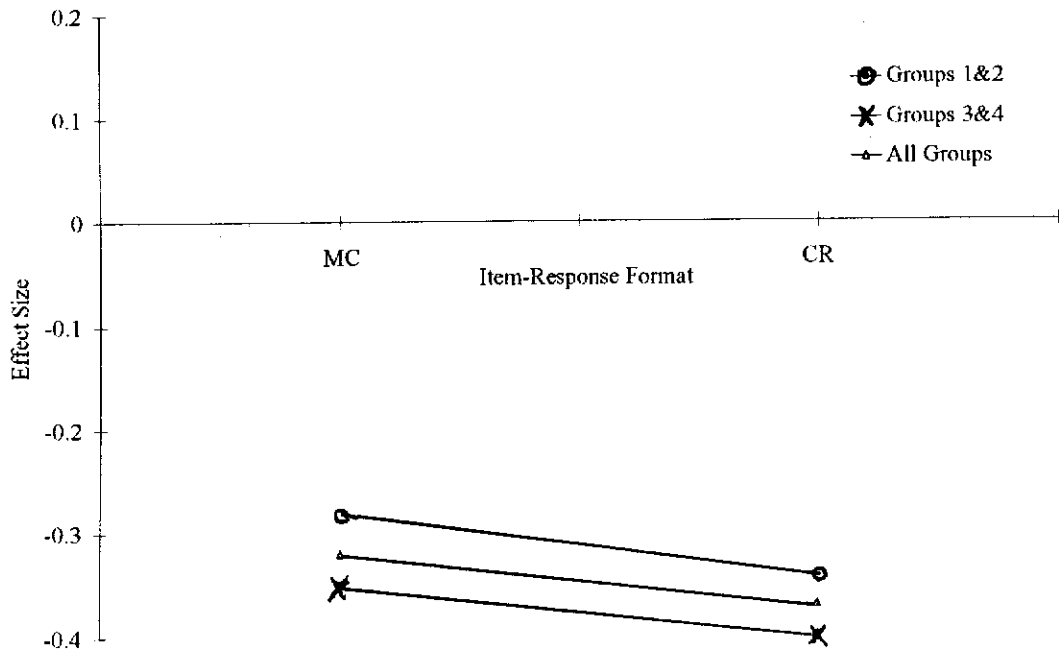


Figure 4.3
Differences between the Extent of Girls' and Boys' Worrisome Cognitions for Multiple-Choice and Constructed-Response Items by Research Group

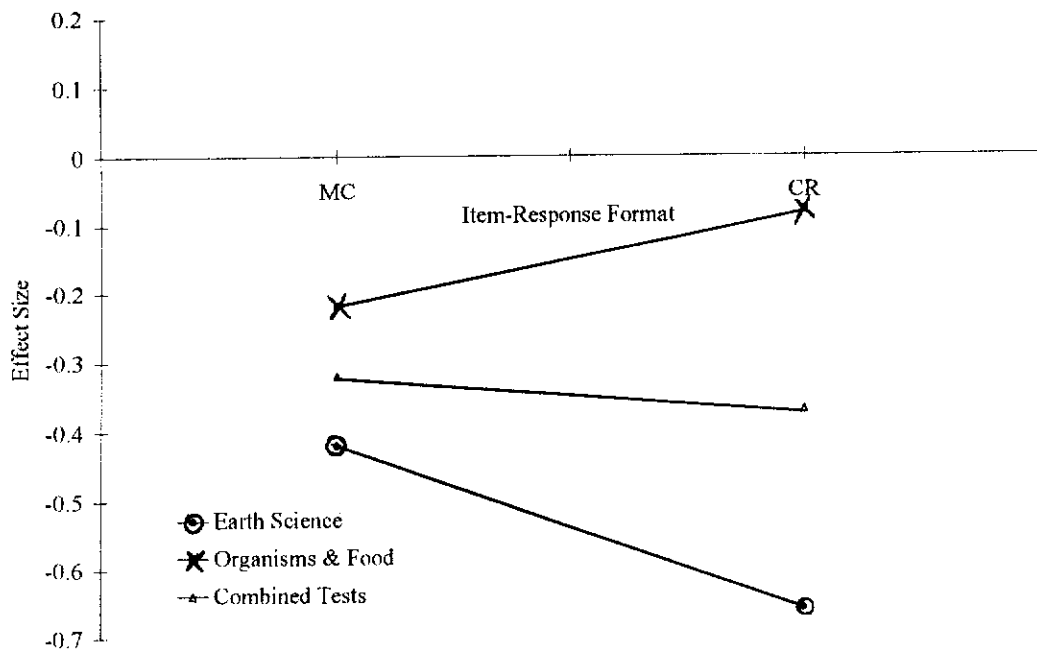


Figure 4.4
Differences between the Extent of Boys' and Girls' Worrisome Cognitions for Multiple-Choice and Constructed-Response Items by Unit Topic

THE RELATIONSHIPS BETWEEN SELF-EFFICACY, WORRY AND ACHIEVEMENT

Research Questions Addressed

This section of the data analyses tests the conceptual framework that underpins this study by addressing the third of the research questions ('How are the boys' and the girls' perceptions of self-efficacy and worry related to their achievement on multiple-choice and constructed-response test formats?'), for all the 154 students who were involved in the data collections. Additionally, findings are also presented for the subsidiary research question, 'Are gender-related differences in self-efficacy dependent on whether achievement items are answered correctly or incorrectly?', which emerged during the analyses.

Results from the quasi-experimental data analyses are used to indicate the extent of the associations between achievement, self-efficacy and worry for each of the response formats, and analyses of the survey data to provide descriptions for the way in which these variables are related.

The Achievement Context

Data sources

Data for these analyses were obtained from the *Earth Science* and the *Organisms and Food* achievement tests. Summary statistics for the boys' and the girls' achievements by test format, achievement test and groups of students are shown in Table 4.12 (page 136). These summary statistics were obtained by aggregating raw scores for the students in each of the research groups that answered the same multiple-choice and constructed-response questions. This procedure was essential because it counterbalanced the order that the multiple-choice and constructed-response questions were answered, and it was justified because each of the research groups contained similar numbers of girls and boys.

Table 4.12
Achievement on Test Item-Response Formats

Groups	n	<u>MC</u> items			<u>CR</u> items		
		<u>M</u>	<u>SD</u>	<u>d</u>	<u>M</u>	<u>SD</u>	<u>d</u>
<i>Earth Science test</i>							
Groups 1&2							
Boys	29	12.10	3.23		9.74	4.28	
Girls	40	12.25	2.96	-0.05	10.79	4.25	-0.25
Groups 3&4							
Boys	25	11.40	3.07		8.32	3.07	
Girls	36	10.95	3.21	+0.15	8.00	3.46	+0.10
<i>Organisms and Food test</i>							
Groups 1&2							
Boys	29	12.41	3.05		9.02	3.16	
Girls	37	12.59	3.10	-0.06	8.97	4.51	+0.01
Groups 3&4							
Boys	24	11.50	3.15		9.73	3.87	
Girls	35	11.86	3.49	-0.11	9.03	4.77	+0.16

Note. Maximum score = 20. Positive effect sizes indicate higher achievement scores for boys than girls. MC = Multiple-Choice; CR = Constructed Response; d = effect size.

General Findings

The boys' and the girls' achievement scores show no clear pattern of gender-related differences (Table 4.12). The girls achieved higher scores than the boys in four of the eight instances, once for the constructed-response items and three times for the multiple-choice items. Moreover, for seven of the eight cases the effect sizes were so small as to be trivial. Otherwise, small gender-related differences ($d = -0.25$) in achievement were detected for the students in Research Groups 1 and 2 on the constructed-response section of the *Earth Science* test. Additionally, the mean achievement scores for the girls and the boys were higher for the multiple-choice items than the parallel constructed-response items in every instance (Appendix J, Table J.3), though in one case this difference was quite small. However, it is

important to note that these latter comparisons involved items that tested the same content but with different groups of boys and girls.

Specific Comparisons of Girls' and Boys' Achievements

Effect sizes for differences between the boys' and the girls' achievements were averaged across the *Earth Science* and the *Organisms and Food* tests for students in Research Groups 1 and 2, and Research Groups 3 and 4. For the students in Research Groups 1 and 2, the mean effect sizes for differences between the girls' and the boys' achievements were negligible on the multiple-choice ($d = -0.06$) and the constructed-response ($d = -0.12$) items. Similarly, the mean effect sizes for differences in achievement for the boys and the girls from Research Groups 3 and 4 were negligible for the multiple-choice ($d = +0.02$) and the constructed-response ($d = +0.13$) items.

Effect sizes for differences between the girls' and the boys' achievements, averaged across the research groups for the *Earth Science* and the *Organisms and Food* tests, show that gender-related differences in achievement are negligible. For the *Earth Science* test, the mean effect sizes for differences between the boys' and the girls' achievement scores were negligible for the multiple choice ($d = +0.05$) and the constructed-response ($d = -0.08$) items. Similarly, there were essentially no gender-related differences in achievement for the multiple-choice ($d = -0.09$) and the constructed-response ($d = +0.09$) items on the *Organisms and Food* test.

Overall, the mean effect sizes for differences between the girls' and the boys' achievements, obtained by averaging effect sizes across research groups and achievement tests, show the absence of any practically important differences between the boys' and the girls' science achievements on multiple-choice ($d = -0.02$) and constructed-response ($d = +0.01$) items.

Interaction Between Gender, Achievement and Item-Response Format

The previous analyses reported the absence of any practically important differences between the girls' and the boys' achievements on multiple-choice and

constructed-response item formats. These patterns of achievement were observed consistently for combinations of tests (Figure 4.5) research groups (Figure 4.6) and combinations of research groups and tests. Hence, no overall interaction was detected between gender and test item-response formats for science achievement.

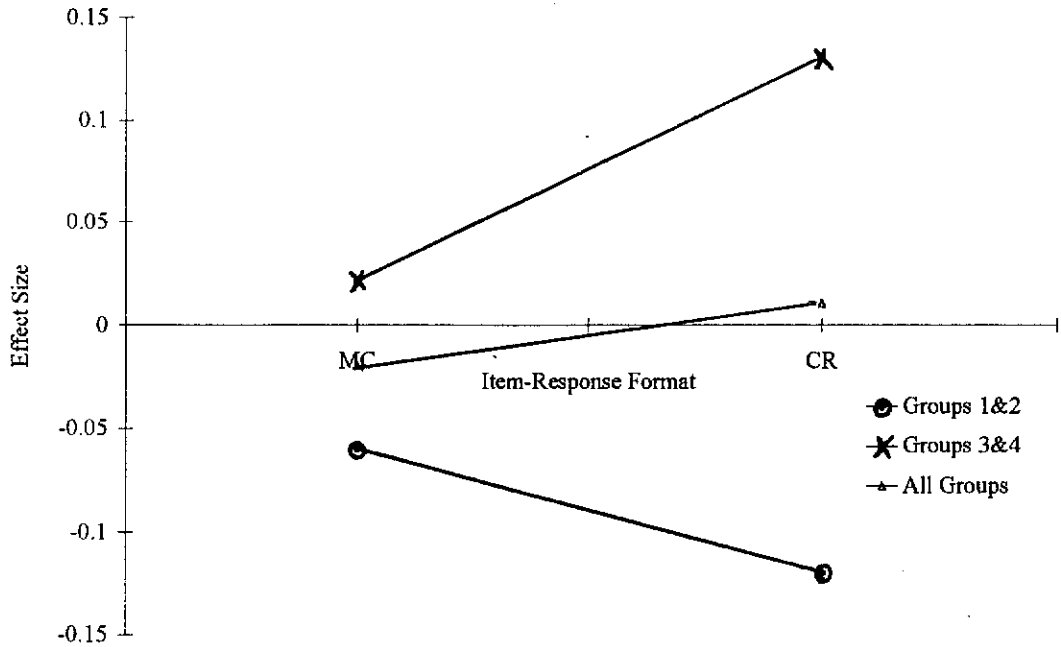


Figure 4.5
Differences Between Girls' and Boys' Achievement for Multiple-Choice and Constructed-Response Items by Research Group

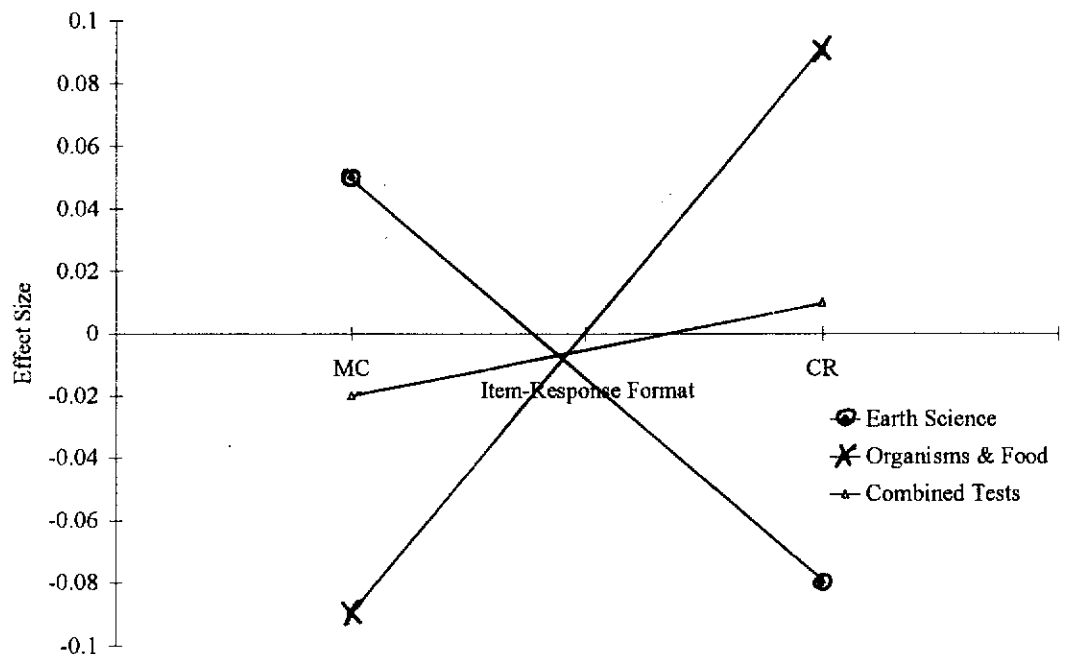


Figure 4.6
Differences Between Boys' and Girls' Achievement for Multiple-Choice and Constructed-Response Items by Unit Topic

Relationship Between Self-Efficacy and Achievement

Influence of Positive Thoughts During Tests

All the boys and almost all the girls (95%) produced responses to Question 3a on the *Survey About Science Tests*, 'Do you think that positive thoughts affect your work during tests?' Similar proportions of the girls (60%) and the boys (59%) reported that their work during tests was influenced by their positive thoughts. Interpretation of these responses is not problematic because the boys' and the girls' response rates to Question 3 were similar.

A quarter of the students (28% of the girls and 21% of the boys) reported that their positive thoughts make them feel good during tests (for example, 'You feel better', 'they boost your morale, make you more enthused'). Only small proportions of the boys (19%) and the girls (11%) indicated that their positive thoughts caused them to try harder ('they push you on', 'get me fired up') and that they helped the girls (12%) and the boys (8%) think more effectively ('You think much more efficiently', 'Helps you focus', 'The questions seem easier'). Although most of the students ascribed a beneficial influence to their positive thoughts, a very small group of seven students indicated that their positive thoughts were a source of distraction for them during tests (for example, 'All thoughts take away my concentration' and 'They make me go off track'). It is important to note in this context that there were no differences in the relative importance of these categories for the boys and the girls. Overall, there is little evidence that the students' positive thoughts have much influence during tests other than to promote a sense of well-being.

Additionally, a small proportion of the girls (29%) and the boys (35%) reported that they are not influenced by positive thoughts in tests. Three categories of responses emerged from the data; that positive thoughts are absent during tests (for example, 'I don't really think positive thoughts during tests'), students are task-focussed ('I don't pay much attention to them [positive thoughts]') and that their positive thoughts have no influence ('they don't help me know any more information', 'They try to help me threw (sic) but my negative thoughts fight them back'). The

small numbers of the boys and the girls responding to each of these categories precluded meaningful analyses of these data.

Effects Associated With a Lack of Confidence in Tests

Almost all of the girls (99%) and the boys (99%) and responded to Question 4 on the *Survey About Science Tests*, 'What types of things happen to you when you lose your confidence during a test?'. Five categories of deleterious responses emerged from this data, namely: interference with thoughts, withdrawal from the testing situation, silly mistakes, bad feelings, and off-task thoughts and behaviours (Table 4.13). Additionally, a final category of responses accommodated the students who reported that they do not experience negative thoughts during tests.

Table 4.13
Students' Responses When They Lose Confidence During Tests

Effects associated with a lack of confidence	Percentage of Students	
	Girls	Boys
Interference with thoughts	46	22
Off-task thoughts and behaviours	24	27
Bad feelings	23	11
Withdrawal	10	21
Silly mistakes	11	3
Not applicable	6	10

Note: Responses are ranked from the most frequent to the least frequent.

One third of the students reported they experience interference with their thoughts when they lack confidence (for example, 'You can't remember answers you would have known', 'I can't think straight' and, 'I cannot seem to focus on my test'), and about one quarter of the students reported the occurrence of off-task thoughts and actions ('I just sit there and play with my pen', 'Your eyes stray and you don't get on with your work', 'worry about the outcome of the exam'). Smaller proportions of the students reported that they feel bad ('Start to feel really bad', 'I become upset'), that they withdraw from the testing situation ('You tend to give up

and stop trying', 'I slow down') and that they make silly mistakes ('do stupid things like circleing (sic) the wrong letters, not reading questions probly (sic)') when they lose their confidence in tests.

Extent of the Association Between Item-Specific Self-Efficacy and Achievement

Quantitative data from the achievement tests were used to calculate the correlations between item-specific self-efficacy and achievement for the boys and the girls on multiple-choice and the constructed-response formats (Table 4.14). All of these correlations were positive, which indicates an association between high levels of self-efficacy and high levels of achievement. This finding is consistent with the conceptual model that was developed in Chapter 2, and this relationship is explored in more detail in the following chapter.

Table 4.14

Relationships Between Item-Specific Self-Efficacy and Achievement

Students	n	<u>MC</u> items		<u>CR</u> items	
		r_{xy}	r_{xy}^2	r_{xy}	r_{xy}^2
<i>Earth Science test</i>					
Groups 1&2	69	+ .26	.07	+ .42	.18
Groups 3&4	61	+ .54	.30	+ .52	.27
<i>Organisms and Food test</i>					
Groups 1&2	66	+ .30	.09	+ .32	.11
Groups 3&4	59	+ .28	.08	+ .43	.19

Note. MC = Multiple-Choice; CR = Constructed-Response.

The extent of the association between self-efficacy and achievement, as indicated by the coefficient of determination, r_{xy}^2 , shows that between 7% and 30% of the variations in achievement were associated with variations in self-efficacy. The mean coefficients of determination for each of the response formats were 14% for the multiple-choice items (17% for the girls and 14% for the boys) and 19% for the constructed-response items (17% for the boys and 22% for the girls). Overall, a mean correlation of +.38 was obtained by averaging the individual correlation coefficients,

which suggests the existence of a moderate relationship (Cohen, 1969) between item-specific self-efficacy and achievement.

Item-Specific Self-Efficacy for Correctly and Incorrectly Answered Items

The magnitude of the gender-related differences in self-efficacy that emerged during the quasi-experimental analyses prompted the subsidiary research question; ‘Are the gender-related differences in self-efficacy dependent on whether the achievement items are answered correctly or incorrectly?’. This research question sought to better understand the origin of these gender-related differences in self-efficacy, and in doing so to provide additional understandings of the relationship between gender, achievement and self-efficacy (see Tables 4.15 & 4.16).

Table 4.15

Boys’ and Girls’ Item-Specific Self-Efficacies for Incorrectly Answered Items

Students	MC items				CR items			
	<u>n</u>	<u>M</u>	<u>SD</u>	<u>d</u>	<u>n</u>	<u>M</u>	<u>SD</u>	<u>d</u>
<i>Earth Science test</i>								
Groups 1&2								
Boys	29	2.81	0.83		27	2.47	0.87	
Girls	40	2.88	0.83	-0.08	35	2.46	0.93	+0.01
Groups 3&4								
Boys	25	2.95	0.63		25	2.96	0.79	
Girls	36	2.50	0.70	+0.65	36	2.58	0.73	+0.48
<i>Organisms and Food test</i>								
Groups 1&2								
Boys	29	2.86	0.77		28	2.38	0.98	
Girls	37	2.57	0.81	+0.36	35	2.16	0.83	+0.24
Groups 3&4								
Boys	24	2.84	0.56		23	2.66	0.83	
Girls	35	2.48	0.73	+0.53	34	2.26	0.76	+0.50

Note. Self-efficacy judgements were made on 5-point scales (1 = very little confidence, 5 = quite a lot of confidence). Positive effect sizes indicate higher self-efficacy scores for boys than girls. MC = Multiple-Choice; CR = Constructed Response; d = effect size.

Table 4.16
Girls' and Boys' Item-Specific Self-Efficacies for Correctly Answered Items

Students	<u>MC</u> items				<u>CR</u> items			
	<u>n</u>	<u>M</u>	<u>SD</u>	<u>d</u>	<u>n</u>	<u>M</u>	<u>SD</u>	<u>d</u>
<i>Earth Science test</i>								
Groups 1&2								
Boys	29	3.62	0.90		25	3.52	0.74	
Girls	40	3.25	0.92	+0.27	37	3.20	0.94	+0.36
Groups 3&4								
Boys	25	3.62	0.88		23	3.78	0.76	
Girls	36	3.09	0.71	+0.65	33	3.26	0.81	+0.63
<i>Organisms and Food test</i>								
Groups 1&2								
Boys	29	3.41	0.69		25	3.32	0.96	
Girls	37	3.13	0.82	+0.37	29	2.94	0.91	+0.41
Groups 3&4								
Boys	24	3.29	0.58		24	3.55	0.89	
Girls	35	2.63	0.70	+0.90	29	2.91	0.82	+0.71

Note. Self-efficacy judgements were made on 5-point scales (1 = very little confidence, 5 = quite a lot of confidence). Positive effect sizes indicate higher self-efficacy scores for boys than girls. MC = Multiple-Choice; CR = Constructed Response; d = effect size.

Correct responses were operationalised as those multiple-choice items where the students scored 1 mark from the maximum score of 1, and those constructed-response items where the students scored either 1.5 or 2 marks from the maximum score of 2 marks. It was necessary to operationalise correct responses on the constructed-response questions in this manner because the numbers of the students scoring the maximum score of 2 marks were insufficient for the analyses. Accordingly, incorrect responses were operationalised as those multiple-choice items where the students scored 0 marks and those constructed-response items where the students scored either 0 or 0.5 marks. The numbers of the boys and the girls involved in these analyses for the correctly answered items were less than the number of the students in the other analyses because not all of the students were able to correctly answer at least one of the constructed-response questions.

The self-efficacy data show that the girls' and the boys' item-specific self-efficacies are higher when they answer achievement items correctly than when they answer achievement items incorrectly (Tables 4.14 & 4.15). Effect sizes for these differences are larger for the boys than the girls, and this is particularly the case for the multiple-choice items ($d = 0.57$ for the girls and $d = 0.91$ for the boys) than the constructed-response items ($d = 0.96$ for the boys and $d = 0.79$ for the girls).

Additionally, the boys reported higher self-efficacies for incorrectly answered items than the girls in all but one instance (Table 4.15), and higher self-efficacies for correctly answered items than the girls in every instance. Overall, the mean effect sizes for differences between the girls' and the boys' self-efficacies when they answered items incorrectly were small for each of the response formats ($d = +0.37$ for multiple-choice items and $d = +0.31$ for constructed-response items), and they were moderately sized when the boys and the girls answered items correctly ($d = +0.55$ for multiple-choice items and $d = +0.53$ for constructed-response items). Hence, the boys reported relatively higher self-efficacies than the girls for multiple-choice and constructed-response achievement items, and particularly when they answered items correctly rather than incorrectly. Although there appears to be a small interaction between gender and whether items are answered correctly or incorrectly, there is clearly no interaction for self-efficacy scores between test item-response formats, gender and the correctness of students' answers.

Relationship Between Self-Efficacy and Worry

Extent of the Association Between Item-Specific Self-Efficacy and Worry

Quantitative data from the achievement tests were used to calculate the correlations between item-specific self-efficacy and worry for the multiple-choice and the constructed-response item formats (see Table 4.17 on page 145). All of these correlations were negative, which indicates an association between high levels of worry and low levels of self-efficacy. Again, this finding is consistent with the conceptual model that was developed in Chapter 2.

The extent of the relationships between worry and self-efficacy, as indicated by the coefficient of determination, r_{xy}^2 , shows that between 2% and 32% of the variation in self-efficacy was associated with variations in worry. The mean coefficients of determination for each of the response formats were 11% for the multiple-choice items and 15% for the constructed-response items. No comparisons were made of the coefficients of determination by gender because the small numbers of the boys and the girls precluded meaningful analyses of these data. Overall, a mean correlation of $-.33$ was obtained by averaging the individual correlation coefficients, which suggests the existence of a moderate relationship (Cohen, 1969) between item-specific self-efficacy and worry.

Table 4.17
Relationships Between Self-Efficacy and Worry

Students	n	MC items		CR items	
		r_{xy}	r_{xy}^2	r_{xy}	r_{xy}^2
<i>Earth Science test</i>					
Group 1	33	-.42	.18	-	-
Group 2	36	-	-	-.39	.15
Group 3	29	-.36	.13	-	-
Group 4	32	-	-	-.30	.09
<i>Organisms and Food test</i>					
Group 1	29	-	-	-.57	.32
Group 2	37	-.21	.05	-	-
Group 3	29	-	-	-.13	.02
Group 4	30	-.24	.06	-	-

Note. Worry data were collected only for the questions on the first part of each achievement test, dashes indicate cells where no data were collected.

MC = Multiple-Choice; CR = Constructed Response.

Relationship Between Worry and Achievement

Influence of Negative Thoughts During Tests

Almost all of the girls (99%) and all of the boys responded to Question 2 on the *Survey About Science Tests*, 'Do you think that negative thoughts affect your work during tests?' A higher proportion of the girls (79%) than the boys (63%) reported that negative thoughts influenced their work during tests. All but one of these students attributed deleterious consequences to their negative thoughts, which were categorised as interruption, bad feelings and withdrawal from the test. One girl, reported to the contrary, however, that negative thoughts have a positive influence on her work because 'they make me work'.

Nearly half of the students (38% of the boys and 54% of the girls) indicated that negative thoughts interrupt their thoughts during tests (for example, 'You can't concentrate because you are too worried about your marks', 'they distract you', and 'They get me confused'). Much smaller proportions of the girls (20%) and the boys (11%) reported that their negative thoughts arouse bad feelings ('I get frustrated', 'They make the weight of the world crash down upon you'), and that negative thoughts cause the boys (11%) and the girls (13%) to withdraw their effort during tests ('You just give up', 'They make me work slower'). Additionally, a higher proportion of the boys (35%) than the girls (16%) reported that they were not influenced by negative thoughts during tests, either because they do not experience negative thoughts (for example, 'I don't really think about it that much', 'I just think about doing the test') or because they can cope with their negative thoughts ('I push them out of my brain', 'I hardly ever take notice of them').

Students' Responses to Their Negative Thoughts During Tests

The review of the literature in Chapter 2 highlighted the importance of the subjective meanings that students attach to their worrisome cognitions during tests rather than the frequency of their worrisome thoughts. Hence, Question 5 on the *Survey About Science Tests* asked students, 'What do you do to help yourself when

you start feeling worried or flustered during a test?’ A very high proportion of the girls (93%) and the boys (90%) produced responses that were categorised as; thought control, off-task thoughts and behaviours and increased effort (Table 4.18).

Table 4.18
Boys’ and Girls’ Responses to Worrisome Situations During Tests

Type of Response	Percentage of Students	
	Girls	Boys
Thought control	55	44
Off-task thoughts and behaviours	21	22
Increased effort	17	8

Note. Response categories are ordered from the most frequent to the least frequent.

Approximately half of the students reported that they engage in some form of thought control when they are flustered in tests. Groups of responses that were subsumed within this category included calming efforts (for example, ‘rub my temples and relax for 2 minutes,’ and ‘I try to calm myself down by breathing slowly’), positive self-talk (‘I say to myself ‘Think positive’, ‘Think positive, tell yourself I can do it. Think about the story the Little Engine that could’), increased concentration (‘I just make myself concentrate (sic) again’) and ignoring negative thoughts (‘Ignore negative thoughts’, ‘I try not to think about it’). Slightly less than a quarter of the students reported off-task thoughts and behaviours such as going on to the next question (‘Go on to the next question and then go back’) and withdrawal from the test situation (‘Pull my pen apart when I am doing bad’, ‘I just sit there’), and a small proportion of the students reported that they increased their effort (‘Just keep going with the test & try to do my best’) when they were worried. Overall, the boys and the girls appeared to respond to their worrisome thoughts in similar ways to each other during tests.

Extent of the Association Between Worry and Achievement

Quantitative data from the achievement tests were used to calculate the correlations between multiple-choice worry and multiple-choice achievement, and

constructed-response worry and constructed-response achievement (see Table 4.19). All of these correlations were negative, which indicates an association between high levels of worry and low achievement. This finding, which is consistent with the conceptual model that was developed in Chapter 2 is examined in more detail in the following chapter.

Table 4.19
Relationships between Worry and Achievement

Research Groups	n	MC items		CR items	
		r_{xy}	r_{xy}^2	r_{xy}	r_{xy}^2
<i>Earth Science test</i>					
Group 1	33	-.62	.39	-	-
Group 2	36	-	-	-.23	.05
Group 3	29	-.05	.00	-	-
Group 4	32	-	-	-.52	.27
<i>Organisms and Food test</i>					
Group 1	29	-	-	-0.45	0.20
Group 2	37	-0.03	0.00	-	-
Group 3	29	-	-	-0.20	0.04
Group 4	30	-0.31	0.10	-	-

Note. Worry data were collected only for the questions on the first part of each achievement test; dashes indicate cells where no worry data were collected. MC = Multiple-Choice; CR = Constructed Response.

The extent of the relationships between worry and achievement, as indicated by the coefficient of determination, r_{xy}^2 , show that between 0% and 39% of the variation in achievement was associated with variations in worry. The mean coefficients of determination for each of the response formats were 12% for the multiple-choice items and 14% for the constructed-response items. No comparisons were made of the correlations between achievement and worry, and the corresponding coefficients of determination by gender because the low numbers of the girls and the boys precluded meaningful analyses of these data. Overall, a mean correlation of -.30 was obtained by averaging the individual correlation coefficients, which suggests the existence of a moderate relationship (Cohen, 1969) between worry and achievement.

Determinants of Future Achievement

A large majority of the boys (75%) and the girls (82%) responded to Question 10 on the *Survey About Science Tests*, 'What would have to change before you could achieve a higher mark on your next science test?'. Any additional comments that the students made at the end of the survey form, which were relevant to this question were included in these analyses. Five categories of responses, namely; effort, attitude, study skills, tests and classroom processes emerged from analyses of the girls' and the boys' responses. Table 4.20 shows the proportions of students responding to each category and selected groups within these categories.

Table 4.20

Types of Changes That are Required to Improve Girls' and Boys' Achievements

Types of Changes	Percentage of Students	
	Girls	Boys
Internal Changes	55	52
• Effort	40	46
• Attitude	9	8
• Study Skills	10	5
External Changes	45	32
• Testing Situation	27	22
• Classroom Processes	26	14
- Class learning	18	8
- Teacher	8	7

Note: Response categories within the internal and external divisions are ranked from the most frequent to the least frequent.

Approximately one half of the students expressed thoughts about changes that are within their control, which are denoted as internal changes in Table 4.20. A little less than half of all the students listed changes relating to expending more effort in class and in study (for example, 'more concentration in class' and 'study harder'). Only a very small proportion of the students mentioned improved attitudes ('my attitude towards the subject') and better study skills ('my study habits').

Approximately a third of the students expressed thoughts about changes that lay

beyond their control, which are denoted as external changes in Table 4.20. About a quarter of the students reported changes to tests relating to time ('more time'), unit size ('have a few small tests rather than one large one'), test formats ('no written questions on the test') and difficulty of tests ('make the test easier'). Slightly less than a quarter of the students mentioned changes relating to classroom processes such as class learning ('we need to cover much more of the work in class', 'more revision sheets or time in class before the test') and the teacher ('If I had a good teacher who didn't get angry when I asked questions when we learn the stuff for the test').

Similar proportions of the boys (52%) and the girls (53%) reported that their improved achievement is contingent on them making internal changes. For example, a little less than half of the girls (40%) and the boys (46%) reported that their achievement would change if they made a greater effort. Much smaller, and similar proportions of the boys (8%) and the girls (9%) reported that an improvement in their achievement was dependent on possessing better attitudes and study skills. It is important to note that there were no differences in the relative importance of any of these internal changes for the girls and the boys.

A higher proportion of the girls (45%) than the boys (32%) reported that their improved achievement is contingent on them making external changes. For example, a slightly higher proportion of the girls (27%) than the boys (22%) suggested changes relating to tests. Additionally, a much higher proportion of the girls (26%) than the boys (14%) suggested changes related to classroom processes. Classroom processes, and particularly aspects of classroom processes related to learning in class appeared to be relatively more important for the girls than the boys in this study.

It is noteworthy that all but one of the responses that the students made about their teachers related to two male science teachers who taught the students in Research Groups 3 and 4. Typical comments from the boys and the girls included, 'Teacher's attitude towards the students', 'If I had a good teacher who didn't get angry when I asked questions', 'I get a dirty look or he rolls his eyes. I never understand what he is talking about,' and 'If the teacher didn't have conversations with only 5 or 6 people, instead of the class'. In all, 1% of the students in Research Groups 1 and 2, and 14% of the students in Research Groups 3 and 4 made unsolicited negative comments about their science teacher.

SUMMARY OF THE FINDINGS

This chapter has completed the 'coarse-grained' analyses of the data that were collected for all the 154 girls and boys who participated in this study. Analyses of the quasi-experimental data have shown the absence of any practically important gender differences for science achievement on multiple-choice and constructed-response formats, and the absence of any interactions between gender and test item-response formats for achievement. However, the boys demonstrated more self-efficacy and less worry than the girls on both the multiple-choice and the constructed-response formats. Additional analyses revealed that the gender-related differences in self-efficacy were slightly larger for questions that were answered correctly than for questions that were answered incorrectly, and that the extent of the gender-related differences in self-efficacy was not consistent across the research groups. No practically significant interactions between gender and item-response format were detected for either self-efficacy or worry in this study.

Analyses of the survey data show that the boys and the girls experienced similar types of negative and positive thoughts as each other during tests, and that there were no gender-related differences in the types of thoughts that contribute toward girls' and boys' generalised self-efficacies and their worrisome thoughts on multiple-choice and constructed-response questions. However, these analyses indicated that the girls were relatively more self-focussed than the boys during tests — the girls were more likely to direct their negative and positive thoughts at themselves and their capabilities whereas the boys tended to direct their negative and positive thoughts at the test, their progress on the test and the consequences of their achievement. Additionally, test-wise strategies were relatively more important for the girls' self-efficacies than the boys' self-efficacies on multiple-choice questions, and perceptions about the greater rigour of constructed-response items contributed relatively more toward the girls' perceptions of inefficacy than the boys' perceptions of inefficacy. Moreover, that constructed-response items provide a greater opportunity for students to demonstrate their learning was relatively more important for the boys' self-efficacies than for the girls' self-efficacies.

Additionally, the conceptual model underpinning this study was supported by moderate correlations between item-specific self-efficacy, worry and achievement, together with students' reports from the survey data. These relationships and other relationships are elaborated further in Chapter 5.

CHAPTER 5

ANALYSIS OF THE DATA ... II

OVERVIEW

This chapter presents the 'fine-grained' analyses of the interpretivist-constructivist data that were collected from the 22 students who were interviewed. These analyses sought to clarify and extend the theory relating to generalised self-efficacy, worry and achievement on multiple-choice and constructed-response test formats through a study of boys' and girls' meaning-perspectives. The interpretative data were analysed with reference to the wider context of the coarse-grained analyses and the extant literature (as reviewed in Chapter 2).

The first section of this chapter describes the procedures, which were informed by F. Erickson's work (1986), to generate and report assertions as patterns of generalisation within the interpretative data; the second section examines the content of girls' and boys' perceptions of self-efficacy and their worrisome cognitions during tests; and the final section presents the analyses relating to the conceptual model that was developed to better understand boys' and girls' achievements according to test item-response formats. The analyses in these last two sections are framed by interpretative commentary that starts with a series of postulates and ends with a consideration of the insights that were yielded by the interpretative data.

CONTEXT FOR THE ANALYSES

Analytical Procedures

Procedures for Analysing the Interview Transcripts

The interview data were analysed according to perspectives from the ethnographic and the interpretative research traditions, after the analyses of the quasi-experimental were complete, by drawing on the work of F. Erickson (1986). The 22 interviews and the 9 follow-up interviews were read several times and provisional

codes were assigned as they emerged from readings of the transcripts. The selection of these provisional codes was partly informed by the preliminary analyses of the interviews, and most likely, they were unconsciously influenced by categories from the coarse-grained analyses. Individual 'chunks' (Miles & Huberman, 1994, p.56) of information, which generally consisted of a phrase or a combination of several phrases, were coded. Then there followed an iterative process of redefining the coding categories and reassigning the chunks of information to progressively broader categories. This process resulted in the formation of 'meta-clusters' (Miles & Huberman, 1994), which subsumed as much of the interview data as possible into independent categories (Merriam, 1988, p.136). According to these procedures, single chunks of information were assigned to as many of the meta-clusters as were appropriate. It is important to note in this context that neither the provisional codes nor the final categories were defined *a priori* by the researcher, but rather, they were allowed to emerge from the data. It was at this point that a tension existed between the focus provided by the original research questions and the emergent nature of the interviews and the analyses. Although there was some correspondence between the meta-clusters and the original research questions, some of the meta-clusters were unrelated to the original purposes of this study. This chapter reports the analyses that were directly related to the original research questions, together with analyses that have heuristic value for the purposes of this research.

Inferential links (that is, assertions) were written to describe the relationships between these meta-clusters. The adequacy of these assertions, as patterns of generalisation within the database, was tested by searching for confirming and disconfirming evidence. Interview data, survey data and quasi-experimental data were consulted for discrepant cases, either to explicate these discrepant cases or to amend the assertions to accommodate the disconfirming data. Hence, the analyses were inductive in the sense that the assertions were allowed to emerge from the data, and deductive in the sense that the assertions were tested against all the research data.

Rigour of the Data Analyses

The fifth of F. Erickson's (1986) criteria for making warranted assertions

(p.140) was addressed during the data analyses to avoid charges of researcher bias and evidentiary inadequacy. This criterion, which relates to the analysis of discrepant cases, is important because it helps adjust the assertions so that they are more consistent with the data base. It is important to recall in this context that students were purposely selected for interviewing to increase the likelihood of obtaining disconfirming data. Moreover, this study continuously sought both confirming and disconfirming data by testing the emerging assertions against the database. Additionally, given that researchers are 'far more likely to see *confirming* instances of original beliefs and perceptions than to see disconfirming instances' (Miles & Huberman, 1994, p.263, italics in the original) the researcher checked the plausibility of his assertions, and also for the presence of alternative explanatory frameworks, with one of his teaching colleagues. Additionally, as described in Chapter 3, this researcher kept a journal that monitored his progressive subjectivity (Guba & Lincoln, 1989) through records of his *a priori* constructions and his emerging findings during the interviewing cycles. One limitation of these analyses, however, is that the final assertions were not taken back to the students for member checks, as they were for the findings that emerged from the first interviewing cycle.

Procedures for Reporting Findings from the Interview Data

The researcher's task, when reporting assertions, is to 'persuade the audience that an evidentiary warrant exists for the assertions made, [and] that patterns of generalization within the data set are indeed as the researcher claims they are' (F. Erickson, 1986, p.149). Therefore, F. Erickson recommends that sufficient detail is provided so that readers can co-analyse cases and judge the plausibility of the analyses for themselves. In other words, readers of qualitative analyses must be able to experience vicariously the research setting and encounter key analytical constructs; to survey the full range of evidence that underpin the analyses; and, to map changes in the researcher's perspective during the analyses. Therefore, the fine-grained analyses of the interview data contain the following elements that are described by F. Erickson (1986); (i) 'particular description', which involves descriptions of individual students and quotes from the raw data; (ii) 'general description', which describes the

occurrence of patterns within the data corpus by indicating the frequency of particular descriptions; and (iii) 'interpretative commentary', which documents the presuppositions of the researcher, from his journal, before the final analyses of the interview data and the changes in his perspective over the course of the analyses. As mentioned in Chapter 1, the journal that was kept during this inquiry enhanced this researcher's capacity for reflection by recording his ongoing analyses of the interpretive data.

It is important to recognise that the interview data were analysed after the coarse-grained data analyses were complete. Hence, the starting point for the fine-grained data analyses included presentiments from the preliminary analyses of the interview data as well as perspectives from the coarse-grained analyses and the review of the literature. It is also important to note that students were asked in the interviews about their confidence in tests rather than the more precisely defined constructs of generalised self-efficacy and achievement expectancy, which were likely to have little meaning for the students. Nevertheless, it was possible in many of the cases to judge from the context of the responses whether students were describing aspects of their generalised self-efficacies or aspects relating to their achievement expectancies.

THE CONTENT OF BOYS' AND GIRLS' WORRISOME THOUGHTS AND THEIR SELF-EFFICACIES DURING TESTS

Research Questions Addressed

This section presents findings relating to the first two research questions for the 22 students who were interviewed, that is,

- Are there gender-related differences in the content and extent of students' self-efficacies for answering science achievement questions set in multiple-choice and constructed-response formats? And consequently: is there a gender-related interaction between girls' and boys' perceptions of self-efficacy on science achievement questions and the response format these questions require?
- Are there gender-related differences in the content and extent of students'

worrisome cognitions for answering science achievement questions set in multiple-choice and constructed-response formats? And consequently: is there a gender-related interaction between boys' and the girls' worrisome cognitions on science achievement questions and the response format these questions require?

Assertion 1

Students' confidences before tests are determined by their perceptions about the adequacy of their preparation. Although teachers have an influence on students' self-efficacies, this influence is mediated by students' perceptions about the adequacy of their preparation.

Preliminary Comments

When students were asked, on the *Survey About Science Tests* about the changes that needed to take place before they could achieve a higher mark on their next test, a little less half of the students reported that they need to do more study (Table 4.19). Additionally, factors relating to the test and classroom processes were less important for students' future achievements than their preparations for tests. In this context, it is important to recall that skills deficit models for test anxiety posit that test anxiety is caused by an awareness of inadequate preparation for tests and inadequate test-taking skills (Tobias, 1985), and that worry is inversely related to self-efficacy.

Additionally, the coarse-grained analyses revealed the existence of small gender-related differences in self-efficacy for the students in Research Groups 1 and 2 and much larger gender-related differences in self-efficacy for the students in Research Groups 3 and 4. Furthermore, the most striking difference between the students in Research Groups 3 and 4 when they were compared to the students in Research Groups 1 and 2, apart from the magnitude of the gender-related differences in self-efficacy, concerned the students' perceptions about their science teachers. In this context, a higher proportion of the students (14%) from Research Groups 3 and 4

expressed dissatisfaction with their science teacher on the *Survey About Science Tests* than the students (1%) from Research Groups 1 and 2. Moreover, almost all of the students in Research Groups 3 and 4 were taught by two male science teachers and almost all the students in Research Groups 1 & 2 were taught by two female science teachers.

Hence, it was postulated that students' self-efficacies before tests are related to their perceptions about the adequacy of test preparations and also to teacher-related variables, except that the influence of these teacher-related variables (including the gender of the science teacher) was speculated to be more potent for the girls than the boys.

Analysis of the Interview Data

The interview responses indicated that the students' confidences before tests, (that is, their self-efficacies and their achievement expectancies) were chiefly related to their perceptions about the adequacy of their preparation. Only after students volunteered this information did they mention sundry aspects that appeared to have less bearing on their confidences. Additionally, none of the students volunteered any information relating to either their teachers or the format of test items.

Preparation for Tests

All 22 students (10 boys and 12 girls) reported that their confidence before tests depends mainly on their perceptions about the adequacy of their preparation. These students reported that their self-efficacies are high when they know and understand their work. Typically, most of these students identified knowledge about their performances from revision activities and quizzes as important for shaping their self-efficacies;

- S. It (confidence) really depends on how much I know and how much I don't know. (Joshua)*
- S. When you've gone through that (the sheet of objectives) and you can do everything then you feel more confident. (Kevin)*
- S. If I do like quizzes with other friends and I get them all right, [and I] got through revision sheets and I know that I'm getting the answers right [I'm confident]. (Lucy)*

Although Rebecca commented, 'I've been good at science all my life and I know that I'll always do well,' she still reported the importance of preparing for tests. Moreover, students generally reported low self-efficacies when they perceived that they lacked either knowledge or understanding of their work.

S. I know I [am not] sure of some things. (Joshua)

S. You don't know anything. (Madeline)

S. You are not very good at the section [of work] that you are doing. (Sarah)

These depressed self-efficacies that arise from not knowing or understanding the work, together with the elevated self-efficacies that accompany students' knowledge and understanding of their work underscored the importance of students' preparations for tests.

Study and revision were identified by almost all of the students (10 girls and 9 boys) as the most important of their preparations for tests. Typically, these students identified the importance of starting their study and revision well before the test and the thoroughness of their preparation as factors that contribute toward their self-efficacy.

S. Instead of leaving everything to the last minute you should study. Like I study about every two weeks. I go over what we've done and try and remember [it] rather than leaving it all to the last minute. It makes you more confident. (Danielle)

S. You've gone through your book and you've done everything you think you possibly can, like the sheet [of objectives]. (Kevin)

S. If you studied well and you know that you know it, it's easy to be confident. But if you haven't studied then there's nothing to be confident about. (Madeline)

As Madeline indicated, these students associated their low self-efficacy with their lack of study. Additionally, other students reported:

S. [I'm not confident when] I haven't studied and I feel that I don't understand the work and I think that I'm going to fail. I'm just not confident with myself. (Danielle)

S. If I haven't looked at any of the revision beforehand, if I've just gone straight into the test without really knowing what is in front of me [then I lack confidence]. (Matthew)

Students' preparations in their classes leading up to the time of the test were much less important to the students' confidences than study and revision. Eight students (5 boys and 3 girls) reported that their self-efficacies depended on working in class and listening to their teacher.

S. You've worked hard, listened in class and you've done some studying and you feel that you know it all so you're feeling confident. (Michael)

- S. *If you've listened well and written everything down and studied over the term you'd feel pretty confident. (Cameron)*
- S. *[I am confident when] I've worked and I've done what I'm meant to do and so I know that I should be right for it (the test). If I've done what I'm told to [do] I should know everything that's needed to be known. (Rebecca)*

Conversely, when students had not been attentive in class or when they had 'mucked about a fair bit,' their self-efficacies were depressed because they perceived that they did not know their work.

Teacher-Related Aspects of Students' Self-Efficacies

In all, 17 students (10 girls and 7 boys) reported that their confidence was influenced by their teacher. Several of these students indicated that their achievement expectancies were dependent on their relationship with their teacher.

- S. *If you don't like your teacher you're thinking, 'Oh she's going to mark me bad'. Things like that [make me feel] less confident. (Wade)*
- S. *If the teacher likes you then you think you get a few extra, not heaps extra, but a few extra marks. (Madeline)*

Otherwise, the remainder of the students linked teacher-related variables to the quality of their learning, and consequently to their self-efficacy before tests. However, four of the students (3 boys and 1 girl) indicated that their confidence before tests was unrelated to their teacher. These students reported that test answers were not marked as subjectively as other forms of assessment, and that 'confidence comes from yourself and not from your teachers.'

The influence that teachers have on students' confidences in tests appeared to be mediated by students' perceptions about the adequacy of their test preparation. Fifteen of the students (8 girls and 7 boys) identified the importance of focussed learning activities, frequent teacher input, guidance and feedback, together with general aspects of teacher competence as important for their confidences.

- S. *If you think that the teacher has prepared you well for it (the test) then you'd be more confident. If you know that you haven't covered as much as you should have then you'd be worried. (Brendan)*
- S. *Some teachers can't control the class as much as some other teachers can, so you don't learn as much, so you don't get much done. (Angela)*
- S. *With teachers that set a lot of homework and things like that you have to learn things yourself it doesn't make me as confident because I'm not sure if I have it or know it right or things like that. (Wade)*

Additionally, 11 of the students (7 boys and 4 girls) identified the ability of the teacher to form positive relationships within the class as important for their self-efficacies.

- S. *If you don't like the teacher you just don't participate that much and you don't put your hand up. You figure, 'why should I bother because I don't like him. Why bother?' But if I do like [the teacher] you tend to put your hand up and ask [and that affects my confidence] because the teacher has actually explained it. (Sophie)*
- S. *I just like being friends with teachers, sort of having some bond I suppose, just so you can learn more things, like they encourage you more. (Kevin)*

These students perceived that they learned more when they had 'some bond' with their teachers, and consequently, they appeared to report higher confidences because they perceived that they were better prepared for their tests.

However, the teachers' gender appeared to have no bearing on students' confidences for tests. All but one of the 15 students who indicated that teachers have an influence on their confidence reported no differences associated with the gender of their science teacher. Matthew, who was the only student to report that he learned better from male teachers commented that 'female teachers don't seem as forceful and they let kids get away with much more [misbehaviour]'. Otherwise, the remainder of the students regarded the gender of their teachers as unimportant for their confidences.

- S. *As long as the teacher does what they are supposed to do and gives all the information then it doesn't make a difference [if they are male or female]. (Michael)*
- S. *You still learn the same work and everyone does the same sort of stuff. (Lisa)*
- S. *It doesn't make a difference to me because I've got Mr. Mason and he's the nicest teacher, but like he's really nice and like I've got heaps of nice female teachers too. But then I've got really bad female teachers and really bad male teachers. (Sophie)*

Concluding Comments

The fine-grained analyses of the interview data indicated that the students' confidences before tests were determined mainly by their perceptions about the adequacy of their preparation. Although the students responded to questions about their science teacher and the format of test items when they were asked about these factors, they perceived that these variables had much less influence on their confidences than their preparation for tests. Furthermore, although teachers were perceived to have some influence on students' self-efficacies, this influence appeared to be mediated through students' perceptions about the adequacy of their preparation.

No evidence was detected, either to support or counter the postulate that girls' and boys' self-efficacies were differentially influenced by their perceptions about their science teacher.

Assertion 2

Boys and girls give similar reasons to each other for their format-related confidences and their format-related worries.

Preliminary comments

The review of the literature indicated that gender-related differences in achievement on test item-response formats are most likely due to the consistent and simultaneous operation of all-pervasive construct-irrelevant factors (Mazzeo et al., 1993). Moreover, self-efficacy and worry were identified in Chapter 2 as factors that could help develop a better understanding of girls' and boys' achievements for multiple-choice and constructed-response items. Accordingly, the content of the boys' and the girls' format-related worries and their format-related self-efficacies were inferred from their responses on the *Survey About Science Tests*. These coarse-grained analyses, for each of the test formats, showed that the girls' and the boys' self-efficacies were influenced by the same factors, although there was some variation in the relative importance of these factors for the boys and the girls. Consequently, it was postulated that the girls and the boys would give similar reasons as each other for their confidences and their worrisome thoughts on multiple-choice and constructed-response test formats.

Analysis of the Interview Data

All 22 students (10 boys and 12 girls) were asked about which of the test item-response formats they were more confident answering, the multiple-choice or the constructed-response format. In most instances, students responded either by describing aspects of one format they regarded as easy or by describing aspects of the

other format they regarded as difficult. Additionally, students in the first interviewing cycle were asked about worry and test item-response formats. However, because the content of the students' format-related worries was the same as their perceptions of inefficacy for each of the response formats, students were not asked about the content of their format-related worries in the second interviewing cycle.

Multiple-Choice Questions

All the students except for one girl described at least one aspect of their confidences relating to multiple-choice test items. The provision of the answer to multiple-choice questions, embedded within the text of the question appeared to be the chief source of these students' confidences. Only one student expressed an alternative viewpoint, that she has a greater facility for answering multiple-choice items because they are similar to the questions on revision sheets.

In all, 15 students (8 girls and 7 boys) reported that they perceived multiple-choice items to be less rigorous than constructed-response items. Typically, these students expressed the view that multiple-choice questions require less thought than constructed-response questions, and that they can use elimination strategies and guessing to arrive at their answers.

- S. With multiple-choice you don't have to think much. (Madeline)*
- S. Multiple-choice are a real lot easier because you don't really have to think hard. You have to pick which one sounds the best and usually you know it straight away. Its just easier. (David)*
- S. I look at each answer, and like if its totally irrelevant and stupid I automatically cross through that one in my mind but most of them, sometimes they sound right and I usually choose the best one. (Kevin)*
- S. If you don't know the question its also easier to guess with the multiple-choice question ones than [with] the other ones. (Lucy)*

Consequently, 5 boys and 7 girls reported higher achievement expectancies for multiple-choice items than constructed-response items, particularly when they do not know the answer.

- S. You have a 1 in 4 chance of doing it and you know that like 'I've got a 25% chance'. (Sophie)*
- S. There is less of a chance of getting them (multiple-choice items) wrong because there is only four possible answers, while on the others there can be millions depending on how well you've studied for them. (Brendan)*

Of less importance, were reports by 4 of the students (2 girls and 2 boys) that they have higher self-efficacy for multiple-choice items than constructed-response items

because ‘you don’t have to write out the answers properly’ and ‘you [can] get the test done quick’.

In all, eight students (6 boys and 2 girls) reported aspects of multiple-choice questions that are associated with their depressed self-efficacies and their lowered achievement expectancies. All but two of these students reported that distinguishing the correct answer from the distractors is problematic.

S. Sometimes they are really similar and you get confused and you can circle the wrong one easily. (Danielle)

S. You’ve got so many choices that it is sometimes hard if you are an indecisive person. (Matthew)

Otherwise, two students expressed the views that multiple-choice items are ‘difficult to understand’ and that they are ‘too hard this way’. Unfortunately, these students were not asked to elaborate on their responses to identify the sources of their difficulty on multiple-choice questions.

Constructed-Response Questions

Almost all of the students (11 girls and 9 boys) described at least one aspect of their confidences relating to constructed-response items. These students reported that having to generate and write out answers to constructed-response questions, rather than select the answer from the distractors is the chief source of their high and low confidences.

In all, 12 students (4 boys and 8 girls) perceived that constructed-response questions are more rigorous than multiple-choice items because they (that is, the students themselves) have to generate their answers. Typically, these students reported that they have to know their work and think more about constructed-response items than multiple-choice items, and that they can neither guess nor work out their answers from the text of the item.

S. With the writing ones you really have to learn your work really well to know how to answer the questions. (Julia)

S. Written answers are a bit harder because you don’t have anything to choose from. You have to know the work so you can write it down. It has to come from your own mind and not from the test. (Anne)

S. With written questions you have to actually think harder about what the answer is going to be. (Angela)

Moreover, six of these students (2 girls and 4 boys) reported that constructed-response items are more rigorous and time consuming than multiple-choice items because constructed-response items require carefully worded and structured answers.

S. Its harder to write down your thoughts than to just circle things that look right. (Wade)

S. You've got to structure the explanation or your message won't get across clearly enough what your answer is. (Sarah)

Additionally, six of the students (3 boys and 3 girls) who reported that they were confident answering constructed-response questions identified the opportunities to write answers they understand and to provide their opinions as the sources of their confidences on this response format. Moreover, several of these students reported that writing an answer is easier than resolving the ambiguity among multiple-choice distractors.

S. [With] the written answer ones you can say your own thing. But with the multiple-choice you have to answer with one of the ones, even if you don't agree with any of them. (Melanie)

S. You know what you are writing. You just write out the answer you think. (Cameron)

Additionally, four (3 girls and 1 boy) of these six students reported that constructed-response items provide them with the opportunity to 'show off' their understanding about a topic, even if their understanding is not directly relevant to the question:

S. With written answers you can express yourself and you can like sort of show off a bit I guess and show that you have learnt this and you know it and you can write out a full answer. (Sophie)

S. You can talk about things that you know and your knowledge, things that are relevant to the [wider] topic that may not be right, but when the marker looks at it they say, 'They've had a go. It's nearly there', so they will give you a mark. (Matthew)

Otherwise, several students reported higher achievement expectancies for constructed-response items because credit is given for partial knowledge, unlike multiple-choice items where answers are scored as either correct or incorrect.

Concluding Comments

The fine-grained analyses of the interview data provided additional evidence that boys and girls give similar reasons to each other for their format-related confidences and their format-related worries. In relation, explicitly, to the initial postulate, this analysis of the fine-grained data added no new insights.

Assertion 3

Girls' and boys' confidences during tests are influenced by the difficulty of test items and by ongoing self-appraisals of their progress. Testwise strategies are used by most students to regulate the influence of their affective states rather than to increase their confidences during tests.

Preliminary Comments

The previous assertions have described factors that contribute toward boys' and girls' confidences before tests and factors that are related to girls' and boys' confidences on multiple-choice and constructed-response test formats. This present assertion is concerned with factors that influence students' confidences while they are working during tests.

The analyses of the quasi-experimental data revealed moderately sized differences in self-efficacy between achievement items that were answered correctly and achievement items that were answered incorrectly. This finding that self-efficacy is inversely related to task difficulty is consistent with self-efficacy theory (Bandura, 1977, 1986). Moreover, the students also reported the presence of negative thoughts relating to their progress on tests for Question 1 on the *Survey About Science Tests* (Table 4.4). However, a small proportion of the students reported on the *Survey About Science Tests* that their negative thoughts did not influence them during tests because they had ways of coping with their deleterious thoughts. When all of the students were asked what they do when they started feeling worried or flustered during tests they described ways of controlling their negative thoughts, going onto the next question and increasing their persistence (Table 4.17). Additionally, the girls reported self-talk as the most frequent of their positive thoughts during tests (Table 4.3). Therefore, the data from the *Survey About Science Tests* seemed to indicate that the students had ways of managing their worrisome thoughts and their lack of confidence in tests.

Hence, it was initially postulated that students' on-going confidences in tests would be determined by the difficulty of the test items and perceptions about their

progress on tests, and that their confidences would be influenced by the use of test-wise strategies.

Analysis of the Interview Data

All 22 of the students provided evidence that their confidences (that is, their self-efficacies and their achievement expectancies) fluctuate during tests according to the difficulty of the items they encounter and the on-going judgements they make about their progress. Additionally, two students mentioned that their confidences are also influenced by interruptions during the test, such as teachers 'just staring at you and watching every single thing that you do.' None of the students volunteered that their confidences are influenced by test-wise strategies. However, the students reported using self-talk, positive thinking and leaving difficult questions until the end of the test when they were directly asked about the strategies they use during tests.

Confidence and the Difficulty of Tasks

In all, 17 of the students (9 boys and 8 girls) reported that their confidences (that is, their self-efficacies and their achievement expectancies) are related to the difficulty of the questions. These students indicated that their confidences are elevated when they perceive that test items are easy:

- S. If you are low on confidence on one question and the next question is easy you get your confidence back. (Melanie)*
- S. You [are confident when you] know all the questions, you know the answers and you go, 'Oh, this is easy, keep going'. (Madeline)*

Conversely, these students' reported depressed confidences when:

- S. The questions are really hard and you don't sort of understand them as well as you should. (Bernard)*
- S. I haven't got a clue what to do. (Julia).*
- S. You don't remember what to on those questions [so] your level of confidence drops. (Nicholas)*

Although this finding may seem trivial, it is important to note that it lends support to the analyses of the survey data, where inferences were made about the content of the girls' and the boys' format-related self-efficacies by examining aspects of multiple-choice and constructed-response questions that they judged as easy and as hard.

Confidence and Perceptions About On-Going Progress

Students' confidences (that is, their self-efficacies and their achievement expectancies) were reported to increase when the students perceived that they were making good progress on tests, and they were reported to decrease when the students perceived that they were making poor progress. In all, 19 of the students (9 boys and 10 girls) provided responses that were consistent with this conceptualisation, and no students provided any evidence to the contrary. Typically, these students reported that their confidences increase when:

- S. *I'm answering all the questions and I know that I've a fair chance of getting them all right. (Angela)*
- S. *You've done a couple of questions and you think you've got them right [and] the next couple of questions are pretty easy too you think. It can help you out because your confidence goes really higher and you think that you'll get a good mark. (Bernard)*
- S. *I like to keep an eye on the time and if I'm like halfway through and I've had only 15 minutes or something, that gives me more confidence and it helps me to relax a bit more. (Lucy)*

Conversely, these students reported that their confidences decrease when:

- S. *You are doing really bad on the test. (Joshua)*
- S. *You get a couple wrong and then you start thinking to yourself, 'I'm going to keep getting all these wrong now'. (Sarah)*
- S. *If I've got lots of questions left when I don't have much time. (Anne)*

It is important to note in this context that the students judge whether they are making favourable progress on tests according to the likelihood that their answers are correct and whether they think they have sufficient time to complete the test.

Self-Talk and Positive Thinking

Thirteen of the students (9 girls and 4 boys) reported that they use either self-talk or positive thinking to deal with their lack of confidence during tests. Generally, these strategies appeared to be used by students to manage their nervousness and worrisome thoughts, rather than to increase their confidences. It is instructive to note that ten of these students reported, on the *Survey About Science Tests*, that their work during tests was influenced by their positive thoughts.

Positive thoughts and self-talk appeared to be used by these students to exert a calming influence during tests and to counter their nervousness and worry. Typically, students told themselves to 'calm down,' that they have studied for the test and that they can do the work.

- S. *If I am worrying I say to myself consciously to stop worrying and just find out what the mark will be after the test and not worry about it while you are doing it. (Melanie)*
- S. *[I] just try and relax. If you get in there and you are all stressed and stuff you worry heaps. If you just go in there relaxed, thinking it over, knowing that you've done the work that will get you through the test easier. (Matthew).*
- S. *Just saying [to myself] that, 'I've studied for this. I know it. I'm Okay,' and you just sort of think to yourself what you read last night and this morning [that helps me feel more confident]. (Sophie).*
- S. *Tell yourself that you can do it. (Brendan)*

Although several students reported self-messages such as 'I have studied as hard as I can and there is nothing more I can do,' these types of thoughts were interpreted to represent expressions of their confidence rather than their resignation to a particular outcome. Additionally, several students reported that they concoct pleasing, task-irrelevant thoughts to promote a sense of well-being, or at least a sense of optimism for the future.

- S. *Well, if I get a negative thought I try to get a positive thought just to try to make it even, and when I go in [to the test] I think positive thoughts so I feel good and I don't have to worry. (Anne).*
- S. *[I focus on the good points] like just getting it back. If you don't do so good you've always got a chance to make it up. (Nicholas).*

Although these students provided little evidence about the efficacy of their positive thoughts and self-talk, it is clear that self-talk is not regarded by these students as a magical incantation that guarantees success.

- S. *You can't just say, 'Oh, I can do it.' You've got to think you can do it and when you think you can do it, it gets rid of them (negative thoughts). (Brendan)*
- S. *[Self talk doesn't work] because you know its not true. (Madeline).*

Instead, Madeline reported that students must 'meaningfully' believe their self-talk.

Working on the Easiest Questions First

Almost all of the students (8 boys and 10 girls) reported the behavioural strategy of skipping difficult questions and doing the easiest questions first. However, Anne expressed the view that she preferred starting tests with the most difficult questions '[So] you can get that over and done with and you can have the easy bit at the end'. Anne was one of the highest achieving girls and she reported relatively high levels of self-efficacy for achievement during tests and for controlling her thoughts on the *Survey About Science Tests*. On the other hand, Michael indicated that he

preferred to work through tests in the order that items were presented 'so I don't leave anything out.'

Nevertheless, starting with the easiest questions and skipping difficult questions was perceived by most students to have a facilitative effect. Typically, students reported their beliefs that this strategy 'warms up the brain', helps overcome nervousness and that it contributes toward efficacious thoughts.

- S. I go onto the next one because if you get the next one right it sort of takes the worry away from that one. (Joshua).*
- S. I'm more nervous at first and I don't do the questions as well as I could do and then towards the end I just get used to it so its fine. (Sarah).*
- S. It gets you a lot more confident (self-efficacy). You think the test is pretty easy and then when you get to the harder ones you try real hard. (David).*

However, the strategy of skipping questions and leaving them until later presented difficulties for some of the students. Five of the students (4 girls and 1 boy) reported that they still worried about the missed questions and about whether they will have sufficient time to go back and complete them.

- S. I do skip it and go onto the next one [but I] am still thinking about it in the back of my brain and it just confuses me. (Danielle).*
- S. If you skip questions you're thinking, 'Oh no, do I have enough time to go back?', and things like that. (Wade)*

These 5 students indicated on the *Survey About Science Tests* that they were influenced by negative thoughts during tests and they also reported relatively high occurrences of worrisome thoughts on the *Attitude Towards Tests Survey*.

Concluding Comments

The fine-grained analyses of the interview data indicated that during tests, the boys' and the girls' confidences were influenced by the difficulty of test items and ongoing self-appraisals of their progress. Moreover, almost all the students used strategies to regulate the influence of their affective responses, rather than to increase their confidences *per se*. The new insights provided by these analyses relate to the students' use of test-wise strategies. Although most students reported using test-wise strategies, it seems that they used these strategies (with varying degrees of success) to manage their affective responses rather than to increase their confidences as postulated by this researcher.

THE RELATIONSHIPS BETWEEN SELF-EFFICACY, WORRY AND ACHIEVEMENT

Research Questions Addressed

This section of the fine-grained analyses tests the conceptual framework that underpins this study by addressing the third of the research questions ('How are the girls' and the boys' perceptions of self-efficacy and worry related to their achievement on multiple-choice and constructed-response test formats?') for the 22 students who were interviewed.

Assertion 4

During tests, confidence and overconfidence are characterised by thoughts and feelings of well-being whereas a lack of confidence is characterised by the presence of worrisome thoughts and nervousness.

Preliminary Comments

When all of the students were asked on the *Survey About Science Tests* to describe the types of things that happen when they lose confidence during tests, they reported the presence of disruptive thoughts, withdrawal from tests, bad feelings, and off-task thoughts and behaviours (Table 4.12). These response categories were similar to the categories that subsumed the students' responses to worry during tests, namely, an inability to think clearly, bad feelings and withdrawal from tests (Table 4.17). Additionally, perspectives from Question 3 on the *Survey About Science Tests* showed that the students' positive thoughts in tests were associated with pleasant feelings, an impetus to try harder and the ability to think clearly. These reports were consistent with the review of the literature, which reported that test anxiety and low self-efficacy are characterised by unpleasant worrisome and emotional responses (Bandura, 1997; Liebert & Morris, 1967; I. Sarason, 1984; Spielberger, 1966; Wine, 1980), and that high self-efficacy is characterised by self-assurance in difficult

situations and an ability to ameliorate aversive emotional states (Bandura, 1997). Moreover, analysis of the quasi-experimental data showed, for the students in this study, that the strength of the students' self-efficacies and the extent of their worrisome cognitions were inversely related.

Hence, these perspectives influenced the development of the initial assertion, that confidence describes a state of well-being whereas a lack of confidence is characterised by the presence of worry and nervousness. It is important to note that the conceptual importance of overconfidence, which is also incorporated in these analyses, emerged during the first interviewing cycle.

Analysis of the Interview Data

Confidence and Overconfidence

Confidence and overconfidence were perceived by the students as transitory states of well-being that are characterised by the absence of nervousness, doubt and worrisome thoughts. Overall, 20 of the students (9 boys and 11 girls) associated various aspects of well-being with their confidence and their overconfidence:

- S. If you are more confident then you don't get nervous [and] you don't worry about it and start thinking about other things like 'I'm going to fail this test'. (Danielle)*
- S. It's just your self-esteem. If you've got high self-esteem then most of the time you're not worrying and you think you are going to get it right. (Cameron)*
- S. If you are feeling confident it takes all the doubts and all the worry away. (Joshua).*

Additionally, this absence of nervousness, doubt and worry was associated with self-perceptions of heightened efficacy and elevated achievement expectancies, together with 'happy', 'positive' and 'quite excited' feelings. No evidence was presented from any of the remaining students that the constructs of confidence and overconfidence describe unpleasant states.

Favourable self-perceptions of efficacy were reported by 14 of the students when they were asked about their confidence and their overconfidence during tests. (Although most students described confidence and overconfidence in similar terms, two students did report that overconfidence differs from confidence, in that, when they are overconfident they 'get really cocky' and they 'think how good [they] are'). Typically, these 14 students commented that their confidence and their

overconfidence was related to their convictions that they know the work and that they are able to answer achievement items.

- S. You think 'I know that' and just circle it because you think you're really good. (Madeline)*
- S. You think 'I can actually do this', and if you're not confident sort of thing, 'I can't do this'. (Sophie)*
- S. [Confidence] makes you feel like you know a lot more, like it makes you feel that you will be able to remember all the stuff that you learned. (David)*
- S. [Confidence] makes me seem like I know what I'm doing and it makes me happier about answering some of the other questions. (Wade)*

Additionally, most of the students imputed causal links between their perceptions of self-efficacy and their progress on tests, a theme that is explored in the next assertion.

Favourable achievement expectancies were reported by 11 of the students when they were asked about their confidence and their overconfidence during tests.

Typically, these constructs were related to the students' assurances that they are going to get a good mark or, at the least, that they expect to pass the test.

- S. You think its all going your way and you think you are going to get a good mark. (Cameron)*
- S. You feel that you know it and you are going to get a good mark. (Sophie)*
- S. You know that you are going to do your best and get a good mark. (Matthew)*

However, four of these students reported that their overconfidence is followed by feelings of disappointment when their unrealistically high achievement expectancies are not realised.

- S. I reckon its a bad thing to be overconfident because then you think you've done really well and you haven't and then when you get the test back you then you go, 'I did get that wrong', and you get real upset because you thought you went really well. (Madeline)*
- S. If you get a bad mark and you felt pretty confident you feel put down. (Cameron)*

Nevertheless, during tests overconfidence appears to be linked to a state of well-being.

Underconfidence

In contrast, the students perceived underconfidence as an unpleasant state. Nearly all of the students (11 girls and 8 boys) reported various aspects of 'thinking down' (that is, depressed self-efficacy), lowered expectancies for success, nervousness and worry to characterise their lack of confidence. Moreover, none of the remaining students reported any data to the contrary.

In all, 14 of the students associated their lack of confidence during tests with nervousness and worry.

- S. *[When] you're underconfident you're not thinking clearly and you are worrying about whether you are going to finish the test or things like that.* (Wade)
- S. *If I'm less confident [about my grade] I'm more nervous as well.* (Lucy)
- S. *Nervousness sometimes overcomes what you know and [it] overcomes the good feelings... Sometimes you might feel 'I'm going to do well, this is great', and when I get into the test I feel, 'This is bad. I'm not going to do well'. Nervousness has just overcome all [those good thoughts].* (Rebecca)

Additionally, 10 of the students reported that they lack the capability to correctly answer achievement items when they are underconfident. Typically, these thoughts included expressions of hopelessness and futility.

- S. *You just don't have any self-esteem. You just think, 'Oh no, I can't do that', and [you] just keep on going either really down in the dumps or, 'This is so hard. I can't do this.'* (Sophie)
- S. *You just think if I do try hard I won't get it.* (David)
- S. *You don't ... sort of feel good about yourself when you're going onto the next questions ... You don't feel good about yourself any more.* (Bernard)

Moreover, 7 of the students reported that they anticipate failure, or at the best diminished performances when they lack confidence during tests.

- S. *You're really going to think down and say 'I'm never going to finish the test, I'm not going to pass'.* (Julia)
- S. *It's bad because then you know that you have ... you just know that you are not going to do well.* (Kevin)
- S. *You just keep thinking I'm going to keep getting all these wrong now.* (Sarah).

Additionally, most of the students' reports of low self-efficacy were causally linked with their progress during tests, a theme that is explored in the next assertion. On the other hand, this causality was less evident in the responses of those students who described the presence of depressed achievement expectancies.

Concluding Comments

The fine-grained analyses of the interview data indicated that during tests, confidence and overconfidence were characterised by thoughts and feelings of well-being whereas underconfidence was perceived as an unpleasant state that is characterised by the presence of worrisome thoughts and nervousness. The additional insight yielded by these analyses relates to overconfidence. Although overconfidence

and confidence appear to describe similar states of well-being, some evidence was presented that the presence or absence of ‘cockiness’ separates these two constructs.

Assertion 5

The quality of students’ engagement with tasks during tests is related to their level of confidence. Students who are either over- or underconfident do not engage with tasks as well as students who are moderately confidence.

Preliminary Comments

The importance of motivational constructs was initially discounted in this study because the researcher assumed that all of the students would maximise their engagement with test items because tests are high stakes pursuits, and also because these students attended a relatively elite, fee-paying school. However, as this study progressed, it became clear that this assumption was grounded in a narrow understanding of motivation and motivational processes. Findings about the students’ positive and negative thoughts from the *Survey About Science Tests* indicated that students’ engagement with test items was related to their confidence. For example, the students in this study reported the presence of disruptive thoughts, off-task thoughts and behaviours, and withdrawal from tasks when they lost their confidence in tests. Furthermore, these students reported increased persistence with tasks and greater clarity of their thought processes when they were confident. This finding is consistent with an ample body of research evidence, which shows that self-efficacy in test situations is related to persistence, better management of time and a lesser likelihood for the premature closure of problems (for example; Bandura, 1997; Bouffard-Bouchard et al., 1991; Multon et al., 1991).

Hence, it was postulated from analyses of the survey data and from the review of the literature, that the quality of the students’ engagement in tests is directly related to their confidence. The conceptual importance of overconfidence, which is also reported in these analyses, emerged during the first of the interviewing cycles.

Analysis of the Interview Data

Overconfidence and Disengagement From Test Items

The boys and the girls who were interviewed in this study indicated that, when they are overconfident, the quality of their engagement with test items is less than when they are moderately confident. This disengagement with test items is characterised by their divided attention and a lack of caution as they rush their way through tests. In all, 18 of the students (11 girls and 7 boys) reported that their overconfidence is related to a lesser engagement with test items. No evidence to the contrary was advanced by the remaining 4 students (3 boys and 1 girl), who were unsure whether overconfidence during tests is a 'bad thing.' Moreover, most of these 18 students ascribed a causal influence to their overconfidence insofar as their engagement with test items is concerned.

Six students reported that the quality of their engagement with test items decreases when they are overconfident because they are unable to think clearly. Typically, they indicated that their minds become confused, their thoughts wander and that they forget answers to questions they know because their attention is divided between the test items and their off-task thoughts. This type of reduced engagement was perceived to arise when the students have unrealistically high outcome and achievement expectancies, as exemplified by Rebecca, or unrealistically high self-appraisals of their capabilities, as exemplified by Lucy. Rebecca reported that:

- R. Because you've been doing really well all throughout the test, like every question you've been confident on and you know you've done really well on every question, that's a bad thing because your confidence gets too high. Because of that you lose some of the information.*
- I. Oh, Okay. How can you lose some of the information when your confidence gets too high?*
- R. I'm not sure about that. I don't know why exactly it happens but I find sometimes I'm in a test and I'm doing really well and then I'm all quite excited because I know I'm going to do well and Mum's going to be happy, and then because I've been thinking about that instead of the questions something that I was thinking about sort of disappears. Like I'm sort of halfway through a question and I think of the answer and I start thinking of this and I've forgotten the answer.*

And Lucy reported that:

- L. Well you get really cocky and you think how good you are. You just wouldn't have your mind on the test.*
- I. Okay. What would your mind be on then?*
- L. How good you are.*

- I. *Okay. You are thinking how good you are and because you've got your mind on that then what happens? Why is that a bad thing?*
- L. *It would be distracting. It would take all your thoughts and you wouldn't have any thoughts on the test.*

Moreover, when students are overconfident they reported not trying as hard or thinking as much as they do when they are moderately confident. Consequently, students perceived that this type of disengagement from tests causes them to make 'silly mistakes' because they either misjudge or misread test items. Fifteen of the 18 students reported that this type of reduced engagement is associated with either their elevated achievement expectancies or their excessively optimistic self-efficacies. For example, these students reported for their unrealistically high outcome expectancies that:

- S. *If you are too overconfident [about your marks] there's a bad side to it. You just rush through it without thinking about the actual questions. You don't really think about the questions. (Cameron)*
- S. *You think you are going to do really well then when you get to a question you won't like read it twice or anything. You'll just circle the answer you think is right and you'll reckon you get it right, but because you didn't read the question carefully enough you'll probably get it wrong. (Angela).*
- S. *You just think you are going to go good and so you don't have to try because you are going to fluke the questions anyway. (Sarah)*

Additionally, these students reported for their excessively optimistic self-efficacies that:

- S. *If you are overconfident you just don't try as hard because you might think you know everything and so you don't have to try as hard. (Sarah)*
- S. *You think, 'I know all this,' and you rush through it but if you are sort of not confident but sort of confident halfway you are sort of careful about your answers. (Madeline)*
- S. *You go through too quick and you don't answer the questions right so you don't do as well. (Melanie)*
- S. *If you go into a test and you think because you're good at a subject, 'This is going to be really easy,' like that and the test is actually hard and you don't notice it and you go through all the questions like you've got half an hour or ten minutes left. Well then you've gone through it too quick and you've done really bad. (Melanie)*

Both elevated self-efficacies and elevated achievement expectancies appeared to be deleterious to girls and boys because they result in a lesser engagement with test items.

Underconfidence and Disengagement from Test Items

Similarly, when students are underconfident during tests, irrespective of whether they have depressed achievement expectancies or unfavourable self-appraisals, they perceived that the quality of their engagement with achievement items is less than when they are moderately confident. This type of disengagement with test items is characterised by their divided attention and by the withdrawal of their effort. All but one of the students (9 boys and 12 girls) reported that underconfidence during tests is associated with some form of reduced engagement. The remaining student was unable to decide whether underconfidence influenced his work during tests. Furthermore, most of the students ascribed a causal influence to their underconfidence insofar as their engagement with test items is concerned.

When students were asked about their underconfidence during tests, 16 of the students reported deleterious effects associated with their divided attention. Typically, these students reported that their underconfidence is characterised by an inability to concentrate, confusion, doubt and forgetfulness. Moreover, students' inability to concentrate was clearly associated with the presence of nervousness and worrisome thoughts. Students reported, typically, for their lowered self-efficacies that:

- S. *You might worry too much and the dangers would be like not thinking the question clearly and thinking of your answer properly. (Emma)*
- S. *You start not doing your best work and your mind freezes and you forget words that you know. (Kevin)*
- S. *You might know something but you might just doubt that it's right because of the other things you didn't know, and you think, 'Oh, yeah this is probably not right either'. (Joshua)*

Similarly, students reported for their depressed achievement expectancies that:

- S. *You don't concentrate fully on the question, you are thinking about two things at once and like you are not fully absorbing the question, you make silly mistakes and things like that. (Danielle)*
- S. *I'm just sitting there and I'll think, you know instead of working on the questions, I'll be thinking about my nervousness like I'll think, 'I'm not going to do well', and because I'm thinking all those thoughts it wastes time and you don't have as much time on the questions because you are so nervous. (Rebecca)*

No evidence was presented from any of the remaining students that underconfidence in any of its forms assisted them to think more clearly during tests.

Students also reported that they withdraw from particular test items when they lack self-efficacy for completing these items. Fourteen of the students reported this type of disengagement, which involves decreased persistence on difficult items and the omission of challenging items.

- S. When you don't sort of feel good about yourself when you are going onto the next questions, so if they're really hard as well you sort of just fall into a heap and you don't understand the questions and you just put a dash in the space and you go onto the next question. (Bernard).*
- S. You just keep thinking, 'I'm going to keep getting all these wrong now.' Because you think you don't know them you just give up. (Sarah)*
- S. You sort of don't try real hard because you just think, 'If I do try hard I won't get it'. (David)*

However, seemingly contradictory responses were produced by ten of the students who reported that they 'slow down' and 'try harder' on difficult questions when they lack confidence.

- S. When you don't have your confidence (self-efficacy) you slow down and you think the question over because some of the questions are pretty hard. (Bernard)*
- S. You could go through a test slowly and be able to read the questions through thoroughly, but [then] you could finish the test with 20 or 30 questions undone. (Matthew)*

It appears for these students, that the additional time which is expended to answer individual items is at the expense of completing the entire test. Thus, 'slowing down' and trying harder on individual items appears to represent another form of disengagement from tests.

Additionally, two students attributed an apparently positive influence to their worrisome thoughts. For example, Rebecca reported that:

- S. Sometimes [worry] can be a good thing because it makes you try harder. It makes you think 'I've got to do really well' and it makes you try harder and you can do well, but then there's other times when you just can't do anything.*

It appears, therefore, that worrisome thoughts can have a facilitative effect for some students. However, it is also important to recognise, as described previously, that increased engagement with individual test items amounts to an overall withdrawal from timed tests. Otherwise, students' worrisome thoughts and their nervousness, which accompanied and confounded their underconfidence, were associated with students' disengagement from test items.

- S. They [worrisome thoughts] set you back and make you think about other things [so] you don't concentrate in tests. (Madeline)*

S. You tend to sit there and twiddle your thumbs and bite your nails instead of actually getting on with the work, instead of setting your mind to the task. (Matthew)

S. You sort of give up sometimes. (Kevin)

Overall, for most students, nervousness and worrisome thoughts were associated with divided attention and the withdrawal of effort, and only a few students reported a facilitative effect for their worrisome thoughts.

Moderate Levels of Confidence and Engagement with Tasks

When the students responded to questions about over- and underconfidence they also reported that moderate levels of confidence enhance the quality of their engagement with test items. A total of 16 students (9 girls and 7 boys) reported that moderate levels of confidence are causally related to increased persistence and their attention to achievement items. No evidence to the contrary was reported by any of the remaining 6 students (3 boys and 3 girls). These reports about moderate levels of confidence were not directly solicited from the students. Instead, the responses that are analysed in this section arose during discussions about over- and underconfidence.

Ten students reported that moderately positive self-appraisals of their capabilities help them to recall and process information better than when they are either over- or underconfident, because they can think 'clearer' and 'straighter'.

S. If I'm really confident then I'll probably think clearer and probably remember more of what I've studied. (Wade)

S. It makes you head seem clearer and ... because you're confident you can think of it and [remember] it. (Emma)

S. When you are confident you seem to explain things better than you would if you are negative towards the exam or overconfident. (Michael)

Additionally, six students reported that appropriate self-appraisals of their capabilities cause them to increase their persistence during tests. Typically, students reported that they 'think harder' and they 'just want to keep going' at test items because they feel that they know their work and that they can remember it.

S. You think, 'I can actually do this,' and you try really hard. (Sophie)

S. If you have confidence it's good and you can do it and you think more about the test and you try to remember what you learnt. (Madeline)

According to students' reports, moderate levels of self-efficacy, rather than realistic achievement expectancies appeared to be associated with students' increased engagement in tests.

Concluding Comments

The fine-grained analyses of the interview data showed that the quality of the students' engagement with test items was related to their level of confidence, and that the students who were either over- or underconfident did not engage with items as well as students who were moderately confident. The major new insight yielded by these analyses relates to the role of overconfidence. These analyses show that unrealistically high self-efficacies and achievement expectancies may have a deleterious influence on the quality of students' engagement with test items. Additionally, some evidence was presented that low levels of worry can have a facilitative effect for some students.

Assertion 6

In timed tests, over- and underconfidence have a deleterious influence on achievement and moderate levels of confidence enhance academic achievement.

Preliminary Comments

As discussed in the previous assertion, when the students were asked, on the *Survey About Science Tests*, about the influence of their positive and negative thoughts, they reported that their engagement with test items and their propensity for making silly mistakes was related to their confidence. Additionally, analyses of the quasi-experimental data determined that item-specific self-efficacy and achievement were moderately correlated ($r_{xy}=+.38$) with each other. Moreover, a recent meta-analysis of the self-efficacy research literature (Multon et al., 1991) determined that self-efficacy beliefs account for an average of 14 percent of the variance in academic achievement.

Hence it was initially postulated that the students' confidences (that is, their self-efficacies and their achievement expectancies) would be positively related to their achievement. The importance of overconfidence, which emerged from the students' reports in the interviews is also incorporated in these analyses.

Analysis of the Interview Data

Analyses in the previous section have shown that students' achievement expectancies and their self-efficacies are related to their engagement with test items. Although the students were not asked to relate confidence specifically to their achievement, many of the students volunteered relevant data when they described the influences of their over- and underconfidence. Evidence for an association between confidence and achievement was supplied by 18 of the students (12 girls and 6 boys). One student (a boy) reported no association between confidence and achievement, and 3 boys failed to provide relevant responses.

Overconfidence and Lowered Achievement

Overconfidence, that is either elevated achievement expectancies or unrealistically high self-appraisals of capabilities were perceived by 8 of the students (2 boys and 6 girls) to be associated with their decreased achievement. Moreover, 6 of these students ascribed a causal role to their overconfidence. Typically, these students reported that when they are overconfident that they rush through tests and that they fail to engage with test items as completely as they would if they were moderately confident. In this context, these students perceived that their overconfidence had a deleterious influence on their achievement.

- S. *There's a bad side to it (elevated achievement expectancy). You rush through without thinking about the actual questions. That's not very good if you rush though it because you don't really think about the questions [and] your mark could go down. (Cameron)*
- S. *If you are too confident (self-efficacy) you just whiz through all the questions and you think, 'Gee that was really easy,' and then you've done really bad because you've thought you are really confident and you did it without thinking. (Danielle)*
- S. *If you are too confident you might do like a stupid mistake and not answer the question properly and then it'll make the marks go down. (Emma)*

Only Kevin reported, to the contrary, that he did not think there were any dangers associated with his overconfidence. However, he was unable to think of any reasons to justify his viewpoint.

Additionally, none of the remaining students provided any evidence to indicate that overconfidence is associated with higher levels of achievement.

- S. *I don't find that it (elevated achievement expectancy) helps me through the test. I find that I may not, I will not do any better than I normally would, but I won't*

do any worse unless I'm thinking of the question and it [my answer] just disappears because I'm thinking of something else. (Rebecca)

- S. *You might go in [with high levels of self-efficacy] just thinking its all right and you just still have to do the work to get the 'A' and you can't go in positive and know that you are going to get it. You've got to do the work to get it. (Anne).*

The remaining 13 students did not provide any confirming or disconfirming evidence for the assertion that overconfidence has a deleterious influence on achievement.

Underconfidence and Lowered Achievement

Underconfidence was also perceived by 14 of the students (10 girls and 4 boys) to be associated with their decreased achievement. All but one of these students attributed a causal influence to their underconfidence, either by way of reduced engagement with test items or by way of their worrisome and self-defeating thoughts, which they perceived as causing them to disengage from test items.

- S. *Its general that you don't really do well because if you are confident you know that you can do it, but if you are not confident (low self-efficacy) you sort of know that you can't and you don't [do it]. (David)*
- S. *You worry too much [when you don't expect a good mark] and you don't do your questions right. (Melanie)*
- S. *If you are less confident (low achievement expectancy) it will effect your whole grade a lot [by] not being able to think clearly. (Lucy).*

An important recurring theme in many of these students' reports was that their achievement is influenced by their self-efficacy because you 'have to think you can do it before you can do it.'

Five students perceived, however, that their underconfidence does not necessarily translate into decreased levels of achievement. Instead, these students reported that they 'slow down' on tests to 'read the questions thoroughly' and 'be a lot more careful' when they lack confidence. However, all but one of these students indicated that this increased engagement with individual items has a deleterious influence on their achievement if they cannot complete the test in the allotted time. Additionally, none of these students perceived that their underconfidence was conducive to higher levels of achievement.

Moderate Levels of Confidence and Enhanced Levels of Achievement

Moderate levels of confidence were perceived by 12 of the students (6 boys and 6 girls) to be associated with enhanced levels of achievement. For example, Sophie

reported, 'If you've got confidence you usually do better.' Furthermore, eight of these students asserted, that when their achievement expectancies, and particularly their self-efficacies, are realistic for the task that their confidence exerts a causal influence on their achievement. This increased achievement was attributed to their better quality engagement with test items.

- S. It would have given us better marks because of our confidence (self-efficacy) building up would have got us on the ball a bit quicker. (Matthew)*
- S. If you are positive (self efficacy) and you are feeling good about how you are going to do it will affect your marks in a positive way so that you will get better marks. (Kevin)*
- S. Yeah [it helps your marks]. Being more confident (self efficacy) just helps you remember things a lot better. (Wade)*

However, three students reported that moderate levels of confidence have no influence on their achievement. Two of these students gave inconsistent responses and the remaining student asserted that confidence 'doesn't help' but that it is a 'good thing' to have in tests. However, this student was unable to give any reasons to support his viewpoint. Otherwise, nine of the students did not make any responses relevant to moderate levels of confidence and their achievement on science tests.

Concluding Comments

The fine-grained analyses of the interview data indicated that during timed tests, that for these students, over- and underconfidence had a deleterious influence on achievement, and that moderate levels of confidence were perceived to enhance academic achievement. The new insights yielded by these analyses relate to the role of overconfidence and time. First, these analyses have shown, by way of students' perceptions, that overconfidence has deleterious influence on achievement. Although the quantitative data analyses supported by scatter plots of achievement and self-efficacy indicated a linear relationship between self-efficacy and achievement, the fine-grained analyses of the interviews suggest that the relationship may be curvilinear. Second, these fine-grained analyses have shown that decreased achievement is partly related to the limited time that is provided for students to complete their tests — when students lack confidence their rate of progress slows to the extent that they may not finish timed tests.

SUMMARY OF THE FINDINGS

This chapter has reported the 'fine-grained' analyses of the interpretivist-constructivist data that were collected from the 22 students who were interviewed. These analyses described and interpreted the girls' and the boys' meaning-perspectives about self-efficacy, worry and achievement for multiple-choice and constructed-response test formats.

For the students in this study, the fine-grained analyses show that the content of the boys' and the girls' confidences, that is, their self-efficacies and their achievement expectancies are similar to each other for science tests. Overall, the students' confidences before tests appeared to be determined chiefly by their perceptions about the adequacy of their preparation, and during tests by the difficulty of items and the judgements they made about their progress. Additionally, format-related effects made a relatively unimportant contribution toward the girls' and the boys' confidences during tests, and no differences were evident between the content of the boys' and the girls' format-related self-efficacies.

The findings from the interpretivist-constructivist research design were consistent with the model that was developed in Chapter 2 for the relationships between worry and self-efficacy, and self-efficacy and achievement. However, little data were collected and analysed for the relationship between worry and achievement because the girls' and the boys' worrisome thoughts, in the main, could not be separated from their perceptions of low self-efficacy. Additionally, the fine-grained analyses extended this model by suggesting that the quality of students' engagement with tasks is an important precursor of their achievement, and that overconfidence and strategies for the regulation of affective states need to be included in this model.

CHAPTER 6

DISCUSSION AND CONCLUSIONS

OVERVIEW OF THE STUDY

As outlined in Chapter 1, this study was prompted by evidence that boys achieve relatively better than do girls on assessment items set in multiple-choice formats, and that the magnitudes of these gender-related differences are either decreased, or their directions reversed when constructed-response formats are used (Anderson, 1989, Bell & Hay, 1987; Bolger & Kellaghan, 1990; Bransky & Qualter, 1993; Breland et al., 1991; Bridgeman & Lewis, 1994; D. Burkam & A. Burkam, 1995; Harding, 1981; Hellekant, 1994; Hoste, 1982; Mazzeo et al., 1991, 1993; R. Murphy, 1980, 1982; Smail & Kelly, 1984; Stumpf & J. Stanley, 1996; Teese et al., 1995; Volkoff & Hocevar, 1995; Whitehouse & Sullivan, 1990; Wood, 1978). With the current emphases on issues relating to womens' and girls' underparticipation in high school science and science-related careers, and on issues relating to validity and fairness in testing, this researcher judged that a greater understanding is required for how constructed-response formats function, in comparison to the multiple-choice response format, so that all students can be afforded equal opportunity to demonstrate their knowledge and understanding of science content. As mentioned in Chapter 1, the major purpose of this study was to contribute to that enhanced understanding.

The study was guided by a model, developed from an extensive review of the literature, incorporating the dimensions of generalised self-efficacy, item-specific self-efficacy and worry. The research questions that were derived from this model were operationalised within separate frameworks constructed within the positivist and the interpretivist-constructivist research traditions. The participants in the study, who comprised all the Year 9 boys and girls at the school where the researcher is employed as a teacher, rated their item-specific self-efficacies immediately before answering a series of equivalent multiple-choice and constructed-response achievement items and they completed a series of worry items for these response formats halfway through their achievement tests. Additionally, data were collected by surveying all the students and by interviewing selected students. Chapter 4 presented the findings from

the data that were collected within the positivist research design and Chapter 5 presented the findings from the data that were collected within the interpretivist-constructivist research design.

The present chapter follows Bernstein's (1991) recommendation by regarding each of the research traditions as a conversational partner of the other. As described in Chapter 3, this process involves a dialectical approach involving each of the research traditions. In this study, dialogical encounter was operationalised as a search for commonalities, together with points of difference and conflict. The first section of this chapter addresses issues relating to the quality of the study; the second section discusses the conclusions from this study; and the final section outlines the methodological, theoretical and educational implications of this study.

QUALITY OF THE STUDY

Quality of the Findings From the Positivist Data

As described in Chapter 3, this study developed a series of procedures to optimise the quality of the quasi-experimental and survey data. All the quasi-experimental data were derived from students' responses to specially designed multiple-choice and constructed-response achievement items. The content validity of these items was addressed by developing the achievement items from a series of existing test items with the assistance of the cooperating teachers, a table of specifications and standard test construction manuals. The reliabilities of the achievement, item-specific self-efficacy and worry data that were generated in response to these achievement items were enhanced by (i) administering the test booklets according to a set of common instructions, (ii) training the students and cooperating teachers regarding how the test booklets should be answered, (iii) using marking schedules for scoring the achievement items and a series of standard procedures for scoring their other responses, (iv) checking the accuracy of the scoring procedures, (v) deleting students from the analyses who either appeared to answer items spuriously, or failed to answer a sufficient number of items, and (vi) replicating the data collections for the *Organisms and Food* unit topic. Furthermore, the

reliability of the self-efficacy and the worry data were enhanced and the unidimensionality of the worry scale established by conducting pilot studies at the research site. Additional evidence from the interviews suggested that the methods for collecting the research data had little influence on students during the tests, and that the reliability and the validity of this study were not threatened by the way that students answered the test booklets. Finally, the internal consistencies of the quasi-experimental data were generally quite high. The lower internal consistencies for some of the achievement data were judged as inconsequential, because it appeared most likely that these data reflected the heterogeneity of the unit topics. Overall, the findings from the quasi-experimental design seemed likely to be highly reliable for the constructs they purported to measure.

The *Survey About Science Tests* was specially developed and piloted for this study. Although the check-recheck reliabilities for coding data from this survey were very high, as reported in Chapter 4, the findings that were drawn from these data may have been compromised. In particular, it was not always possible to link students' responses on the survey form to the constructs that were of interest to this study. For example, it was unclear whether some of the students' responses related to the worry construct or self-efficacy. Therefore, high levels of worry were confounded with low levels of self-efficacy and low levels of worry with high levels of self-efficacy. This ambiguity was addressed by analysing these data in terms of the self-efficacy construct. This approach appears to be justified because of the overlap between these (that is, the self-efficacy and worry) constructs.

Quality of the Findings From the Interpretivist-Constructivist Data

As described in Chapters 3 and 5, this study addressed the five criteria that F. Erickson (1986) lists for reducing researcher bias and ensuring the existence of evidentiary warrants. Nevertheless, it is important to recognise two limitations of the interpretive data. First, the findings from the fine-grained analyses were reliant on students' self-reported perceptions. Although participant observation is a key element of interpretive research designs (F. Erickson, 1986), it was not employed in this study because this researcher judged that participant observation was unlikely to make a

substantial contribution toward the specific problems addressed in this study. Second, it was not always possible to draw clear links between the language that was used in the interviews and the theoretical constructs that were described in the literature. For example, the term 'confidence' was used by students to refer to their self-efficacy, outcome expectancies and achievement expectancies. Consequently, some of the findings from the fine-grained analyses were framed in terms of confidence rather than self-efficacy. Additionally, as reported for data from the *Survey About Science Tests*, it was not always possible to distinguish between the worry and self-efficacy constructs. Again, this issue was addressed by analysing ambiguous data in terms of the self-efficacy construct rather than worry.

Transferability of the Findings

It is important to recall that the students who participated in this research were not representative of the Year 9 students in the state of Queensland. Hence, this study cannot and does not subscribe to positivist notions of sample-to-population generalisation. Rather, the position taken is that Merriam's (1988) notions of transferability are well suited to the purpose and design of this study. These conceptualisations include (i) regarding the findings as working hypotheses rather than as definitive generalisations, (ii) locating the general within descriptions and analyses of the particular, and (iii) user generalisability, whereby each reader judges for themselves the case-to-case applicability of studies such as this one. Additionally, it is important to note that the transferability of any of the findings from this study is contingent on satisfying concerns relating to reliability and internal validity (F. Erickson, 1986; Guba & Lincoln, 1981; Merriam, 1988).

The transferability of the findings from this study was addressed, as described earlier in this section, by adopting procedures to ensure the quality, that is, the internal validity and the reliability of the findings from the positivist design and the existence of evidentiary warrants for the assertions that were made within the interpretive-constructivist design. Although it would have been desirable to collect data from a number of research sites to establish the typicality of this study (F. Erickson, 1986; Merriam, 1988; Miles & Huberman, 1994), such an undertaking was well beyond the

scope of this study. Therefore, the findings from this study are regarded as a series of working hypotheses rather than as new theory, and this report was written to facilitate case-to-case generalisations. Accordingly, thorough descriptions are provided of the research setting, the participants and their instructional context, together with the methodology, analyses and findings so that readers can identify commonalities (that is, the general within the particular) between their context and so the context of this study, and so that they can construct their own inferential bridges in order to judge the applicability of the findings for themselves.

DISCUSSION OF THE CONCLUSIONS

The Content and Extent of Girls' and Boys' Self-Efficacies

Statement of the Research Question

The first of the research questions guiding this study sought to determine, for the students who participated in this research, whether there were gender-related differences in the content and the extent of students' self-efficacies for answering science achievement questions set in multiple-choice and constructed-response formats; and, consequently, whether there was a gender-related interaction between perceptions of self-efficacy for answering science achievement questions and the response format these questions required.

The Content of Boys' and Girls' Self-Efficacies

The first of the working hypotheses emerging from this study is: *There are no substantial gender-related differences in the content of girls' and boys' self-efficacies for answering multiple-choice and constructed-response tests.* This hypothesis subsumes assertions relating to the boys' and the girls' generalised confidences before tests and during tests with findings (from each of the research traditions) about the content of students' generalised self-efficacies for answering multiple-choice and constructed-response questions.

At the most general level, the girls' and the boys' confidences were related, before tests, to their perceptions about the adequacy of their preparation, and during tests, to their perceptions about the difficulty of the items and their progress towards completion of the test. The literature (for example; Birenbaum & Nasser, 1994; Culler & Holahan, 1980; Hembree, 1988; P. Kalechstein et al., 1989; Wittmaier, 1972) strongly supports this association between ineffective study habits and procrastination, especially for highly test-anxious students who harbour low perceptions of their capabilities. This finding is consistent with skills deficit models for test anxiety, as described in Chapter 5, which conceptualise students' worrisome thoughts (and also their perceptions of low self-efficacy) as an awareness that they are not well prepared for tests and that they are likely to do poorly (Tobias, 1985). Additionally, Bandura's (1986) claims about the importance of mastery experiences for the development of self-efficacy perceptions are underscored by two of the findings from this study. First, the students' reflections about the extent of their preparation and their performance on quizzes and revision activities contributed toward their self-efficacy. Second, the boys' and the girls' self-efficacies during tests were moderated by perceptions about their progress and the difficulty of the test items. These findings are also consistent with previous research (Bandura, 1986; Pajares, 1997), which demonstrates that outcomes which are interpreted as successful heighten self-efficacy perceptions, whereas outcomes that are interpreted as failures lead to the lowering of self-efficacy perceptions.

For multiple-choice items, both the girls and the boys in this present study reported the same factors as contributing toward their perceptions of high and low self-efficacy. These factors, which have been listed in Chapter 4, are consistent with those factors which emerged from another recent study of Australian high school students (Rennie & L. Parker, 1998). The small, gender-related differences in the relative importance of these factors provides an interesting point of comparison with extant hypotheses about gender and item-response formats. For example, although both the girls and the boys reported that distinguishing between the distractors was the chief factor contributing toward their inefficacy on multiple-choice items, this present study did detect some evidence consistent with the view that the multiple-choice items favoured the boys' ways of thinking (Harding, 1994; P. Murphy, 1988,

1991, 1996). In describing reasons for their self-efficacy on multiple-choice items, the boys indicated that the provision of the correct answer was the most important factor contributing toward their sense of self-efficacy, whereas the most important factor for the girls related to their capabilities for guessing multiple-choice answers. This finding, which can be interpreted as an indication of the girls' relatively greater uncertainty and the boys' relatively greater certainty of deducing an answer for multiple-choice items, is consistent with P. Murphy's (1988, 1991) assertion that girls are more likely than boys to see either several or no correct answers among the distractors. This comparison is interesting in the context of reports that boys are more likely than girls to guess multiple-choice answers when they don't know the answer (Anderson, 1989; Ben-Shakhar & Sinai, 1991; Chopin, 1975; Hanna, 1986; Linn et al., 1987; S. Sherman, 1974). In this present study, the girls seemed to have had little compunction in guessing answers to multiple-choice items. However, further discussion of guessing tendencies is unwarranted, partly because this study did not address this phenomenon directly, and thus did not, for example, compare the ratios of omitted items for the boys and the girls.

For constructed-response items, again the girls and the boys in this study reported similar factors to each other as contributing toward their high and low self-efficacies, and gender-related differences in the relative importance of these factors were small. Again, the reasons that the students gave for their self-efficacies on constructed-response items were broadly similar to those categories identified by Rennie and L. Parker (1998). It is important to recognise for the purposes of this discussion, however, that this present study's operationalisation of constructed-response items as requiring short, factual and episodic responses was more attuned to boys' writing styles than to those of girls (Gorman et al., 1988; P. Murphy, 1988, 1991, 1996). Even so, the boys identified having to write rather than select their responses as relatively more problematic than the girls. Nevertheless, there were several indicators that the boys perceived relatively more self-efficacy than the girls for answering constructed-response items. For example, the girls were more likely to regard the constructed-response items as relatively more rigorous than the boys, and the boys were more likely than the girls to report that the constructed-response items provided them with more opportunity than the multiple-choice items to show off their

understanding. Overall, the content of the boys' and the girls' self-efficacies for answering the constructed-response items was characterised more by their similarities than by their differences.

The Extent of Girls' and Boys' Self-Efficacies

The second of the working hypotheses emerging from this study is: *Boys report relatively more self-efficacy for science tests than do girls, and the extent of these gender-related differences is not moderated by test item-response formats.* This hypothesis, which was derived from the positivist research design, subsumes individual findings relating to the boys' and the girls' generalised and item-specific self-efficacies for multiple-choice and constructed-response items.

As discussed earlier in this section, small differences in the relative importance of the factors contributing toward the girls' and the boys' self-efficacies on multiple-choice and constructed-response items suggested that the boys perceived more self-efficacy than the girls for each of these response formats. Additionally, the boys reported higher item-specific self-efficacies than the girls for each of the item-response formats. In statistical terms, these gender-related effects were moderately sized and practically important (Cohen, 1969).

The working hypothesis stated here is consistent with a large body of research, which shows that males generally report relatively higher confidences, higher generalised and item-specific self-efficacies and greater expectancies for success than females in sex-typed fields such as science and mathematics than their achievements would suggest (Fennema & J. Sherman, 1977, 1978; Fullarton, 1993; Furst et al., 1985; Hyde, Fennema, Ryan et al., 1990; Kelly, 1988; Lenney, 1977; Licht 1987; Licht & Dweck, 1983; Licht et al., 1989; Linn & Hyde, 1989; M. Lundeberg et al., 1994; Maccoby & Jacklin, 1974; Mednick & Thomas, 1993; Vollmer, 1986). However, it was somewhat surprising that the extent of these self-efficacy differences was larger for the *Organisms and Food* items than the *Earth Science* items because there is ample evidence that girls have relatively more interest and out-of-school experiences than boys in life related science topics (for example; Archenhold, 1988; S. Johnson, 1987; Kahle & Lakes, 1983; Sjoberg & Imsen, 1988). Nevertheless, in

statistical terms, this interaction between gender and unit context was very small, and it was unlikely to have any practical importance for the students in this study (Cohen, 1969).

Also, in relation to self-efficacy, the finding that the boys' relatively greater self-efficacies were more pronounced when they answered items correctly rather than incorrectly is of interest. Direct comparisons with previous research are not possible because there is a paucity of literature to describe boys' and girls' self-efficacies on items they answer correctly and items they answer incorrectly, particularly for high school students. The finding is, however, at odds with the work of M. Lundeberg and her associates (M. Lundeberg, A. Brown, Fox, Elbedour, 1998; M. Lundeberg et al., 1994), who report that the relative confidences of male university students tend to be exaggerated to a greater extent when they are wrong. The need for additional research in this area is underscored by the critical importance of confidence discrimination in educational settings (M. Lundeberg et al., 1994).

In this present study, a considerable amount of evidence was presented, from each of the research traditions, that the girls and the boys who participated in this study perceived more self-efficacy for the multiple-choice items than the constructed-response items. This finding subsumes evidence from the quasi-experimental design (relating to the boys' and the girls' higher item-specific self-efficacies for multiple-choice items than constructed-response items) and the survey design (where the girls and the boys tended to report more generalised self-efficacy and a preference for answering multiple-choice items over constructed-response items). This finding is consistent with other research which shows that high school students prefer multiple-choice items over explanatory, calculation and essay items (Rennie & L. Parker, 1998; Zeidner, 1987). However, this pattern does not appear to be as clear for university students (compare; Gellman & Berkowitz, 1993; L. Smith, J. Smith & Lackland, 1998) as it does for high school students. Importantly, no interaction was detected in this study between gender and item-response formats for the boys' and the girls' item-specific self-efficacies on multiple-choice and short-answer constructed-response items. This finding makes a significant and new contribution to the research literature.

The Content and Extent of Boys' and Girls' Worrisome Cognitions

Statement of the Research Question

The second of the research questions guiding this study sought to determine, for the students who participated in this research, whether there were gender-related differences in the content and extent of students' worrisome cognitions for answering science achievement questions set in multiple-choice and constructed-response formats; and, consequently, whether there was a gender-related interaction between worrisome cognitions on science achievement questions and the response format these questions required.

The Content of Girls' and Boys' Worrisome Cognitions

The third of the working hypotheses emerging from this study is: *There are no substantial gender-related differences in the content of boys' and girls' format-related worrisome cognitions for answering science tests.* This hypothesis subsumes findings from the survey design about the content of students' positive and negative thoughts during tests, together with findings (from each of the research traditions) about the content of students' format-related worries for the multiple-choice and the constructed-response questions. It is important to recall in the following discussion that separating the content of the boys' and the girls' format-related worrisome cognitions from their format-related self-efficacies was problematic.

At the most general level, the girls and the boys who participated in this study reported similar types of negative and positive thoughts when answering science tests. However, there was some evidence from the relative importance attributed to these thoughts to indicate that the girls perceived themselves as experiencing more worry during tests than the boys. First, a slightly higher proportion of the girls than of the boys reported the presence of worrisome thoughts when describing their negative thoughts during tests. Second, the girls were more likely than the boys to report the presence of self-focused negative thoughts (that is, self-deprecatory thoughts, self-blame and self-doubt) and self-focussed positive thoughts (that is, self-messages of

encouragement, reassurance and exhortation). It is instructive to recall, in this context, that highly test-anxious students are characterised by the presence of self-focussed thoughts, irrespective of whether these thoughts consist of negative self-evaluations or coping cognitions (Blankstein et al., 1991; Bruch et al., 1983; Deffenbacher, 1980; Galassi et al., 1981, Ganzer, 1968; Hembree, 1988; M. Many & W. Many, 1975; Prins et al., 1994; I. Sarason, 1960; I. Sarason et al., 1991; Wine, 1971, 1980; Zatz & Chassin, 1985). Additionally, it is also important to recall from the analyses of the interpretive data that self-talk and positive thinking were used by some of the students to regulate the influence of their affective, and therefore their worrisome responses to tests.

As mentioned previously, findings relating to the content of the boys' and the girls' worrisome cognitions for the multiple-choice and the constructed-response items are problematic because they are confounded with the content of the girls' and the boys' self-efficacies for each of these test formats. However, because there is an overlap between these two constructs (Wine, 1980) such that self-efficacy variables appear to be subsumed by the worry construct, it can be assumed that the contents of the boys' and the girls' format-related worrisome thoughts are similar to those aspects of the response formats for which they perceive little capability. Therefore, it also seems highly likely that there are no substantial differences between the content of the girls' and the boys' worrisome cognitions for answering multiple-choice and constructed-response items.

The Extent of Boys' and Girls' Worrisome Cognitions

The fourth of the working hypotheses emerging from this study is: *Girls perceive a relatively higher frequency of worrisome cognitions during science tests than do boys, and the extent of these gender-related differences is not moderated in any substantial way by test item-response formats.* This hypothesis, which was derived from the positivist research design, subsumes individual findings relating to the girls' and the boys' positive and negative thoughts from the survey design, and measures for the extent of the boys' and the girls' worrisome thoughts from the quasi-experimental design.

As indicated earlier in this section, small differences in the relative importance of the girls' and the boys' negative and positive thoughts suggested that the girls in this study perceived they were more susceptible to worrisome cognitions during science tests. This finding, that the girls reported relatively more worry in test settings than the boys is well documented in the literature (Bridgeman & Schmitt, 1997; Hembree, 1988). Moreover, gender-related differences for worrisome cognitions, as measured by the quasi-experimental design, were small to moderately sized for each of the response formats. These effects were similar in size to those reported by Hembree (1988) in his meta-analysis, and statistically, they were likely to be practically important for the students in this study (Cohen, 1969). Additionally, evidence from the quasi-experimental design showed that the girls' and the boys' worrisome cognitions were moderated by test item-response formats. This finding, that students express more worry for constructed-response items than multiple-choice items has been reported elsewhere (for example, Crocker & Schmitt, 1987; Shaha, 1982; Weare, 1984) and it is consistent with high school students' preferences for answering multiple-choice items over constructed-response items (Rennie & L. Parker, 1998; Zeidner, 1987), together with the well-documented finding that associates higher levels of test anxiety with more difficult tasks (for example, P. Becker, 1982; Deffenbacher, 1980; Hembree, 1988; Morris & Liebert, 1969; Morris & Fulmer, 1976).

Importantly, this present study shows that the extent of the gender-related differences for students' worrisome cognitions was moderated only slightly by the item-response formats, and then in the opposite direction to that expected. In this study, the girls reported relatively more worry than the boys on the constructed-response items than they did for the multiple-choice items. Overall, the extent of the interaction for worry scores between gender and test item-response formats was very small, and in statistical terms, it was unlikely that it had any practical importance for the students in this study (Cohen, 1969). Although very few studies have compared girls' and boys' worrisome cognitions for different test formats, the findings of this present study contrast with one report that 15-year old boys perceived relatively more anxiety about constructed-response questions than 15-year old girls (P. Murphy, 1981, cited by Gipps & P. Murphy, 1994). However, additional comparisons are

problematic because neither the extent of these differences nor the methodology of the study are delineated.

The Relationship Between Self-Efficacy, Worry and Achievement for Boys and Girls on Multiple-choice and Constructed-response Test Formats

Statement of the Research Question

The third of the research questions guiding this study sought to determine, for the students who participated in this study, how the girls' and the boys' perceptions of self-efficacy and worry relate to their achievement on multiple-choice and constructed-response test formats.

The Achievement Context

The fifth of the working hypotheses emerging from this study is: *There is no practically important interaction between gender and test item-response formats for boys' and girls' achievements on multiple-choice and short-answer constructed-response tests.* This hypothesis, which is derived from the quasi-experimental design, provides evidence that helps clarify some of the uncertainties in the literature that was reviewed in Chapter 2.

The equivalent multiple-choice and constructed-response items that were used in this study were specially developed so that they tested the same knowledge and understanding of the subject matter. Further, the two types of items were couched in the same contexts and they referred to the same figural information. Despite this extensive 'matching' of item stems, the constructed-response questions were more difficult and discriminating than the multiple-choice questions. This pattern is consistent with findings from the United States (Martinez, 1998; Mehrens & Lehmann, 1984) and it implies the operation of format-related, construct-irrelevant factors (see Scheuneman, 1991). The fact that, in this study, credit was given for partially correct answers to the constructed-response items but not for the multiple-choice items underscores the greater difficulty of the constructed-response items and

the extent of these construct-irrelevant factors.

In statistical terms, this study found no evidence of any practically important gender-related differences in the girls' and the boys' achievements for either of the response formats (Cohen, 1969). Nonetheless, the boys received a marginal advantage over the girls by the use of multiple-choice items on the *Earth Science* test, and the girls received a marginal advantage over the boys on the *Organisms and Food* test by the use of multiple-choice questions. However, none of these interactions between gender and unit context was likely to have any practical importance for the students in this study (Cohen, 1969).

In some ways this lack of an interaction between gender and item-response formats for the students' achievements is hardly surprising. As noted in Chapter 2, the research literature revealed a lack of consensus about the extent of the interaction between gender and item-response formats. Whereas some researchers report a strong interaction (for example, Bolger & Kellaghan, 1990; Breland et al., 1991; Hellekant, 1994), others (for example, Teese et al., 1995; Whitehouse & Sullivan, 1990; Willingham & Cole, 1997a) suggest that the interaction is quite small. Part of this confusion seems to be related to the broad range of tasks defined by the 'constructed-response' categorisation (Bennett, 1993; Martinez, 1998) and a tendency for researchers to quote the statistical significance, rather than the practical importance of their findings.

This failure to find a practically important interaction between gender and item-response format is also consistent with earlier indications in the literature (for example, Teese et al., 1995; Whitehouse & Sullivan, 1990) and two reviews that emerged since this study was conceptualised. In one of these reviews, Willingham and Cole (1997a) located 12 studies that compared achievement scores for paired stem-equivalent multiple-choice and constructed-response items, where the constructed-response items required short answer responses. None of the five studies that were motivated by an interest in girls' and boys' relative achievements detected any gender-by-format interactions for boys' and girls' achievement scores. Consequently, Willingham and Cole (1997a) concluded that '[r]emoving the answer alternative from a MC question and requiring the examinee to produce the answer unaided appear[s] to have little if any effect on what is measured or on the relative

performance of females and males' (p.351). Similarly, Ryan and Franz (1998) report inconsequential gender-related differences from their recent meta-analysis of 14 studies that were drawn mainly from high-school samples. According to their review, boys achieve relatively higher than girls on selected-response items than constructed-response items. However, most of the format-related effect sizes were only about one-tenth of a standard deviation. It is also important to note in this context that Ryan and Franz (1998) employed definitions that were considerably broader than those adopted in this study, thus making direct comparisons complex — they operationalised the free-response formats to include writing as well as performance assessments and the selected-response format to include multiple-choice, matching and true-false formats.

The Relationships Between Self-Efficacy, Worry, and Achievement

The final working hypotheses of this study are: *The relationships between the variables in the theoretical model underpinning this study are supported; and, the quality of students' engagement with test items mediates the relationship between self-efficacy and achievement, particularly when time is limited.*

Interrelationships Between the Variables in the Model

The finding that self-efficacy was positively associated with achievement subsumes a series of indicators from each of the research traditions. First, confidence before tests and achievement seemed to have similar antecedents. The adequacy of students' preparedness for tests, which included variables relating to their effort and their study, together with variables associated with their teacher and the quality of their instructional experiences were identified from the interpretive database as factors that contribute toward the students' confidences before tests, and from the survey data as factors that contribute toward their achievement. Second, the students perceived the existence of relationships between confidence, achievement-related behaviours and achievement. In particular, underconfidence was perceived from the survey and the interpretive data to have a deleterious influence on achievement because it influenced the quality of the students' attention to tasks and also because it caused the students to work slowly or withdraw their effort. Consequently,

underconfident students may fail to complete achievement items to the extent of their capabilities during timed tests. Additionally, analyses of the interpretive data identified the deleterious influence that overconfidence was perceived to have on achievement through students' divided attention and their lack of caution. Finally, the quasi-experimental data yielded positive correlations between the students' item-specific self-efficacies and their achievement for each of the item-response formats. Overall, item-specific self-efficacy and achievement were moderately correlated with each other. An average of 14% of the variations in multiple-choice achievement were associated with variations in item-specific self-efficacy; and an average of 19% of the variations in constructed-response achievement were associated with variations in item-specific self-efficacy. It is instructive to recall in this context that Multon et al. (1991) calculated a similarly sized mean association of 14% between self-efficacy and achievement in their meta-analysis.

This finding that self-efficacy and achievement were related to each other is in agreement with Bandura's (1986, 1997) theoretical framework that informed this study and an extensive body of research findings from the last 20 years that has 'reasonably secured' (Pajares, 1997, p.563) the empirical connection between self-efficacy and achievement. Moreover, the influence that self-efficacy beliefs have on achievement through motivational variables such as effort, persistence and perseverance has also been documented (Bandura, 1997; Bouffard-Bouchard et al., 1991; Multon et al., 1991). There has, however, been little research investigating the relationship between overconfidence and achievement. Although Bandura (1986) theorises that unnecessarily optimistic self-efficacies have a beneficial influence through increased effort and persistence, this present study raises questions regarding the extent to which excessively optimistic efficacy appraisals serve to increase human functioning. This study suggests a curvilinear relationship between self-efficacy and achievement, and that excessively high levels of self-efficacy are maladaptive, as are underestimations of self-efficacy. Additionally, it must be noted that the causality which students imputed to their self-efficacy is a matter of their perceptions and that more appropriate research designs (see Pajares, 1997) are required to test the understandings provided by their reports.

The finding that students' worrisome cognitions were inversely related to their

achievement for each of the response formats subsumes several indicators from the positivist tradition. First, in the survey data, the students imputed a deleterious influence to their negative thoughts (chiefly worry) through their divided attention during tests, bad feelings and withdrawal from tests in ways that were consistent with cognitive-attentional (I. Sarason, 1980, 1984; Wine, 1971, 1980) descriptions for test anxiety. Although several students reported that worry has a facilitative effect through their increased engagement with individual items, this increased engagement with individual items is at the expense of completing timed tests. Second, the quasi-experimental data yielded an inverse relationship between the frequency of the girls' and the boys' worrisome cognitions and their achievement for each of the response formats — 12% of the variations in the students' multiple-choice achievement scores and 14% of the variations in their constructed-response achievement scores were associated with worry. Overall, worry and achievement were moderately correlated with each other. This finding, that the worry component of test anxiety is inversely related to achievement is reported in the research literature (for example; Doctor & Altman, 1969; Liebert & Morris, 1969; Morris & Liebert, 1970), and the extent of the association between these variables is similar to that of 10% reported by Hembree (1988) in his meta-analysis.

The finding that self-efficacy and worry were inversely related to each other subsumes several indicators from each of the research traditions. The first indicator relates to this researcher's difficulty with the confounding of low self-efficacy with high levels of worry, and high levels of self-efficacy with low worry when he was searching for format-related effects among the interpretive data. Second, the students cited the presence of worrisome thoughts, or thoughts that are best conceptualised as worry, when they were asked to describe, in the survey and the interpretive designs, effects associated with their lack of confidence. Moreover, the interpretive data also indicated that confidence and overconfidence in tests were characterised by thoughts and feelings of well-being and the absence of worrisome thoughts. Finally, from the quasi-experimental design, worry was inversely related to item-specific self-efficacy — 11% of the variations in multiple-choice self-efficacy and 15% of the variations in constructed-response self-efficacy were associated with students' format-related worrisome thoughts. Overall, item-specific self-efficacy and worry were moderately

correlated with each other. Hence, it is quite clear from the extent of this overlap that high self-efficacy cannot be regarded simply as the absence of worry, or that worry provides an adequate description for low self-efficacy. There is a paucity of studies that have directly investigated the relationship between these two variables because it appears that most researchers are more interested in the relationship that these variables have with achievement rather than with each other. Nevertheless, the studies that have incorporated both of these variables (for example, Benson et al., 1994; Pajares & M. Johnson, 1994) report inverse relationships between test anxiety and self-efficacy that are consistent with this study.

Application of the Theoretical Model

The model underpinning this study presumes that boys' and girls' item-specific self-efficacies and the extent of their worrisome cognitions provide a heuristic for understanding their patterns of achievement according to test item-response formats. However, according to this model, the boys should have achieved higher scores than the girls for each of the item-response formats because they perceived higher item-specific self-efficacies and less worry than the girls. However, no practically important gender-related differences in achievement were detected for either of the response formats, even though the gender-related differences in item-specific self-efficacy and worry were practically significant. Instead, the girls' and the boys' achievement scores mirrored the similarities in their IQ scores and their prior science achievement.

Several explanations can be advanced for the failure of the model to correctly predict the boys' and the girls' relative achievements. The gender-related differences in students' self-reported, item-specific self-efficacies and the extent of their worrisome cognitions could have reflected real differences between the sexes. In this case, the influence of these differences may have been ameliorated by gender-related variables that are not described in this model. One such variable could relate to the girls' relatively greater tendency than the boys to engage in self-talk and positive thinking during tests. Therefore, the deleterious influence of the girls' lower self-efficacies and the higher frequency of their worrisome cognitions may have been ameliorated. Although appealing, this explanation appears to be flawed because there

is scant evidence from this study and other research (Bandura, 1997; Prins et al., 1994; Zatz & Chassin, 1985) that these strategies have much success.

Another possible explanation could relate to the more than ample time that was allocated for the achievement tests. The deleterious influences of the girls' lower self-efficacies and the higher frequency of their worrisome cognitions may have been ameliorated because sufficient time was available for them to complete the test to the level of their capabilities. This explanation is supported by evidence from the interpretive design, which indicated that the students who harbour low perceptions of their capabilities make a lower quality engagement with test items, that they work through tests more slowly than students with higher self-efficacies, and that they may fail to finish tests. Although Bridgeman and Schmitt (1997) report from their review that the speededness of tests does not appear to explain gender-related differences in test scores, it must be noted that Bridgeman and Schmitt were unwilling to generalise beyond the studies they reviewed, and that none of the studies they consulted investigated the role of self-efficacy. Additional research could profitably investigate the relative achievements for students of varying levels of self-efficacy in speeded and non-speeded tests to test the plausibility of this explanation.

Alternatively, the girls in this study may have under-reported their self-efficacy and over-reported the frequency of their worrisome thoughts relative to the boys. It is important to note here that Willingham and Cole (1997a) warn about inferring gender-related differences from self-reported data. Thus it may have been that the boys and the girls responded to the research instruments in the ways that they perceived were acceptable for them to do so. If this was the case, then their responses would have reflected prevailing notions that the study of science and competitive situations are male domains (Kelly, 1985; Lenney, 1977; Linn & Hyde, 1989; Mednick & Thomas, 1993), together with the generalised stereotypes that girls are submissive, timid and diffident, and that boys are aggressive, ebullient and confident (Easlea, 1986; Kelly, 1985; P. Murphy, 1988).

Overall, it is difficult to ascertain the veracity of the most likely of these explanations, or whether one or more of these explanations are implicated in this study. These difficulties in reconciling the data gathered in this study with the model underpinning this study underscore the problems associated with relying on self-

reported data, and they point to the need for other ways of measuring worry and self-efficacy. However, notwithstanding the previous discussion, the model successfully explains the girls' and the boys' patterns of achievement when interactions between gender and item-response format are considered. According to this approach, the absence of any practically important interactions between gender and test item-response formats for self-efficacy and worry predicted the absence of a similar interaction for the boys' and the girls' achievement scores.

IMPLICATIONS AND RECOMMENDATIONS

Implications for Research Methodology

This study is methodologically significant because it has gone beyond the integration of qualitative and quantitative data to a dialogical encounter (Bernstein, 1991) between the positivist and the interpretivist-constructivist research paradigms. In doing so, this study has moved beyond the inherently positivist approach that characterises many mixed-method designs (Guba & Lincoln, 1989). As mentioned previously, in Chapter 3, very few single research studies appear to have been designed around principles consistent with Bernstein's (1991) notion of dialogical encounter. Additionally, this study raises issues relating to the measurement of self-efficacy.

As outlined earlier, the significance of the methodology employed by this study is located not in the techniques that were used for collecting qualitative and quantitative data, but rather, in the dialogical use of methods from each of the research traditions. In this sense the design of the overall study differs from most mixed-method designs (see Guba & Lincoln, 1989, pp.160-162; Miles and Huberman, 1994, pp.41-42), which integrate qualitative and quantitative data without giving due regard to epistemological matters (J. Smith & Heshusius, 1986). For example, in mixed-method designs, qualitative data may be consulted to provide a description of the context within which the 'more rigorous' quantitative findings may be interpreted; qualitative data may be collected from representative subjects to 'make real' or illustrate what has been uncovered at a more general level in an experimental design

or survey design; quantitative data may be used for the selection of representative samples for interviewing so that generalisations can be made about populations; and, qualitative data may be collected in discovery/verification designs which use quantitative methods to test hypotheses that emerge from analyses of qualitative data. Ultimately, these types of research designs are concerned with making sample-to-population generalisations, and consequently, they subsume qualitative data into positivist frameworks (Guba & Lincoln, 1989; J. Smith & Heshusius, 1986). Moreover, these types of designs, which purport to occupy the middle ground between the research paradigms (Miles & Huberman, 1994), confuse research methods (that is, strategies derived from particular epistemologies) with research techniques (Guba & Lincoln, 1989).

This study differed from mixed-method designs in that it attempted to do justice to each of the positivist and interpretivist-constructivist research traditions and then discuss the findings using Bernstein's (1991) notions of dialogical encounter. Accordingly, this study utilised methodologies for collecting positivist data (both qualitative and quantitative) and interpretivist-constructivist data within separate frameworks, which were analysed separately from each other in Chapters 4 and 5. Dialogical encounter, as operationalised in this study, was focused around the research questions guiding this study and it involved searching for commonalities, together with points of difference and conflict when discussing findings from each of the research traditions. The outcome of this dialogue was reported as a series of working hypotheses.

Like mixed-method designs, the approach taken in this study enabled a comprehensive view of the inherently complex research questions. The positivist design provided a frame of reference (through the findings from the coarse-grained data) and the interpretivist-constructivist design provided some of the finer detail. For example, the positivist design was essentially theory testing to determine if the model had any explanatory power in relation to the stated research questions and the emergent nature of the interpretivist-constructivist design led to new insights about overconfidence and the quality of students' engagement with test items during tests. Additionally, there is little doubt in the mind of this researcher that the 'binocular' (Reichardt & Rallis, 1994b) view of the research questions gained by using qualitative

and quantitative data provided a richer and more comprehensive understanding of the research questions than could have been obtained by using only one of these types of data. For example, quantitative data were collected to compare girls' and boys' item-specific self-efficacies and correlate them with achievement, and qualitative data were collected to describe the content of their self-efficacies, the meanings that students attach to their self-efficacies during tests, and how their self-efficacies are related to achievement through achievement-related behaviours.

The design of this study posed one advantage over mixed-method designs although its implementation was at times problematic. In the first instance, the significance of the design used in this study is located in the attention it gave to each of the research traditions rather than with a pragmatic concern for results. Smith and Heshusius (1986, p.10) assert that the idea 'research decisions can be made on the basis of what works is not the bottom line.' Instead, in their view, together with the view of other 'conceptual purists' (for example; Guba & Lincoln, 1989) and this researcher, what works must take into account one's adherence to paradigmatic assumptions about the nature of the world and the purpose, role and approach of the researcher. In this study the integrity of each of the research traditions was maintained instead of blending them with each other or making one subservient to the other. However, this researcher experienced some difficulty maintaining the focus of this present study, about which Bogdan and Biklen (1992) warn, in that it has tended to become a study in method as well as the original research topic. Further, this study required considerably extra attention to methodological matters than studies located in only one of the research traditions to avoid becoming 'a zooey mixture that undermines the quality [of either research tradition]' (S. Miller et al., 1998, p.397), and therefore compromise the quality of the entire study.

Additionally, questions related to the generalisability of the entire study created a tension between the status of the two research traditions. First, there was the question whether notions of generalisability be drawn from the positivist tradition or from the interpretive tradition? Second, there was the question of what consequences are associated with making such a choice? It seems that these questions are not addressed in the literature because dialogical encounter represents one way that researchers and philosophers from differing traditions can communicate with each

other rather than a method for discussing findings within a single study. Nevertheless, the choice to locate notions of generalisability for the entire study within the interpretivist-constructivist tradition was relatively easy given the unrepresentativeness of the students who participated in this study. However, this decision begs the question of whether this choice locates this study, ultimately, within the interpretivist-constructivist paradigm (and therefore does an injustice to the notion of dialogical encounter). It would seem, to this researcher, that notions of generalisability must be reconceptualised before genuine dialogical encounter can be achieved between the research traditions. Although this weighty issue is important and stimulating, it must be left for others to debate because additional discussion is well beyond the scope of this study.

This research also sheds some light on the measurement of self-efficacy and worry. Although the procedures for measuring these variables were straightforward, interpretation of boys' and girls' scores was problematic. One limitation of this study was its reliance on self-report measures for self-efficacy and worry. This researcher echoes Pajares' (1997) recent caution that, 'One must not confuse differences (or lack of differences) on the metric of a self-efficacy scale as differences in confidence,' and Pajares' advice that self-efficacy researchers need to develop ways of measuring and describing self-efficacy through direct observation. Clearly, additional research is required to develop and pilot categories so that observational data for students working during tests can be interpreted alongside their self-reported data for either worry or self-efficacy.

Implications for Theory

As outlined earlier in this chapter, the findings from this study are presented as a series of working hypotheses rather than as new theory. Consequently, this study provides additional support for other recent findings in the area of gender, achievement and test item-response formats, and it suggests several avenues that may be fruitfully explored for the development of new theory.

This study provides additional support for the recently reported finding (Willingham & Cole, 1997a) that there is no interaction between gender and item-

response formats for achievement on parallel multiple-choice and short-answer constructed-response items. However, as outlined in Chapter 3, one of the chief limitations of this study related to the operationalisation of constructed-response items. Although very few science teachers in Queensland schools, according to this researcher's experience, write test items that require extended-answer responses, it would be profitable to conduct additional research to compare girls' and boys' achievements, self-efficacies and the extent of their worrisome cognitions for multiple-choice/short-answer response formats with extended-answer response formats, where the literature (for example, Breland et al., 1991; Harding, 1981; Whitehouse & Sullivan, 1990; Willingham & Cole, 1997a) appears to locate boys' and girls' differential patterns of achievement.

This study is also theoretically significant because the model that was proposed to explain girls' and boys' patterns of achievement for test item-response formats provided a tenable description for the absence of an interaction between gender and test item-response formats, despite the uncertainties associated with the self-reported data. Additionally, the descriptions that this study provided for the role of self-efficacy beliefs during testing are highly significant. First, little research appears to have provided qualitative descriptions for the relationship between self-efficacy beliefs and achievement-related behaviours during tests — the descriptions contained within this study provide a greater understanding of how self-efficacy and achievement are related, and they pave the way for additional research into strategies to counter inappropriate efficacy appraisals during tests. Second, the mediating influence of students' engagement with test items provides additional support for Bandura's (1986, 1997) conceptualisation for the relationship between self-efficacy beliefs and achievement.

This study also raises questions about the role of overconfidence during tests. Very little appears to have been reported on this aspect of self-efficacy judgements in the literature. Although Bandura (1997) asserts that human functioning is best served by unnecessarily optimistic self-efficacy appraisals, additional research could investigate to what extent self-efficacies can be elevated before they become associated with deleterious effects.

Implications for Educational Practice

As introduced in Chapter 1, much of the debate relating to test item-response formats is scathing in its criticisms of multiple-choice testing and it lauds the perceived strengths of constructed-response and performance assessments. This study is able to contribute toward this debate in relation to the claim that there is a multiple-choice effect which confers an advantage to boys (Harding, 1994; P. Murphy, 1988, 1991, 1996) in the comparisons it draws between multiple-choice and short-answer response formats. It is important to emphasise in this context that this study did not investigate the use of extended-answer formats.

This study failed to detect any reasons to suggest that short-answer constructed-response formats are inherently superior to multiple-choice response formats. There were no practically important gender-by-format interactions for either self-efficacy, worry or achievement, and each of the response formats had particular advantages and disadvantages associated with their use. The finding, that there were no gender-related consequences associated with choices between the two item-response formats is of central importance to this study. However, it must be recognised that this study cannot comment on whether both of these response formats advantage boys, whether they both advantage girls, or whether they are fair to both sexes. Nevertheless, this study shows that there are other implications associated with the selection of one of these response formats over the other. First, the short-answer questions were more difficult than the equivalent multiple-choice items and they had greater discriminatory power, which was at the expense of the greater difficulty of scoring these items. (This researcher had to score all the constructed-response items himself to ensure comparability between the research groups, even though the cooperating teachers were given detailed marking schedules for each of the tests). The literature (for example, Gronland, 1993; Mehrens & Lehmann, 1984) is replete with the advice that multiple-choice items are superior to constructed-response formats for their greater ease, efficiency and lower cost of scoring responses. Second, item format restricts the range of constructs that can be measured. Part of the experience of this researcher is reflected in Frederiksen's (1984) assertion that it is easier to cast multiple-choice items into short-answer formats than it is to cast short-

answer items into multiple-choice formats. Therefore, short-answer items appear to be superior to multiple-choice items in that they can be used more easily to tap a wider range of constructs (Childs, 1990; Ebel, 1979; Frederiksen, 1984; Martinez, 1998), and they are certainly preferable to poorly constructed multiple-choice items that introduce construct-irrelevant variance by forcing subject matter competencies into an inappropriate format. Moreover, it is important to heed the advice that item-response formats should be congruent with the purposes of the test and the constructs that are assessed (Martinez, 1998; Mehrens, 1998; J. Powell & Gillespie, 1990).

Additionally, this study suggests that time constraints are a potential source of invalidity in students' test scores. According to the findings from this study, students who harbour low perceptions of their capabilities are denied the opportunity to achieve at the level of their capabilities when time is constrained because they are unable to devote sufficient attention to all the test items. Moreover, the desirability that students have sufficient time during tests is underscored by reports that answer changing tendencies are associated with higher achievement scores (Bridgeman & Schmitt, 1997). Students can only change their answers when they have the opportunity to engage with test items beyond their initial impressions, which is most likely to occur toward the end of tests when students go back to check their work. Although individual students require different amounts of time for completing their tests, untimed tests may be neither practicable nor desirable in class settings. Instead, it is incumbent for teachers to identify the point of diminishing returns beyond which the allocation of extra time is unlikely to bring about any appreciable benefits to any of the students in their classes. This judgement should not be too problematic for tests of content knowledge and simple understanding, although it is likely to be more difficult for tests of complex reasoning processes (Bridgeman and Schmitt, 1997).

As outlined in Chapter 1, there is some evidence to suggest that girls' decisions to enrol in science courses and pursue scientific careers are linked to their self-efficacies. Therefore, the gender-related differences in self-efficacy that were detected in this study have clear implications for educational researchers and science teachers, even though it is not clear (i) to what extent these differences arise from sociocultural factors, schooling, the study of science or the competitiveness associated with test settings; (ii) whether the girls, as a group, underestimated their self-efficacy or

whether the boys made overly optimistic appraisals of their capabilities; and, (iii) whether these gender-related differences in self-efficacy reflect real differences in boys' and girls' perceptions of their capabilities or whether they reflect differences in girls' and boys' perceptions about what they should report. Some evidence was presented in this study that gender-related differences in self-efficacy may be associated with instructional and teacher-related variables at the classroom level, which suggests that they are responsive to change. Hence, it is important that all science teachers receive adequate in-service training so that they can act to transform, rather than allow the reproduction of the gender identities that are associated with the study of science (Kelly, 1985). Social cognitive theory (Bandura, 1986) suggests that interventions to enhance girls' self-efficacy beliefs relative to the boys must include the provision of authentic mastery experiences in the classroom, the use of verbal persuasion (which must not be confused with patronising girls) and exposure to vicarious influences that highlight the achievements of successful girls and female role models. It is also important to heed Kelly's (1985) work, which suggests that effective interventions cannot address girls' lower self-efficacies without addressing classroom processes whereby boys dominate class activities and act in ways that put down girls and their accomplishments. Hence, interventions must also provide teachers with strategies for managing science classes so that all students can participate equally in the learning activities. The importance of such interventions is underscored by the trend that more Australian students are completing high school than ever before, which makes girls increasingly dependent on school outcomes for their futures (Teese et al., 1995).

CONCLUDING COMMENTS

This study has determined, for the students who participated in this research, that the boys and girls gave similar reasons to each other for the content of their self-efficacies and their worrisome cognitions as they relate to multiple-choice and short-answer constructed-response test items. However, the girls reported less self-efficacy and more worry than the boys for each of these response formats. Nevertheless, there were no practically important gender-by-format interactions, for achievement, worry

or self-efficacy. The model that was developed for this study to explain girls' and boys' patterns of achievement according to test item-response formats was supported by the data, despite some uncertainties that were associated with the self-reported data. Moreover, the students imputed a causal role to their self-efficacy perceptions insofar as their achievement was concerned, which appeared to be mediated through the quality of their engagement with test items.

The findings from this study suggest that there are no practically important consequences for boys' and girls' relative achievements associated with the choice between multiple-choice and short-answer constructed-response formats. It is recommended, however, that sufficient time is provided during testing so that all students have the opportunity to complete tests to the extent of their capabilities. Additionally, gender-related differences in girls' and boys' perceptions of their capabilities in science need to be addressed by way of additional research and the professional development of science teachers. The importance of addressing this issue is underscored by Willingham and Cole's (1997d, p.357) conclusion, after they finished reviewing the research literature on gender and fair assessment, that

Young people, regardless of gender, develop talent and proficiency in particular areas when they devote the time it takes. Programs that encourage less gender-specific experience, training and careers have succeeded to a remarkable degree, and more of that effort will be needed ... [because] equal opportunity cannot be ensured through passive means. If young men and women are not encouraged and accepted into fields of work and study at levels beyond current underrepresented norms, then their opportunity in those fields is likely to be defined by the status quo.

REFERENCES

- Adkins, D.C. (1974). *Test construction* (2nd ed.). Columbus, OH: Merrill.
- Ajzen, I. (1988). *Attitudes, personality and behaviour*. Milton Keynes: Open University Press.
- Alpert, R. & Haber, R. (1960). Anxiety in academic achievement situations. *Journal of Abnormal and Social Psychology*, 61, 207-215.
- Anastasi, A. (1958). *Differential psychology* (3rd ed.). New York: Macmillan.
- Anderson, J. (1989). Sex-related differences on objective tests among undergraduates. *Educational Studies in Mathematics*, 20, 165-177.
- Archenhold, F.W. (1988). *Science at age 15: A review of APU findings 1980-84*. London: HMSO.
- Armstrong, A. (1993). Cognitive style differences in testing situations. *Educational Measurement: Issues and Practice*, 12, 17-22.
- Arndt, B. (1994, February 19-20). Gender wars in class. *The Weekend Australian*, Weekend Review, p.1.
- Australian Council for Educational Research (1978). *Australian item bank: Science items* (Books 2 & 3). Hawthorn, Victoria: Australian Council for Educational Research.
- Australian Council for Educational Research (1982). *ACER Intermediate Test G: Manual for administration and interpretation*. Hawthorn, Victoria: Australian Council for Educational Research.
- Australian Science Teachers Association. (1984). *The Dircks report: The place of science in the K-12 curriculum of Australian schools*. Canberra Publishing and Printing: Canberra.
- Baker, D.R. (1998). Equity issues in science education. Parker, L.H. & Rennie, L.J. (1998). In B.J. Fraser & K.G. Tobin (Eds.), *International Handbook of Science Education* (pp. 869-895). Dordrecht: Kluwer.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84, 191-215.
- Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist*, 37, 122-147.
- Bandura, A. (1984). Recycling misconceptions of perceived self-efficacy. *Cognitive*

- Therapy and Research*, 8, 231-255.
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall
- Bandura, A. (1989). Regulation of cognitive processes through personal perceived self efficacy. *Developmental Psychology*, 25, 729-735.
- Bandura, A. (1991). Self-efficacy conception of anxiety. In R. Schwarzer & R.A. Wicklund (Eds.), *Anxiety and self-focused attention* (pp. 89-123). Chur: Harwood.
- Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist*, 28, 117-148.
- Bandura, A. (1995). On rectifying conceptual ecumenicism. In J.E. Maddux (Ed.), *Self-efficacy, adaptation, and adjustment* (pp. 347-375). New York: Plenum Press.
- Bandura, A. (1997). *Self-Efficacy: The exercise of control*. New York: Freeman.
- Becker, P. (1982). Towards a process analysis of test anxiety: Some theoretical and methodological observations. In R. Schwarzer, H.K. van der Ploeg & C.D. Spielberger (Eds.), *Advances in test anxiety research* (pp. 11-17). Lisse: Swets and Zeitlinger.
- Bell, R.G. & Hay, J.A. (1987). Differences and biases in English language examination formats. *British Journal of Educational Psychology*, 57, 212-220.
- Benbow, C.P. & Stanley, J.C. (1980). Sex differences in mathematical ability: Fact or artifact? *Science*, 210, 1262-1264.
- Bennett, R.E. (1993). On the meanings of constructed response. In R.E. Bennett & W.C. Ward (Eds.), *Construction versus choice in cognitive measurement: Issues in constructed response, performance testing and portfolio assessment* (pp. 213-240). Hillsdale, NJ: Erlbaum.
- Bennett, R.E., Rock, D.A. & Wang, M. (1991). Equivalence of free-response and multiple-choice items. *Journal of Educational Measurement*, 28, 77-92.
- Ben-Shakhar, G. & Sinai, Y. (1991). Gender differences in multiple-choice tests: The role of differential guessing tendencies. *Journal of Educational Measurement*, 28, 23-35.
- Benson, J. (1989). Structural components of statistical test anxiety in adults: An exploratory study. *Journal of Experimental Education*, 57, 247-261.

- Benson, J., Bandalos, D. & Hutchinson, S. (1994). Modeling test anxiety among men and women. *Anxiety, Stress, and Coping*, 7, 131-148.
- Bernstein, R.J. (1991). *The new constellation: The ethical-political horizons of modernity/postmodernity*. Cambridge, UK: Polity Press.
- Berry, M. & West, R.L. (1993). Cognitive self-efficacy in relation to personal mastery and goal setting across the lifespan. *International Journal of Behavioral Development*, 16, 351-379.
- Beyer, S. (1990). Gender differences in the accuracy of self-evaluations of performance. *Journal of Personality and Social Psychology*, 59, 960-970.
- Birenbaum, M. & Nasser, F. (1994). On the relationship between test anxiety and test performance. *Measurement and Evaluation in Counseling and Development*, 27, 293-301.
- Blankstein, K.R., Flett, G.L., Boases, P & Toner, B.B (1991). Thought listing and endorsement measures of self-referential thinking in test anxiety. In R. Schwarzer & R.A. Wicklund (Eds.), *Anxiety and self-focused attention* (pp. 133-142). Chur: Harwood.
- Bloom, B.S. (1956). *A taxonomy of educational objectives: Handbook I, the cognitive domain*. New York: Longman.
- Bloom, B.S. (1976). *Human characteristics and school learning*. New York: McGraw-Hill.
- Blum, A & Azencot, M. (1986). Multiple-choice versus equivalent essay questions in a national examination. *European Journal of Science Education*, 8, 225-228.
- Bogdan, R.C. & Biklen, S.K. (1992). *Qualitative research for education: An introduction to theory and methods* (2nd ed.). Boston: Allyn & Bacon.
- Bolger, N. (1984, August). *Gender difference in academic achievement according to method of measurement*. Paper presented at the 92nd annual convention of the American Psychological Association, Toronto, Canada.
- Bolger, N. & Kellaghan, T. (1990). Method of measurement and gender differences in scholastic achievement. *Journal of Educational Measurement*, 27, 165-174.
- Bouffard-Bouchard, T., Parent, S. & Larivee, S. (1991). Influence of self-efficacy on performance in a cognitive task. *Journal of Social Psychology*, 130, 353-363.
- Bracey, G.W. (1990). Teachers, thinking and testing. *Phi Delta Kappan*, 71, 404-407.

- Bransky, J. & Qualter, A. (1993). Applying physics concepts - Uncovering the gender differences in Assessment of Performance Unit results. *Research in Science and Technological Education, 11*, 141-151.
- Breland, H.M., Danos, D.O., Kahn, H.D., Kubota, M.Y. & Sudlow, M.W. (1991). *A study of gender and performance on Advanced Placement history examinations* (Report No. 91-4). New York: College Entrance Examination Board.
- Breland, H.M. & Griswold, P.A. (1982). Use of a performance test as a criterion in a differential validity study. *Journal of Educational Psychology, 74*, 713-721.
- Bridgeman, B. & Lewis, C. (1994). The relationship of essay and multiple-choice scores with grades in college courses. *Journal of Educational Measurement, 31*, 37-50.
- Bridgeman, B. & Schmitt, A. (1997). Fairness issues in test development and administration. In W.W. Willingham & N.S. Cole (Eds.), *Gender and fair assessment* (pp. 185-226). Mahwah, NJ: Erlbaum.
- Briggs, D. (1980). A study of the influence of handwriting upon grades using examination scripts. *Educational Review, 32*, 183-193.
- Bruch, M.A. (1981). Relationship of test-taking strategies to test anxiety and performance: Toward a task analysis of examination behavior. *Cognitive Therapy and Research, 5*, 41-56.
- Bruch, M.A., Juster, H.R. & Kaflowitz, N.G. (1983). Relationships of cognitive components of test anxiety to test performance: Implications for assessment and treatment. *Journal of Counseling Psychology, 30*, 527-536.
- Bruch, M.A., Kaflowitz, N.G. & Kuethe, M. (1986). Beliefs and the subjective meaning of thoughts: Analysis of the role of self-statements in academic test performance. *Cognitive Therapy and Research, 10*, 51-69.
- Burkam, D.T. & Burkam, A.S. (1995). Is item format important? In D.R. Baker & K. Scantlebury (Eds.), *Science "coeducation": Viewpoints from gender, race and ethnic perspectives*. NARST monograph No. 7, 140-159.
- Burkhalter, B.B. & Schaer, B.B. (1984-85). The effect of cognitive style and cognitive learning in a nontraditional educational setting. *Educational Research Quarterly, 9* (4), 12-18.
- Burnett, S.A. (1986). Sex-related differences in spatial ability: Are they trivial? *American Psychologist, 41*, 1012-1014.

- Burstein, B., Bank, L., & Jarvik, L.F. (1980). Sex differences in cognitive functioning: Evidence, determinants, implications. *Human Development*, 23, 289-313.
- Byrne, B.M. & Shavelson, R.J. (1987). Adolescent self-concept: Testing the assumptions of equivalent structure across gender. *American Educational Research Journal*, 24, 365-385.
- Caplan, P.J., MacPherson, G.M. & Tobin, P. (1985). Do sex-related differences in spatial abilities exist? *American Psychologist*, 40, 786-789.
- Carver, C.S. & Scheier, M.F. (1991). A control-process perspective on anxiety. In R. Schwarzer & R.A. Wicklund (Eds.), *Anxiety and self-focused attention* (pp. 3-8). Chur: Harwood.
- Childs, R.A. (1990). *Gender bias and fairness* (Report No. EDO-TM-90-9). Washington, DC: American Institutes for Research. (ERIC document reproduction Service No. ED 328 610).
- Chopin, B. (1975). Guessing the answer on objective tests. *British Journal of Educational Psychology*, 45, 206-213.
- Cizek, G.J. (1993). Rethinking psychometricians' beliefs about learning. *Educational Researcher*, 22(4), 4-9.
- Cohen, J. (1969). *Statistical power analysis for behavioral sciences*. New York: Academic Press.
- Collis, K.F. (1992). Curriculum and assessment: A basic cognitive model. In G. Leder (Ed.), *Assessment and learning of mathematics*. Australian Council for Educational Research: Hawthorn, Victoria.
- Commonwealth Schools Commission (1984). *Girls and tomorrow: The challenge for schools*. Commonwealth Schools Commission: Canberra, Australian Capital Territory.
- Coopersmith, S. (1967). *The antecedents of self-esteem*. San Francisco: Freeman.
- Coorey, M. (1998, October 17-18). Year 12 exams a sham: Academic. *The Weekend Australian*, p.3.
- Covington, M.V. (1992). *Making the grade*. Cambridge: Cambridge University Press.
- Crocker, L. & Schmitt, A. (1987). Improving multiple-choice test performance for examinees with different levels of test anxiety. *The Journal of Experimental*

- Education*, 55, 201-205.
- Crockett, G.V. (1996). Shifts in gender segregation in professional occupations in Australia - 1981 to 1991. *Australian Bulletin of Labour*, 22, 265-274.
- Crooks, T.J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, 58, 438-481.
- Crow, L.W. & Piper, M.K. (1986). A study of field independent-biased mental ability tests in community college science classes. *Journal of Research in Science Teaching*, 23, 817-822.
- Culler, R.E. & Holahan, C.J. (1980). Test anxiety and academic performance: The effects of study-related behaviors. *Journal of Educational Psychology*, 72, 16-20.
- Czerniak, C. & Chiarelott, L. (1994, April). *Science anxiety: An investigation of science achievement, sex and grade level factors*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
- Davis, J.K. (1991). Educational implications of field dependence-independence. In S.J. Wapner & J. Demick (Eds.), *Field dependence-independence: Cognitive styles across the lifespan* (pp. 149-175). Hillsdale, NJ: Erlbaum.
- De Laeter, J.R., Malone, J. & Dekkers, J. (1989). Female science enrolment trends in Australian senior secondary schools. *Australian Science Teachers Journal*, 35(3), 23-33.
- Deffenbacher, J.L. (1977). Relationship of worry and emotionality to performance on the Miller Analogies Test. *Journal of Educational Psychology*, 69, 191-195.
- Deffenbacher, J.L. (1978). Worry, emotionality, and task-generated interference in test anxiety: An empirical test of attentional theory. *Journal of Educational Psychology*, 70, 248-254.
- Deffenbacher, J.L. (1980). Worry and emotionality in test anxiety. In I.G. Sarason (Ed.), *Test Anxiety: Theory, research and applications* (pp. 111-128). Hillsdale, NJ: Erlbaum.
- Demick, J. (1991). Organismic factors in field dependence-independence: Gender, personality, psychopathology. In S.J. Wapner & J. Demick (Eds.), *Field dependence-independence: Cognitive styles across the lifespan* (pp. 209-243). Hillsdale, NJ: Erlbaum.

- Denzin, N.K. (1988). Triangulation. In J.P. Keeves (Ed.), *Educational research, methodology, and measurement: An international handbook* (pp. 511-513). Sydney: Pergamon Press.
- DeVito, A.J. (1988). Test anxiety inventory. In D.J. Kayser & R.C. Sweetland (Eds.), *Test critiques*, Vol. 1 (pp. 673-681). Kansas City, MO: Test Corporation of America.
- Doctor, R.M. & Altman, F. (1969). Worry and emotionality as components of test anxiety: Replication and further data. *Psychological Reports*, 24, 563-568.
- Donmoyer, R. (1996). Educational research in an era of paradigm proliferation: What's a journal editor to do? *Educational Researcher*, 25(2), 19-25.
- Donmoyer, R. (1997, March). *The qualitative/quantitative distinction: Is it still a matter of different epistemologies?* Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.
- Doran, R.L., Lawrenz, F. & Helgeson, S. (1994). Research on assessment in science. In D. Gabel (Ed.), *Handbook on research on science teaching and learning* (pp. 388-442). New York: Macmillan.
- Eagly, A.H. (1987). Reporting sex differences. *American Psychologist*, 42, 756-757.
- Easlea, B. (1986). The masculine image of science with special reference to physics: How much does gender really matter? In J. Harding (Ed.), *Perspectives on gender and science* (pp. 132-158). London: Falmer Press.
- Ebel, R.L. (1979). *Essentials of educational measurement* (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Eccles Parsons, J., Adler, T.F. & Kaczala, C.M. (1982). Socialization of achievement attitudes and beliefs: Parental influences. *Child Development*, 53, 310-321.
- Eccles Parsons, J., Adler, T.F. & Meece, J.L. (1984). Sex differences in achievement: A test of alternate theories. *Journal of Personality and Social Psychology*, 46, 26-43.
- Elwood, J & Comber, C. (1995). Gender differences in A level examinations: New complexities or old stereotypes?. *British Journal of Curriculum and Assessment*, 6(1), 24-28.
- Erickson, F. (1986). Qualitative methods in research on teaching. In M.C. Wittrock (Ed.), *Handbook of research on teaching* (3rd. ed.) (pp. 119-161). New York: Macmillan.

- Erickson, G.L. & Erickson, L.J. (1984). Females and science achievement: Evidence, explanations, and implications. *Science Education*, 68, 63-89.
- Erickson, G. & Farkas, S. (1991). Prior experience and gender differences in science achievement. *The Alberta Journal of Educational Research*, 37, 225-239.
- Everson, H., Millsap, R.E. & Browne, J. (1989). Cognitive interference or skills deficit: An empirical test of two competing theories of test anxiety. *Anxiety Research*, 1, 313-325.
- Everson, H.T., Millsap, R.E. & Rodriguez, C.M. (1991). Isolating gender differences in test anxiety: A confirmatory factor analysis of the Test Anxiety Inventory. *Educational and Psychological Measurement*, 51, 243-251.
- Eysenck, H. (1978). An exercise in mega-silliness. *American Psychologist*, 33, 517.
- Eysenck, M.W. (1982). *Attention and arousal: Cognition and performance*. New York: Springer-Verlag.
- Feingold, A. (1988). Cognitive gender differences are disappearing. *American Psychologist*, 43, 95-103.
- Fennema, E.H. & Sherman, J. (1977). Sex-related differences in mathematics achievement, spatial visualization and affective factors. *American Educational Research Journal*, 14, 51-71.
- Fennema, E.H. & Sherman, J.A. (1978). Sex-related differences in mathematics achievement and related factors: A further study. *Journal for Research in Mathematics Education*, 9, 189-203.
- Ferguson, G.A. & Takane, Y. (1989). *Statistical analysis in psychology and education* (6th ed.). New York: McGraw Hill.
- Firestone, W.A. (1987). Meaning in method: The rhetoric of quantitative and qualitative research. *Educational Researcher*, 16(7), 16-21.
- Fishbein, M. & Ajzen, I. (1975). *Belief, attitude, intention and behavior: An introduction to theory and research*. Reading, MA: Addison-Wesley.
- Flaughter, R.L. (1978). The many definitions of test bias. *American Psychologist*, 33, 671-679.
- Fleming, M.L. & Malone, M.R. (1983). The relationship of student characteristics and student performance in science as viewed by meta-analysis research. *Journal of Research in Science Teaching*, 20, 481-495.
- Fraser, B.J. (1982). How strongly are attitude and achievement related? *School*

- Science Review*, 63, 557-559.
- Frederiksen, N. (1984). The real test bias: Influences of testing on teaching and learning. *American Psychologist*, 39, 193-202.
- Fullarton, S (1993). *Confidence in mathematics: The effect of gender* (Research Monograph No. 2). Geelong, Victoria: Faculty of Education, Deakin University National Centre for Research and Development in Mathematics Education.
- Furst, D., Tenebaum, G. & Weingarten, G. (1985). Test anxiety, sex and exam type. *Psychological Reports*, 56, 663-668.
- Gable, R.K. & Wolf, M.B. (1993). *Instrument development in the affective domain: Measuring attitudes and values in corporate and school settings*. Boston, MA: Kluwer.
- Gage, N.L. (1989). The paradigm wars and their aftermath. *Educational Researcher*, 18(7), 4-10.
- Galassi, J.P., Frierson, H.T. Jr. & Sharer, R. (1981). Behavior of high, moderate, and low test anxious students during an actual test situation. *Journal of Consulting and Clinical Psychology*, 49, 51-62.
- Ganzer, V.J. (1968). Effects of audience presence and test anxiety on learning and retention in a serial learning situation. *Journal of Personality and Social Psychology*, 8, 194-199.
- Geen, R.G. (1987). Test anxiety and behavioral avoidance. *Journal of Research in Personality*, 21, 481-488.
- Gellman, E.S. & Berkowitz, M. (1993). Test-item type: What students prefer and why. *College Student Journal*, 27, 17-26.
- Gipps, C.V. (1994). *Beyond testing: Towards a theory of educational assessment*. London: Falmer.
- Gipps, C. & Murphy, P. (1994). *A fair test? Assessment, achievement and equity*. Milton Keynes: Open University Press.
- Glenberg, A.M. & Epstein, W. (1987). Inexpert calibration of comprehension. *Memory and Cognition*, 15, 84-93.
- Gorman, T.P., White, J., Brooks, G., Maclure, M. & Kispal, A. (1988). *Language performance in schools: Review of APU language monitoring 1979-1983*. London: HMSO.

- Gorrell, J. (1990). Some contributions of self-efficacy research to self-concept theory. *Journal of Research and Development in Education*, 23, 73-81.
- Gronland, N.E. (1993). *How to make achievement tests and assessments*. Boston: Allyn & Bacon.
- Gross, T.F. (1990). General test and state anxiety in real examinations: State is not test anxiety. *Educational Research Quarterly*, 14(3), 11-20.
- Guba, E.G. & Lincoln, Y.S. (1989). *Fourth generation evaluation*. Newbury Park, CA: Sage.
- Guthrie, J.T. (1984). Testing higher level skills. *Journal of Reading*, 28, 188-190.
- Hackett, G. & Betz, N.E. (1989). An exploration of the mathematics self-efficacy/mathematics performance correspondence. *Journal for Research in Mathematics Education*, 20, 261-273.
- Halpern, D.F. (1992). *Sex differences in cognitive abilities*. Hillsdale, NJ: Erlbaum.
- Hamilton, L. (1997, March). *Sources of gender differences on the NELS:88 constructed-response and multiple-choice science tests*. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.
- Hammermaster, C.S. (1989). Levels of performance and cognitive interference in test-anxious subjects. *The Alberta Journal of Educational Research*, 35, 164-170.
- Hammersley, M. (1998). Telling tales about educational research: A response to John K. Smith. *Educational Researcher*, 27(7), 18-21.
- Hanna, G. (1986). Sex differences in the mathematics achievement of eighth graders in Ontario. *Journal for Research in Mathematics Education*, 17, 231-237.
- Hansford, B.C. & Hattie, J.A. (1982). The relationship between self and achievement/performance measures. *Review of Educational Research*, 52, 123-142.
- Harding, J. (1981). Sex-differences in science examinations. In A. Kelly (Ed.), *The missing half* (pp. 192-204). Manchester: Manchester University Press.
- Harding, J. (July, 1994). *Gender and assessment*. Seminar presented at the McClintock Collective, Richmond, Victoria, Australia
- Harding, J. & Sutoris, M. (1987). An object relations account of the differential involvement of boys and girls in science and technology. In A. Kelly (Ed.), *Science for girls?* (pp. 24-36). Milton Keynes: Open University Press.

- Harter, S. (1982). The perceived competence scale for children. *Child Development*, 53, 87-97.
- Hedl, J.J., Jr. (1982, April). *A factor analytic study of the Test Anxiety Inventory and A-state worry and emotionality items from the State-Trait Anxiety Inventory (form Y)*. Paper presented at the annual meeting of the Southwestern Psychological Association, Dallas, TX.
- Hellekant, J. (1994). Are multiple-choice tests unfair to girls? *System*, 22, 349-352.
- Hembree, R. (1988). Correlates, causes, effects and treatment of test anxiety. *Review of Educational Research*, 58, 47-77.
- Hitchcock, G. & Hughes, D. (1989). *Research and the teacher*. London: Routledge.
- Hitchcock, G. & Hughes, D. (1995). *Research and the teacher* (2nd ed.). London: Routledge.
- Hogan, T.P. (1981). *Relationship between free-response and choice-type tests of achievement: A review of the literature*. Washington, DC: National Institute of Education. (ERIC document reproduction Service No. ED 224 811).
- Holroyd, K.A. & Appel, M.A. (1980). Test anxiety and physiological responding. In I.G. Sarason (Ed.), *Test anxiety: Theory, research and applications* (pp. 129-151). Hillsdale, NJ: Erlbaum.
- Holroyd, K.A., Westbrook, T., Wolf, M. & Badhorn, E. (1978). Performance, cognition, and physiological responding in test anxiety. *Journal of Abnormal Psychology*, 87, 442-451.
- Hoste, R. (1982). Sex differences and similarities on performance in a CSE biology examination. *Educational Studies*, 8, 141-153.
- Howe, K.R. (1988). Against the quantitative-qualitative incompatibility thesis or dogmas die hard. *Educational Researcher*, 17(8), 10-16.
- Hunsley, J. (1985). Test anxiety, academic performance and cognitive appraisals. *Journal of Educational Psychology*, 77, 678-682.
- Hyde, J.S. (1981). How large are cognitive gender differences? A meta-analysis using ω^2 and d. *Educational Studies*, 8, 141-153.
- Hyde, J.S. (1986). Introduction: Meta-analysis and the psychology of gender. In J.S. Hyde & M.C. Linn (Eds.), *The psychology of gender: Advances through meta-analysis* (pp. 1-13). Baltimore, MD: John Hopkins University Press.

- Hyde, J.S., Fennema, E. & Lamon, S.J. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin*, *107*, 139-155.
- Hyde, J.S., Fennema, E., Ryan, M., Frost, L.A. & Hopp, C. (1990). Gender comparisons of mathematics attitudes and affect. *Psychology of Women Quarterly*, *14*, 299-324.
- Hyde, J.S., Geringer, E.R. & Yen, W.M. (1975). On the empirical relation between spatial ability and sex differences in other aspects of cognitive performance. *Multivariate Behavioral Research*, *10*, 289-309.
- Hyde, J.S. & Linn, M.C. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin*, *104*, 139-155.
- Jackson, L.A., Hodge, C.N. & Ingram, J.M. (1994). Gender and self-concept: A reexamination of stereotypic differences and the role of gender attitudes. *Sex Roles*, *30*, 615-630.
- Jackson, W. (1995). *Methods: Doing social research*. Scarborough, Canada: Prentice-Hall.
- Johnson, S. (1987). Gender differences in science: Parallels in interest, experience and performance. *International Journal of Science Education*, *9*, 467-481.
- Jones, C. (1995, June 27). Exam code would let parents decide. *The Australian*, p.4.
- Jones, L.G. & Jones, L.P. (1989). Context, confidence and the able girl. *Educational Research*, *31*, 189-194.
- Jovanovic, J., Solano-Flores, G. & Shavelson, R.J. (1994). Performance-based assessments: Will gender differences in science achievement be eliminated? *Education and Urban Society*, *26*, 352-366.
- Kahle, J.B. & Lakes, M.K. (1983). The myth of equality in science classrooms. *Journal of Research in Science Teaching*, *20*, 131-140.
- Kahle, J.B. & Meece, J. (1994). Research on gender issues in the classroom. In D. Gabel (Ed.), *Handbook of research on science teaching and learning* (pp. 542-557). New York: Macmillan.
- Kalechstein, P., Hocesvar, D., Zimmer, J.W. & Kalechstein, M. (1989). Procrastination over test preparation and test anxiety. In R. Schwarzer, H.K. van der Ploeg & C.D. Spielberger (Eds.), *Advances in test anxiety research* (pp. 63-76). Lisse: Swets and Zeitlinger.
- Keeves, J.P. (1992). The design and conduct of the Second Science Study. In

- J.P. Keeves (Ed.), *The IEA study of science III* (pp. 141-164). Oxford, England: Pergamon.
- Keeves, J.P. & Kotte, D. (1992). Disparities between the sexes in science education: 1970-84. In J.P. Keeves (Ed.), *The IEA study of science III* (pp. 141-164). Oxford, England: Pergamon.
- Kelly, A. (1985). The construction of masculine science. *British Journal of Sociology of Education*, 6, 133-154.
- Kelly, A. (1988). The customer is always right ...: girls' and boys' reactions to science lessons. *The School Science Review*, 69, 662-676.
- Kendall, P.C. (1984). Behavioral assessment and methodology. In G.T. Wilson, C.M. Franks, K.D. Brownell & P.C. Kendall (Eds.), *Annual review of behavior therapy: Theory and practice* (pp. 39-94). New York: Guildford Press.
- Kim, S.H. & Rocklin, T.(1994). The temporal patterns of worry and emotionality and their differential effects on test performance. *Anxiety, Stress, and Coping*, 7, 117-130.
- Kirk, R.E. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement*, 55, 746-759.
- Koballa, T.R., Jr., Crawley, F.E. & Shrigley, R.L. (1988). *A summary of research in science education - 1988*. (ERIC document reproduction Service No. ED 321 969).
- Kremer, B.K. & Walberg, H.J. (1981). A synthesis of social and psychological influences on science learning. *Science Education*, 65, 11-23.
- Krockover, G.H. & Shepardson, D.P. (1995). The missing links in gender equity research. *Journal of Research in Science Teaching*, 32, 223-224.
- Langenfeld, T.E. & Pajares, F. (1993, April). *The mathematics self-efficacy scale: A validation study*. Paper presented at the annual meeting of the American Educational Research Association, Atlanta, GA.
- Leder, G. (1985). Sex-related differences in mathematics: An overview. *Educational Studies in Mathematics*, 16, 304-309.
- Lenney, E. (1977). Women's self-confidence in achievement settings. *Psychological Bulletin*, 84, 1-13.
- Licht, B.G. (1987, April). *The interaction between children's achievement-related beliefs and the characteristics of different tasks*. Paper presented at the annual

- meeting of the American Educational Research Association, Washington, D.C.
- Licht, B.G. & Dweck, C.S. (1983). Sex differences in achievement orientations: Consequences for academic choices and attainments. In M. Marland (Ed.), *Sex differentiation and schooling* (pp. 72-97). London: Heinemann.
- Licht, B.G., Stader, S.R. & Swenson, C.C. (1989). Children's achievement-related beliefs: Effects of academic area, sex, and achievement level. *Journal of Educational Research*, 82, 253-260.
- Liebert, R.M. & Morris, L.W. (1967). Cognitive and emotional components of test anxiety: A distinction and some initial data. *Psychological Reports*, 20, 975-978.
- Lincoln, Y.S. & Guba, E.G. (1985). *Naturalistic inquiry*. Newbury Park, CA: Sage.
- Linn, M.C. (1991). Gender differences in educational achievement. In *Sex equity in educational opportunity, achievement and testing: Proceedings of the 1991 ETS invitational conference* (pp. 11-50). Princeton, NJ: Educational Testing Service.
- Linn, M.C., De Benedictus, T., Delucchi, K. Harris, A. & Stage, E. (1987). Gender differences in national assessment of educational progress items: What does "I don't know" really mean? *Journal of Research in Science Teaching*, 24, 267-278.
- Linn, M.C. & Hyde, J. (1989). Gender, mathematics and science. *Educational Researcher*, 18(8), 17-19, 22-27.
- Linn, M.C. & Petersen, A.C. (1985a). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development*, 56, 1479-1498.
- Linn, M.C. & Petersen, A.C. (1985b). Facts and assumptions about the nature of sex differences. In S. Klein (Ed.), *Handbook for achieving sex equity through education* (pp. 53-57). Baltimore, MD: John Hopkins University Press.
- Linn, M.C. & Petersen, A.C. (1986). A meta-analysis of gender differences in spatial ability: Implications for mathematics and science achievement. In J.S. Hyde & M.C. Linn (Eds.), *The psychology of gender: Advances through meta-analysis* (pp. 67-101). Baltimore, MD: John Hopkins University Press.
- Lu, C-H & Suen, H.K. (1995). Assessment approaches and cognitive styles. *Journal of Educational Measurement*, 32, 1-17.

- Lukhele, R., Thissen, D. & Wainer, H. (1994). On the relative value of multiple-choice, constructed response, and examinee-selected items on two achievement tests. *Journal of Educational Measurement*, 31, 234-250.
- Lundeberg, M.A., Brown, A.C., Fox, P.W. & Elbedour, S. (1998, April). *Cultural influences on confidence: Country and gender*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.
- Lundeberg, M.A., Fox, P.W. & Puncochar, J. (1994). Highly confident but wrong: Gender differences and similarities in confidence judgments. *Journal of Educational Psychology*, 86, 114-121.
- Lyon, M.A. (1993). Academic self-concept and its relationship to achievement in a sample of junior high school students. *Educational and Psychological Measurement*, 53, 201-210.
- Maccoby, E.E. & Jacklin, C.N. (1974). *The psychology of sex-differences*. Stanford, CA: Stanford University Press.
- Madaus, G.F. (1988). The influence of testing on the curriculum. In L.N. Tanner (Ed.), *Critical issues in curriculum*, Eighty-seventh Yearbook of the National Society for the Study of Education (pp. 83-121). Chicago, IL: University of Chicago Press.
- Maddux, J.E. (1995a). Looking for common ground. In J.E. Maddux (Ed.), *Self-efficacy, adaptation, and adjustment* (pp. 377-385). New York: Plenum Press.
- Maddux, J.E. (1995b). Self-efficacy theory: An introduction. In J.E. Maddux (Ed.), *Self-efficacy, adaptation, and adjustment* (pp. 3-33). New York: Plenum Press.
- Maddux, J.E. & Lewis, J. (1995). Self-efficacy and adjustment: Basic principles and issues. In J.E. Maddux (Ed.), *Self-efficacy, adaptation, and adjustment* (pp. 37-68). New York: Plenum Press.
- Maki, R.H. & Berry, S.L. (1984). Metacomprehension of test material. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10, 663-679.
- Mandler, G. & Sarason, S.B. (1952). A study of anxiety and learning. *Journal of Abnormal and Social Psychology*, 47, 166-173.
- Many, M.A. & Many, W.A. (1975). The relationship between self-esteem and anxiety in grades four through eight. *Educational and Psychological Measurement*, 35, 1017-1021.
- Marsh, H.W. (1987). The hierarchical structure of self-concept and the application of

- hierarchical confirmatory factor analysis. *Journal of Educational Measurement*, 24, 17-39.
- Marsh, H.W., Parker, J. & Barnes, J. (1985). Multidimensional adolescent self-concepts: Their relationship to age, sex, and academic measures. *American Educational Research Journal*, 22, 422-444.
- Marsh, H.W. & Shavelson, R. (1985). Self-concept: Its multifaceted, hierarchical structure. *Educational Psychologist*, 20, 107-123.
- Marsh, H.W., Walker, R. & Debus, R. (1991). Subject-specific components of academic self-concept and self-efficacy. *Contemporary Educational Psychology*, 16, 331-345.
- Martinez, M.E. (1998, April). *Cognition and the question of test item format*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.
- Massey, A. (1983). The effects of handwriting and other incidental variables on GCE 'A' level marks in English literature. *Educational Review*, 35, 45-50.
- Mathison, S. (1988). Why triangulate? *Educational Researcher*, 17(2), 13-17.
- Mazzeo, J., Schmitt, A.P. & Bleistein, C.A. (1991, April). *Do women perform better, relative to men on constructed-response tests or multiple-choice tests? Evidence from the Advanced Placement examinations*. Paper presented at the annual meeting of the National Council of Measurement in Education, Chicago, IL.
- Mazzeo, J., Schmitt, A.P. & Bleistein, C.A. (1993). *Sex-related performance differences on constructed-response and multiple-choice sections of Advanced Placement examinations* (College Board Report No. 92-7). New York: College Entrance Examination Board.
- Mboya, M.M. (1989). The relative importance of global self-concept and self-concept of academic ability in predicting academic achievement. *Adolescence*, 24, 39-46.
- Mednick, M.T. & Thomas, V.G. (1993). Women and the psychology of achievement: A view from the eighties. In F.L. Denmark & M.A. Paludi (Eds.), *Psychology of women: A handbook of issues and theories* (pp. 585-626). Westport, CT: Greenwood Press.
- Mehrens, W. (1998, April). *Consequences of assessment: What is the evidence?*

- Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.
- Mehrens, W.A. & Lehmann, I.J. (1984). *Measurement and evaluation in education and psychology*. New York: CBS.
- Merriam, S. B. (1988). *Case study research in education*. San Francisco, CA: Jossey-Bass.
- Messick, S. (1984). The nature of cognitive styles: Problems and promise in educational practice. *Educational Psychologist*, 19, 59-74.
- Miles, M.B. & Huberman, A.M. (1984). Drawing valid meaning from qualitative data: Toward a shared craft. *Educational Researcher*, 13(5), 20-30.
- Miles, M.B. & Huberman, A.M. (1994). *Qualitative data analysis: An expanded source book* (2nd ed.). Thousand Oaks, CA: Sage.
- Miller, G.A. (1986). *Explanatory skills*. Paper commissioned by the Study Group on National Assessment of Student Achievement. (ERIC document reproduction Service No. ED 279 688).
- Miller, S.M., Nelson, M.W. & Moore, M.T. (1998). Caught in the paradigm gap: Qualitative researchers' lived experience and the politics of epistemology. *American Educational Research Journal*, 35, 377-416.
- Milliner, K. (1998, August 27). Literally losing. *The Courier Mail*, p.15.
- Mone, M.A., Baker, D.D. & Jeffries, F. (1995). Predictive validity and time-dependency of self-efficacy, self-esteem, personal goals, and academic performance. *Educational and Psychological Measurement*, 55, 716-727.
- Morris, L.W., Davis, M.A. & Hutchings, C.H. (1981). Cognitive and emotional components of anxiety: Literature review and a revised worry-emotionality scale. *Journal of Educational Psychology*, 73, 541-555.
- Morris, L.W. & Fulmer, R.S. (1976). Test anxiety (worry and emotionality) changes during academic testing as a function of feedback and test importance. *Journal of Educational Psychology*, 68, 817-824.
- Morris, L.W. & Liebert, R.M. (1969). Effects of anxiety on timed and untimed intelligence tests: Another look. *Journal of Consulting and Clinical Psychology*, 33, 240-244.
- Morris, L.W. & Liebert, R.M. (1970). Relationship of cognitive and emotional components of test anxiety to physiological arousal and academic performance.

- Journal of Consulting and Clinical Psychology*, 35, 332-337.
- Moss, P.A. (1996). Enlarging the dialogue in educational measurement: Voices from interpretive research traditions. *Educational Researcher*, 25(1), 20-28, 43.
- Multon, K.D., Brown, S.D. & Lent, R.W. (1991). Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation. *Journal of Counseling Psychology*, 38, 30-38.
- Murphy, P. (1988). Gender and assessment. *Curriculum*, 9, 165-171.
- Murphy, P. (1991). Assessment and gender. *Cambridge Journal of Education*, 21, 203-214.
- Murphy, P. (1994, April). *Sources of inequity: Understanding students' responses to assessment*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
- Murphy, P. (1996). Assessment practices and gender in science. In L.H. Parker, L.J. Rennie & B.J. Fraser (Eds.), *Gender, science and mathematics: Shortening the shadow* (pp. 105-117). Dordrecht: Kluwer.
- Murphy, R.J.L. (1980). Sex differences in GCE examination entry statistics and success rates. *Educational Studies*, 6, 169-178.
- Murphy, R.J.L. (1982). Sex differences in objective test performance. *British Journal of Educational Psychology*, 52, 213-219.
- Neil, D. & Medina, M. (1989). Standardised testing: Harmful to educational health. *Phi Delta Kappan*, 70, 688-697.
- Newton, D., Tymms, P & Carrick, N. (1996). Is the GCSE fair? Examination success and cognitive style. *British Journal of Curriculum and Assessment*, 6(2), 21-25.
- Okebukola, P.A. (1986). The influence of preferred learning styles on cooperative learning in science. *Science Education*, 70, 509-517.
- Osborn, M. (1994). Status and prospects of women in science in Europe. *Science*, 263, 1389-1391.
- Owen, S.V., & Froman, R.D. (1988, April). *Development of a college academic self-efficacy scale*. Paper presented at the National Council of Measurement in Education, New Orleans, LA.
- Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research*, 66, 543-578.

- Pajares, F. (1997). Current directions in self-efficacy research. In M. Maehr & P.R. Pintrich (Eds.), *Current directions in self-efficacy research* (pp. 1- 49). Greenwich, CT: JAI Press.
- Pajares, F. & Johnson, M.J. (1994). Confidence and competence in writing: The role of self-efficacy, outcome expectancy, and apprehension. *Research in the Teaching of English*, 28, 313-331.
- Pajares, F., & Miller, M.D. (1994). Role of self-efficacy and self-concept beliefs in mathematical problem solving: A path analysis. *Journal of Educational Psychology*, 86, 193-203.
- Parents want role in assessments. (1995, August 14). *The Canberra Times*, p.5.
- Parker, L.H. (1984). Science for all? The role of science teachers in eliminating gender differences in access to scientific knowledge. *Australian Science Teachers Journal*, 30(1), 21-28.
- Parker, L.H. & Rennie, L.J. (1998). Equitable assessment strategies. In B.J. Fraser & K.G. Tobin (Eds.), *International Handbook of Science Education* (pp. 897-910). Dordrecht: Kluwer.
- Parker, L.H. & Tims, J.E. (1995, April). *Different modes of assessment in science and mathematics: A systematic interaction with gender*. Paper presented at the annual meeting of the National Association for Research in Science Teaching, San Francisco, CA.
- Paulman, R.G. & Kennelly, K.J. (1984). Test anxiety and ineffective test-taking: Different names, same construct? *Journal of Educational Psychology*, 76, 279-288.
- Peshkin, A. (1993). The goodness of qualitative research. *Educational Researcher*, 22(2), 24-30.
- Poiner, G. (1995). Whose education? Questions of equity, gender, schools and society. *Current Affairs Bulletin*, 71(5), 23-31.
- Powell, J. & Gillespie, C. (1990, December). *All tests are not created equally*. Paper presented at the annual meeting of the American Reading Forum, Sarasota, FA.
- Pressley, M., Levin, J.R., Ghatala, E.S. & Ahmad, M. (1987). Test monitoring in young grade school children. *Journal of Experimental Child Psychology*, 43, 96-111.
- Prins, P.J.M., Groot, M.J.M. & Hanewald, G.J.F.P. (1994). Cognition in test-anxious

- children: The role of on-task and coping cognition reconsidered. *Journal of Consulting and Clinical Psychology*, 62, 404-409.
- Queensland Board for Senior Secondary School Studies (1997). *Statistics bulletin for year ended 1996*. Spring Hill, Qld: Queensland Board for Senior Secondary School Studies.
- Randhawa, B.S., Beamer, J.E., & Lundeberg, I. (1993). Role of mathematics self-efficacy in the structural model of mathematics achievement. *Journal of Educational Psychology*, 85, 41-48.
- Reichardt, C.S. & Rallis, S.F. (1994a). Qualitative and quantitative inquiries are not incompatible: A call for a new partnership. In C.S. Reichardt & S.F. Rallis (Eds.), *The qualitative-quantitative debate: New perspectives* (pp. 85-91). San Francisco, CA: Jossey-Bass.
- Reichardt, C.S. & Rallis, S.F. (1994b). The relationship between the qualitative and the quantitative research traditions. In C.S. Reichardt & S.F. Rallis (Eds.), *The qualitative-quantitative debate: New perspectives* (pp. 5-11). San Francisco, CA: Jossey-Bass.
- Rennie, L.J. & Parker, L.H. (1991). Assessment of learning in science: The need to look closely at item characteristics. *The Australian Science Teachers Journal*, 37(4), 56-59.
- Rennie, L.J. & Parker, L.H. (1998). Equitable measurement of achievement in physics: High school students' responses to assessment tasks in different formats and contexts. *Journal of Women and Minorities in Science and Engineering*, 4, 113-127.
- Rennie, L.J. & Punch, K.F. (1991). The relationship between affect and achievement in science. *Journal of Research in Science Teaching*, 28, 193-209.
- Renshaw, P. (1990). Self-esteem research and equity programs for girls: A reassessment. In J. Kenway & S. Willis (Eds.), *Hearts and minds: Self-esteem and the schooling of girls* (pp. 17-33). London: Falmer.
- Rogers, C.R. (1951). *Client-centered therapy*. Boston: Houghton Mifflin.
- Rosenthal, R & Rubin, D.B. (1982a). Further meta-analytic procedures for assessing cognitive gender difference. *Journal of Educational Psychology*, 74, 708-712.
- Rosenthal, R & Rubin, D.B. (1982b). A simple, general purpose display of magnitude of experimental effect. *Journal of Educational Psychology*, 74, 166-169.

- Rosser, P., Brown, S., Greenberger, M., Johnson, S., Medaus, G., Welsch, M., & Wolfe, L. (1989). *Gender bias in testing: Current debates for future priorities. A public policy dialogue*. Proceedings of the Ford Foundation Women's Program Forum. New York: Ford Foundation.
- Ryan, J.M. & Franz, S. (1998, April). *A gender-by-format interaction perspective on the use of free-response and selected-response assessment formats*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.
- Saracho, O.N. (1989). Cognitive style: Individual differences. *Early Childhood Development and Care*, 53, 75-81.
- Sarason, I.G. (1960). Empirical findings and theoretical problems in the use of anxiety scales. *Psychological Bulletin*, 57, 403-415.
- Sarason, I.G. (1980). Introduction to the study of test anxiety. In I.G. Sarason (Ed.), *Test anxiety: Theory, research and applications* (pp. 3-14). Hillsdale, NJ: Erlbaum
- Sarason, I.G. (1984). Stress, anxiety, and cognitive interference: Reactions to tests. *Journal of Personality and Social Psychology*, 46, 929-938.
- Sarason, I.G. (1991). Anxiety, self-preoccupation and attention. In R. Schwarzer & R.A. Wicklund (Eds.), *Anxiety and self-focused attention* (pp. 9-13). Chur, Switzerland: Harwood
- Sarason, I.G., Sarason, B.R., & Pierce, G.R. (1991). Anxiety, cognitive interference, and performance. In M. Booth-Butterfield (Ed.), *Communication, cognition and anxiety* (pp. 1-18). Newbury Park, CA: Sage.
- Sarason, S.B., Davidson, K., Lighthall, F. & Waite, R. (1958). A test anxiety scale for children. *Child Development*, 29, 105-113.
- Sax, G. (1989). *Principles of educational and psychological measurement and evaluation*. Belmont, CA: Wadsworth.
- Scheier, M.A. & Kraut, R.E. (1979). Increasing educational achievement via self-concept change. *Review of Educational research*, 49, 131-150.
- Scheuneman, J.D. (1991). Item bias and individual differences. In P.L. Dean, S.H. Irvine & J.M. Collis (Eds.), *Advances in computer-based human assessment*. Dordrecht: Kluwer.
- Schibeci, R.A. (1984). Attitudes to science: An update. *Studies in Science*

- Education*, 11, 26-59.
- Schrag, F. (1992). In defense of positivist research paradigms. *Educational Researcher*, 21(5), 5-11.
- Schunk, D.H. (1989). Self-efficacy and cognitive skill learning. In C. Ames & R. Ames (Eds.), *Research on motivation in education*, Vol. 3 (pp. 13-44). San Diego, CA: Academic Press.
- Schunk, D.H. (1990). Goal setting and self-efficacy during self-regulated learning. *Educational Psychologist*, 25, 71-86.
- Schunk, D.H. (1991). Self-efficacy and achievement motivation. *Educational Psychologist*, 26, 207-231.
- Schunk, D.H. (1995). Self-efficacy and education and instruction. In J.E. Maddux (Ed.), *Self-efficacy, adaptation, and adjustment* (pp. 281-303). New York: Plenum Press.
- Schwarzer, R., van der Ploeg, H.M. & Spielberger, C.D. (1982). Test anxiety: an overview of theory and research. In R. Schwarzer, H.K. van der Ploeg & C.D. Spielberger (Eds.), *Advances in test anxiety research* (pp. 3-9). Lisse: Swets and Zeitlinger.
- Seipp, B. (1991). Anxiety and academic performance: A meta-analysis of findings. *Anxiety Research*, 4, 27-41.
- Shaha, S.H. (1982). *Reduced anxiety and increased assessment effectiveness with matching test formats*. Los Angeles, CA: California University Center for the Study of Evaluation. (ERIC document reproduction service No. ED 228 281).
- Shavelson, R.J., Hubner, J.J. & Stanton, G.C. (1976). Self-concept: Validation of construct interpretations. *Review of Educational Research*, 46, 407-441.
- Sherer, M., Maddux, J.E., Mercandante, B., Prentice-Dunn, S., Jacobs, B., & Rogers, R.W. (1982). The self-efficacy scale: Construction and validation. *Psychological Reports*, 51, 663-671.
- Sherman, J. (1967). Problems of sex differences in space perception and aspects of intellectual functioning. *Psychological Review*, 74, 290-299.
- Sherman, S.W. (1974). *Multiple choice test bias uncovered by use of an 'I don't know' alternative*. (ERIC document reproduction Service No. ED 121 824).
- Shrigley, R.L., Koballa, T.R., Jr., & Simpson, R.D. (1988). Defining attitude for science educators. *Journal of Research in Science Teaching*, 25, 659-678.

- Sieber, J.E. (1980). Defining test anxiety: Problems and approaches. In I.G. Sarason (Ed.), *Test anxiety: Theory, research and applications* (pp. 15-42). Hillsdale, NJ: Erlbaum
- Sieber, J.E., O'Neil, H.F., Jr. & Tobias, S. (1977). *Anxiety, learning and instruction*. Hillsdale, NJ: Erlbaum.
- Siegel, R.G., Galassi, J.P., & Ware, W.B. (1985). A comparison of two models for predicting mathematics performance: Social learning versus math aptitude-anxiety. *Journal of Counseling Psychology*, 32, 531-538.
- Simpson, R.D., Koballa, T.R., Jr., Oliver, J.S. & Crawley, F.E. III (1994). Research on the affective dimension of science learning. In D. Gabel (Ed.), *Handbook on research on science teaching and learning* (pp. 211-234). New York: Macmillan.
- Simpson, R.D. & Oliver, J.S. (1990). A summary of major influences on attitude toward and achievement in science among adolescent students. *Science Education*, 74, 1-18.
- Simpson, R.D. & Troost, K.M. (1982). Influences to and commitment to and learning of science among adolescent students. *Science Education*, 66, 763-781.
- Sinclair, K.E. (1991). Students' affective characteristics and classroom behaviour. In K. Marjoribanks (Ed.), *The foundations of students' learning* (pp. 167-176). Oxford: Pergamon.
- Sincoff, J. & Sternberg, R. (1987). Two faces of verbal ability. *Intelligence*, 11, 263-276.
- Sincoff, J. & Sternberg, R. (1988). Development of verbal fluency abilities and strategies in elementary-school-age children. *Developmental Psychology*, 24, 646-653.
- Sjoberg, S. & Imsen, G. (1988). Gender and science education I. In P.J. Fensham (Ed.), *Developments and dilemmas in science education* (pp. 218-248). London: Falmer.
- Smail, B. & Kelly, A. (1984). Sex differences in science and technology among 11-year old school children: II-affective. *Research in Science and Technological Education*, 2, 87-106.
- Smist, J.M. (1993, August). *General chemistry and self-efficacy*. Paper presented at

- the national meeting of the American Chemical Society, Chicago, IL.
- Smith, J.K. (1997). The stories educational researchers tell about themselves. *Educational Researcher*, 26(5), 4-11.
- Smith, J.K. & Heshusius, L. (1986). Closing down the conversation: The end of the quantitative-qualitative debate among educational inquirers. *Educational Researcher*, 15(1), 4-12.
- Smith, L.F., Smith, J.K. & Lackland, A.G. (1998, April). *Be careful what you wish for: When students choose item formats*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.
- Smith, M.L. (1994). Qualitative plus/versus quantitative: The last word. In C.S. Reichardt & S.F. Rallis (Eds.), *The qualitative-quantitative debate: New perspectives* (pp 37-44). San Francisco, CA: Jossey-Bass.
- Smith, R.J., Arnkoff, D.B. & Wright, T.L. (1990). Test anxiety and academic competence: A comparison of alternative models. *Journal of Counseling Psychology*, 37, 313-321.
- Snow, R.E. (1993). Construct validity and constructed-response tests. In R.E. Bennett & W.C. Ward (Eds.), *Construction versus choice in cognitive measurement: Issues in constructed response, performance testing and portfolio assessment* (pp. 45-60). Hillsdale, NJ: Erlbaum.
- Sonnert, G. (1996). Gender equity in science: Still an elusive goal. *Issues in Science and Technology*, 12(2), 53-58.
- Spear, M. (1984). Sex bias in teachers' ratings of work and pupil characteristics. *European Journal of Science Education*, 6, 369-377.
- Spear, M.G. (1989). Differences between the written work of boys and girls. *British Educational Research Journal*, 15, 271-277.
- Spiegler, M.D., Morris, L.W. & Liebert, R.M. (1968). Cognitive and emotional components of test anxiety: Temporal factors. *Psychological Reports*, 22, 451-456.
- Spielberger, C.D. (1966). Theory and research on anxiety. In C.D. Spielberger (Ed.), *Anxiety and behavior* (pp. 3-20). New York: Academic Press.
- Spielberger, C.D. (1980). *Preliminary professional manual for the Test Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.
- Spielberger, C., Anton, W. & Bedall, J. (1976). The nature and treatment of test

- anxiety. In M. Zuckerman & C.D. Spielberger (Eds.), *Emotion and anxiety: New concepts, methods and applications* (pp. 137-345). Hillsdale, N.J.: Erlbaum.
- Stanley, K.D. & Murphy, M.R. (1997). A comparison of general self-efficacy with self-esteem. *Genetic, Social, and General Psychology Monographs*, 123(1), 81-99.
- Steinkamp, M.W. & Maehr, M.L. (1983). Affect, ability and science achievement: A quantitative synthesis of correlational research. *Review of Educational Research*, 53, 369-396.
- Stobart, G., Elwood, J. & Quinlan, M. (1992). Gender bias in examinations: How equal are the opportunities? *British Educational Research Journal*, 18, 261-276.
- Stumpf, H. & Stanley, J.C. (1996). Gender-related differences on the College Board's Advanced Placement and achievement tests, 1982-1992. *Journal of Educational Psychology*, 88, 353-364.
- Taylor, S.J. & Bogdan, R. (1984). *Introduction to qualitative research methods* (2nd ed.). New York: Wiley.
- Teese, R., Davies, M., Charlton, M. & Polesel, J. (1995). *Who wins at school?* Melbourne: Department of Education Policy and Management, University of Melbourne
- Thissen, D., Wainer, H. & Wang, X-B. (1994). Are tests comprising both multiple-choice and free-response items necessarily less unidimensional than multiple-choice tests? An analysis of two tests. *Journal of Educational Measurement*, 31, 113-123.
- Time for a new boy? (1998, September). *Education Review*, p. 1.
- Tobias, S. (1985). Test anxiety: Interference, defective skills, and cognitive capacity. *Educational Psychologist*, 20, 135-142.
- Traub, R.E. (1993). On the equivalence of the traits assessed by multiple-choice and constructed-response tests. In R.E. Bennett & W.C. Ward (Eds.), *Construction versus choice in cognitive measurement: Issues in constructed response, performance testing and portfolio assessment* (pp. 29-34). Hillsdale, NJ: Erlbaum.
- Tryon, G.S. (1980). The measurement and treatment of test anxiety. *Review of*

- Educational Research*, 50, 343-372.
- Vispoel, W.P., & Chen, P. (1990, April). *Measuring self-efficacy: The state of the art*. Paper presented at the annual meeting of the American Educational Research Association, Boston, MA.
- Volkoff, J. & Hocevar, D. (1995, April). *Multiple-choice vs. performance based testing and gender*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.
- Vollmer, F. (1986). Why do men have higher expectancy than women? *Sex Roles*, 14, 351-362.
- Walding, R., Fogliani, C., Over, R. & Bain, J.D. (1994). Gender differences in response to questions on the Australian National Chemistry Quiz. *Journal of Research in Science Teaching*, 31, 833-846.
- Wapner, S.J. & Demick, J. (1991). Some open research problems on field dependence-independence: Theory and methodology. In S.J. Wapner & J. Demick (Eds.), *Field dependence-independence: Cognitive styles across the lifespan* (pp. 401-429). Hillsdale, NJ: Erlbaum.
- Ware, W.B., Galassi, J.P. & Dew, K.M.H. (1990). The Test Anxiety Inventory: A confirmatory factor analysis. *Anxiety Research*, 3, 205-212.
- Weare, J. (1984). *Test anxiety in adults*. Research report. (ERIC document reproduction service No. ED 244 116).
- Weinburgh, M. (1995). Gender differences in student attitudes toward science: A meta-analysis of the literature from 1970 to 1991. *Journal of Research in Science Teaching*, 32, 387-398.
- Westerback, M.E. & Primavera, L.H. (1992, March). *A science educators' and a psychologists perspective on research about science anxiety*. Paper presented at the annual meeting of the National Association for Research in Science Teaching, Boston, MA.
- White, R. & Gunstone, R. (1992). *Probing understanding*. London: Falmer Press.
- Whitehouse, H. & Sullivan, M. (1990). *Girls and year 12 science examinations: A study of the participation and performance of girls in publicly examined science subjects in South Australia*. SSABSA Research Report No. 1/1990. Wayville, South Australia: Senior Secondary Assessment Board of South Australia.

- Whitley B.E., Jr., McHugh, M.C. & Frieze, I.H. (1986). Assessing the theoretical models for sex differences in causal attributions of success and failure. In J.S. Hyde & M.C. Linn (Eds.), *The psychology of gender: Advances through meta-analysis* (pp. 102-135). Baltimore, MD: John Hopkins University Press.
- Wiersma, W. (1969). *Research methods in education: An introduction*. Philadelphia, PA: Lippincott.
- Wigfield, A. & Eccles, J.S. (1989). Test anxiety in elementary and secondary school students. *Educational Psychologist*, 24, 159-183.
- Wigfield, A. & Karpathian, M. (1991). Who am I and what can I do? Childrens' self-concepts and motivation in achievement situations. *Educational Psychologist*, 26, 233-261.
- Wilder, G.Z. & Powell, K. (1989). *Sex differences in test performance: A survey of the literature* (College Board Report No. 88-1). New York: College Entrance Examination Board.
- Williams, J. (1991). Modelling test anxiety, self concept and high school students' academic achievement. *Journal of Research and Development in Education*, 25, 51-57.
- Williams, J.E. (1994). Gender differences in high school students' efficacy-expectation/performance discrepancies across four subject matter domains. *Psychology in the Schools*, 31, 232-237.
- Williams, S.L. (1995). Self-efficacy, anxiety and phobic disorders. In J.E. Maddux (Ed.), *Self-efficacy, adaptation, and adjustment* (pp. 69-107). New York: Plenum Press.
- Willingham, W.W. & Cole, N.S. (1997a). Fairness issues in test design and use. In W.W. Willingham & N.S. Cole (Eds.), *Gender and fair assessment* (pp. 227-345). Manwah, NJ: Erlbaum.
- Willingham, W.W. & Cole, N.S. (1997b). Introduction. In W.W. Willingham & N.S. Cole (Eds.), *Gender and fair assessment* (pp. 1-15). Manwah, NJ: Erlbaum.
- Willingham, W.W. & Cole, N.S. (1997c). Research on gender differences. In W.W. Willingham & N.S. Cole (Eds.), *Gender and fair assessment* (pp. 17-54). Manwah, NJ: Erlbaum.
- Willingham, W.W. & Cole, N.S. (1997d). Summary and implications. In

- W.W. Willingham & N.S. Cole (Eds.), *Gender and fair assessment* (pp. 347-365). Mahwah, NJ: Erlbaum.
- Willson, V.L. (1983). A meta-analysis of the relationship between science achievement and science attitude: Kindergarten through college. *Journal of Research in Science Teaching*, 20, 839-850
- Wine, J. (1971). Test anxiety and direction of attention. *Psychological Bulletin*, 76, 92-104.
- Wine, J.D. (1980). Cognitive-attentional theory of test anxiety. In I.G. Sarason (Ed.), *Test anxiety: Theory, research and applications* (pp. 349-385). Hillsdale, NJ: Erlbaum.
- Witkin, H.A., Moore, C.A., Goodenough, D.R. & Cox, P.W. (1977). Field-dependent and field-independent cognitive styles and their educational implications. *Review of Educational Research*, 47, 1-64.
- Wittmaier, B. (1972). Test anxiety and study habits. *Journal of Educational Research*, 65, 352-354.
- Wolf, D.P. (1993). Assessment as an episode of learning. In R.E. Bennett & W.C. Ward (Eds.), *Construction versus choice in cognitive measurement: Issues in constructed response, performance testing and portfolio assessment* (pp. 213-240). Hillsdale, NJ: Erlbaum.
- Wolf, L.F. & Smith, J.K. (1995). The consequence of consequence: Motivation, anxiety, and test performance. *Applied Measurement in Education*, 8, 227-242.
- Wood, R. (1978). Sex differences in answers to English language comprehension items. *Educational Studies*, 4, 157-165.
- Wynstra, S. & Cummings, C. (1993). High school science anxiety: Easing common classroom fears. *The Science Teacher*, 60(7), 19-21.
- Zatz, S. & Chassin, L. (1985). Cognitions of test-anxious children under naturalistic test-taking conditions. *Journal of Consulting and Clinical Psychology*, 53, 393-401.
- Zeidner, M. (1987). Essay versus multiple-choice type classroom exams: The students' perspective. *Journal of Educational Research*, 80, 352-358.
- Zeldin, A.L. & Pajares, F. (1997, March). *Against the odds: Self-efficacy beliefs of women in mathematical, scientific, and technological careers*. Paper presented at the annual meeting of the American Educational Research Association,

Chicago, IL.

Zoller, U. & Ben-Chaim, D. (1988). Interaction of examination type, anxiety state and academic achievement in college science: an action-oriented research.

Journal of Research in Science Teaching, 26, 65-77.

APPENDIX A

DEVELOPMENT OF THE SELF-EFFICACY SCALE

PILOT INSTRUCTIONS FOR COMPLETING THE SELF-EFFICACY SCALE

Instructions

1. You will be asked in this test how confident you are that you can answer each question correctly.
2. For each question:
 - Firstly, read the question. **Do not start to work out the answer.**
 - Rate your **confidence** that you can answer it **correctly**.
 - Circle the appropriate response on the scale.
 - Solve the question.
3. **Do not** go back and change your confidence rating after you have started solving the question.
4. There are no right or wrong answers. Your answers on this scale will not influence your mark.

Examples

Bill has very little confidence that he can solve a particular question.

He circles the letter 'A'.

A	B	C	D	E
very	----->			Quite
little	CONFIDENCE			a lot

Jenni is not sure that she can solve a particular question.

She circles the letter 'C'.

A	B	C	D	E
very	----->			Quite
little	CONFIDENCE			a lot

Section A: Content section.

True/False Questions. (1/2 mark each)
Write T or F in the space next to the statement.

1. A precipitate is a soluble salt. _____
2. Melting ice is an example of a chemical change. _____
3. All chemical reactions produce heat. _____
4. A solution with a pH of 7 is neutral. _____
5. Marble is the common name for calcium carbonate. _____
6. Acidic substances accept hydrogen ions _____

Multiple Choice Section: Select the Most Correct Answer.

Question 1

Which is not a property of acids?

- a. They have a bitter taste.
- b. They are corrosive.
- c. They dissolve in water.
- d. They react with metals to produce hydrogen.

Rate your confidence that you can answer this question correctly

A B C D E
very <-----> Quite
little CONFIDENCE a lot

Question 2

A piece of blue litmus paper was placed in a solution. It remained blue. It can be concluded that the solution was:

- a. acidic
- b. neutral
- c. alkaline
- d. neutral or acidic.

Rate your confidence that you can answer this question correctly

A B C D E
very <-----> Quite
little CONFIDENCE a lot

Question 3

When red cabbage is boiled with water, the water becomes strongly coloured. If sodium bicarbonate, a soluble base, is added to the juice it turns a purple colour. When an acid, like hydrochloric acid, is added the solution becomes red. Which is the most likely explanation for these observations?

- a. With sodium bicarbonate the other colours are boiled out of the cabbage and therefore changing the colour.
- b. The cabbage turns red when it is stored.
- c. The acid stains the cabbage a red colour.
- d. The colour of the cabbage is an acid-base indicator.

Rate your confidence that you can answer this question correctly

A B C D E
very <-----> Quite
little CONFIDENCE a lot

Question 4

Which group of substances would not have acidic properties when they are dissolved in water?

- a. ammonia, magnesium oxide, sodium chloride.
- b. lemon juice, sulphur dioxide, sodium chloride.
- c. sulphur dioxide, lemon juice, magnesium oxide.
- d. ammonia, sulphur dioxide, sodium chloride.

Rate your confidence that you can answer this question correctly

A B C D E
very <-----> Quite
little CONFIDENCE a lot

Question 5

A chemist is attempting to develop a new hair shampoo. The first attempt to make the shampoo results in a product with a pH of 9. This product is...

- a. acidic and it would turn universal indicator red.
- b. acidic and it would turn universal indicator green.
- c. alkaline and it would turn universal indicator red.
- d. alkaline and it would turn universal indicator blue-purple.

Rate your confidence that you can answer this question correctly

A B C D E
very <-----> Quite
little CONFIDENCE a lot

Question 6.

To test for carbon dioxide, which of the following would be of no use at all?

- a. Bubbling CO_2 through distilled water and then testing the solution using pH paper.
 b. Pouring CO_2 over a lighted flame and observing the results.
 c. Bubbling CO_2 through lime water.
 d. Bubbling CO_2 through carbonic acid.

Rate your confidence that you can answer this question correctly

A B C D E

very <-----> Quite

little CONFIDENCE a lot

Question 7.

An alkali can be produced by:

- a. oxidising a non-metal and dissolving the product formed in water.
 b. oxidising a metal and dissolving the product formed in water.
 c. adding a metal to water.
 d. adding a non-metal to water.

Rate your confidence that you can answer this question correctly

A B C D E

very <-----> Quite

little CONFIDENCE a lot

Question 8.

When white phosphorous, a very reactive non-metal, is burned in air the product is:

- a. a basic oxide
 b. an acidic oxide
 c. a neutral oxide
 d. none of the above.

Rate your confidence that you can answer this question correctly

A B C D E

very <-----> Quite

little CONFIDENCE a lot

Question 9.

When a metal oxide reacts with an acid one of the products formed is:

- a. hydrogen
 b. carbon dioxide
 c. water
 d. none of the above.

Rate your confidence that you can answer this question correctly

A B C D E

very <-----> Quite

little CONFIDENCE a lot

Question 10.

The common constituent (part) of acids is:

- a. Oxygen
 b. Hydrogen
 c. Metal
 d. Zinc carbonate

Rate your confidence that you can answer this question correctly

A B C D E

very <-----> Quite

little CONFIDENCE a lot

Question 11.

I am given the following word equation for a chemical reaction with the substance "X".



I can say that X is most likely to be...

- a. an alkali.
 b. an acid.
 c. a metal.
 d. a carbonate.

Rate your confidence that you can answer this question correctly

A B C D E

very <-----> Quite

little CONFIDENCE a lot

Question 12.

When hydrochloric acid is neutralised by sodium hydroxide in the presence of universal indicator...

- a. a red solution is formed.
 b. a blue-purple solution is formed.
 c. no particular colour change is noted.
 d. a green-yellow colour is formed.

Rate your confidence that you can answer this question correctly

A B C D E

very <-----> Quite

little CONFIDENCE a lot

Question 13.

A candle of known mass is burned completely in air. The gases which are given off are collected. The mass of the gas would be...

- a. more than the starting mass of the candle.
 b. the mass of the candle at the start less the mass of the candle at the finish.
 c. the same as the candle.
 d. the same as the candle less the mass of the wick.

Rate your confidence that you can answer this question correctly

A B C D E

very <-----> Quite

little CONFIDENCE a lot

Question 14.

Chemicals X and Y are both soluble in water and are made up of charged particles called ions. The ions in chemical X and Y can be represented as shown.



When the two chemicals are mixed together, a dense precipitate forms. Which of the following combinations could represent the precipitate?

- (a) (b) (c) (d)

Rate your confidence that you can answer this question correctly

A B C D E

very <-----> Quite

little CONFIDENCE a lot

Question 15.

Which of the following will react with sulphuric acid to form carbon dioxide?

- a. calcium sulphate
b. copper oxide
c. magnesium carbonate
d. calcium chloride

Rate your confidence that you can answer this question correctly

A B C D E

very <-----> Quite

little CONFIDENCE a lot

Question 16.

The pH of an acid is:

- a. above 14
b. exactly 7
c. between 1 and 6
d. between 6 & 9

Rate your confidence that you can answer this question correctly

A B C D E

very <-----> Quite

little CONFIDENCE a lot

Question 17.

A gas is given off when a metal reacts with an acid. You would expect this gas to...

- a. turn limewater a milky white colour.
b. extinguish a burning splint.
c. have a strong smell.
d. produce a popping sound with a burning splint.

Rate your confidence that you can answer this question correctly

A B C D E

very <-----> Quite

little CONFIDENCE a lot

Question 18.

Explain a laboratory procedure you could follow to produce a salt of copper nitrate if you were provided with some nitric acid and copper oxide. (4 marks).

.....

.....

.....

.....

Rate your confidence that you can answer this question correctly

A B C D E

very <-----> Quite

little CONFIDENCE a lot

Question 19.

Write out observations you would expect to make if:

- a. sodium hydroxide and universal indicator are mixed

- b. sulphuric acid and magnesium react (1 mark)

- c. a piece of zinc is placed into a test tube containing hydrochloric acid and universal indicator (1 mark)

..... (1 mark)

Rate your confidence that you can answer this question correctly

A B C D E

very <-----> Quite

little CONFIDENCE a lot

d. the solution from Part b in this question is evaporated

(1 mark)

e. calcium carbonate is added to nitric acid

(1 mark)

f. a burning splint is placed into the test tube in Part b of this question

(1 mark)

Section B: Process Section

1. Write word equations for the reactions between:

a. zinc + nitric acid

.....

b. copper oxide + sulphuric acid

.....

c. calcium hydroxide + nitric acid

.....

d. magnesium + sulphuric acid

.....

e. magnesium carbonate + sulphuric acid

.....

(5 marks)

Rate your confidence that you can answer this question correctly

A B C D E

very <-----> Quite

little CONFIDENCE a lot

2. A chemist performs the following experiment:

She burns a piece of magnesium in oxygen. The resulting white powder is placed in a test tube and hydrochloric acid is added to it.

a. Write a word equation for the burning of magnesium.

(1 mark)

b. Write a word equation for the reaction between the product of the burning and the hydrochloric acid.

(1 mark)

Rate your confidence that you can answer this question correctly

A B C D E

very <-----> Quite

little CONFIDENCE a lot

3. Describe how you could prepare:

a. Copper sulphate from copper carbonate

.....

b. Magnesium sulphate from magnesium oxide

.....

(2 marks)

Rate your confidence that you can answer this question correctly

A B C D E
 very <-----> Quite
 little CONFIDENCE a lot

4. You are given the alkalis, sodium hydroxide and calcium hydroxide, and two acids, nitric acid and sulphuric acid.

a. How many salts could you prepare, by neutralisation, using those chemicals?

_____ (1 mark)

b. List the names of two of the salts you could form?

 _____ (1 mark)

Rate your confidence that you can answer this question correctly

A B C D E
 very <-----> Quite
 little CONFIDENCE a lot

5. Complete the following word equations for each of the following reactions and use the solubility rules on the front of the test paper to indicate the presence of any precipitates?

a. Sodium chloride + Silver nitrate →

_____ (2 marks)

b. Zinc sulphate + Barium chloride →

_____ (2 marks)

Rate your confidence that you can answer this question correctly

A B C D E
 very <-----> Quite
 little CONFIDENCE a lot

6. Use your knowledge of the solubility rules to state whether the following compounds are soluble in water. (2 marks)

a. potassium nitrate _____

b. zinc hydroxide _____

c. magnesium sulphate _____

d. mercury chloride _____

Rate your confidence that you can answer this question correctly

A B C D E
 very <-----> Quite
 little CONFIDENCE a lot

7. Suppose you were given three bottles labelled A, B and C, and you were told that one contained hydrochloric acid, another sodium hydroxide and the other water. Explain a procedure you could use to identify the three liquids if you were given some red litmus paper and a small piece of zinc. (3 marks)

.....

Rate your confidence that you can answer this question correctly

A B C D E
 very <-----> Quite
 little CONFIDENCE a lot

REVIEW QUESTIONS ABOUT THE SELF-EFFICACY SCALE

1. Did you understand how to use the confidence scale?
2. Were the instructions how to use the scale clear?
3. How could the procedure be improved?
4. How often did you answer the confidence scale **before** you solved the question?
 - a) Always.
 - b) Mostly.
 - c) Sometimes.
 - d) Never.

Comment on your answer.

5. What did you mean about your confidence when you circled the letter 'D'?

REVISED INSTRUCTIONS FOR ANSWERING THE SELF-EFFICACY
SCALE

Instructions For Rating Your Confidence

1. You will be asked in this test how confident you are that you can answer each question correctly.
2. For each question:
 - Firstly, read the question. **Do not start to work out the answer.**
 - Rate your **confidence** that you can answer the question **correctly.**
 - Circle the appropriate response.
 - Solve the question.
3. **Do not** rate your confidence in your answer.
Do not go back and change your confidence rating after you have started solving the question.
4. There are no right or wrong answers. Your answers on this scale will not influence your mark.
5. You will have ample time to complete the test.

Examples

John is not sure that he can solve a particular question.

He circles number 3 **before** he starts to solve it.

1	2	(3)	4	5
very	----->			Quite
little	CONFIDENCE			a lot

Jenni is fairly confident that she can solve a particular question.

She circles number 4 **before** she starts to solve it.

1	2	3	(4)	5
very	----->			Quite
little	CONFIDENCE			a lot

APPENDIX B
DEVELOPMENT OF THE *ATTITUDES TOWARD*
TESTS SURVEY

WORRY ITEMS FROM THE *TAI* (SPIELBERGER, 1980)

1. Thinking about my mark in a course interferes with my work on tests (Item 3)
2. I freeze up on important exams (Item 4)
3. During exams I find myself thinking about whether I'll ever get through school (Item 5)
4. The harder I work at taking a test, the more confused I get (Item 6)
5. Thoughts of doing poorly interfere with my concentration on tests (Item 7)
6. I seem to defeat myself while working on important tests (Item 14)
7. During tests I find myself thinking about the consequences of failing (Item 17)
8. During examinations I get so nervous that I forget facts I really know (Item 20)

ATTITUDES TOWARDS TESTS SURVEY (PILOT VERSION)

Name _____ Class _____

Gender (Please circle): Male Female

A number of statements which people have used to describe themselves are given below. Read each of the statements and then circle the appropriate number to the right of the statement to indicate how you feel about the test questions **you have just completed**.

1 = Almost Never, 2 = Sometimes, 3 = Often, 4 = Almost Always

There are no right or wrong answers. Do not spend too much time on one statement but give the answer which seems to describe how you feel. Please answer every statement.

- | | | | | | |
|----|--|---|---|---|---|
| 1 | Thinking about my marks in a course interferes with my work on tests | 1 | 2 | 3 | 4 |
| 2 | I do as well, or better than others on tests | 1 | 2 | 3 | 4 |
| 3 | I freeze up on important exams | 1 | 2 | 3 | 4 |
| 4 | I don't have to rush through questions | 1 | 2 | 3 | 4 |
| 5 | During exams I find myself thinking about whether I'll ever get through school | 1 | 2 | 3 | 4 |
| 6 | I am in control of my thoughts during tests | 1 | 2 | 3 | 4 |
| 7 | The harder I work at taking a test, the more confused I get | 1 | 2 | 3 | 4 |
| 8 | I am confident in my abilities when I work on tests | 1 | 2 | 3 | 4 |
| 9 | Thoughts of doing poorly interfere with my concentration on tests | 1 | 2 | 3 | 4 |
| 10 | I find it easy to understand what the questions ask | 1 | 2 | 3 | 4 |
| 11 | I seem to defeat myself while working on important tests | 1 | 2 | 3 | 4 |
| 12 | I am able to keep a clear mind during tests | 1 | 2 | 3 | 4 |
| 13 | During tests I find myself thinking about the consequences of failing | 1 | 2 | 3 | 4 |
| 14 | I do all right on tests | 1 | 2 | 3 | 4 |
| 15 | During examinations I get so nervous that I forget facts I really know | 1 | 2 | 3 | 4 |

RESULTS FROM THE PILOT STUDY OF THE *ATTITUDE*
TOWARDS SCIENCE TESTS

Table B.1
Item Functioning Statistics (All Students)

Item	<u>M</u>	<u>SD</u>	<u>r</u>
1	2.02	0.86	0.36
3	2.04	0.93	0.48
5	1.59	0.90	0.49
7	2.11	0.97	0.43
9	2.08	0.93	0.60
11	1.90	0.72	0.49
13	1.83	1.00	0.52
15	1.96	0.86	0.56

Note. 191 cases were analysed. r = correlation of item with rest of scale.

Table B.2
Item Functioning Statistics for Boys and Girls

Item	<u>M</u>	<u>SD</u>	<u>r</u>
Boys ^a			
1	2.01	0.91	0.41
3	1.91	0.81	0.39
5	1.62	0.90	0.50
7	2.15	0.98	0.38
9	2.03	0.86	0.56
11	1.86	0.74	0.33
13	1.92	1.05	0.56
15	1.89	0.88	0.51
Girls ^b			
1	2.03	0.82	0.32
3	2.16	1.02	0.56
5	1.56	0.90	0.49
7	2.07	0.97	0.48
9	2.12	0.99	0.63
11	1.94	0.71	0.63
13	1.74	0.95	0.50
15	2.02	0.83	0.60

Note. r = correlation of item with rest of scale. ^aScores from 91 boys were analysed. ^bScores from 100 girls were analysed.

Table B.3
Principal Components Analysis of Worry Items (All Students)

Item Number	Communality	Eigenvalue	Percent of Variance	Cumulative Percent of variance
1	1.00	3.24	40.4	40.4
3	1.00	0.94	11.7	52.2
5	1.00	0.92	11.5	63.6
7	1.00	0.80	10.0	73.6
9	1.00	0.63	7.9	81.6
11	1.00	0.56	7.0	88.6
13	1.00	0.49	6.1	94.8
15	1.00	0.42	5.2	100.0

Note. 191 cases were analysed.

Table B.4
Loading of Worry Items (All Students)

Item Number	Loading on Factor 1
1	0.50
3	0.64
5	0.64
7	0.57
9	0.73
11	0.63
13	0.66
15	0.70

Note. 191 cases were analysed.

Table B.5
Principal Components Analysis of Worry Items for Girls and Boys

Item Number	Communality	Eigenvalue	Percent of Variance	Cumulative Percent of variance
Girls^a				
1	1.00	3.50	43.7	43.7
3	1.00	0.97	12.1	55.9
5	1.00	0.85	10.6	66.5
7	1.00	0.68	8.5	75.1
9	1.00	0.62	7.7	82.7
11	1.00	0.54	6.8	89.5
13	1.00	0.42	5.3	94.8
15	1.00	0.41	5.2	100
Boys^b				
1	1.00	3.01	37.6	37.6
3	1.00	1.13	14.2	51.8
5	1.00	0.99	11.8	63.6
7	1.00	0.85	10.6	74.2
9	1.00	0.71	8.9	83.1
11	1.00	0.54	6.7	89.8
13	1.00	0.49	6.1	95.9
15	1.00	0.33	4.1	100.0

Note. ^aScores from 91 boys were analysed. ^bScores from 100 girls were analysed.

Table B.6
Loading of Worry Items for Girls

Item Number	Loading on Factor 1
1	0.44
3	0.69
5	0.63
7	0.62
9	0.75
11	0.74
13	0.63
15	0.72

Note. 100 cases were analysed

Table B.7
Loadings of Worry Items for Boys After Rotation
of the Axes

Item Number	Loading on Factor 1	Loading on Factor 2
1	0.38	0.42
3	-0.04	0.90
5	0.78	0.10
7	0.67	0.03
9	0.65	0.33
11	0.26	0.41
13	0.68	0.32
15	0.23	0.75

Note. 91 cases were analysed.

ATTITUDES TOWARDS TESTS SURVEY (REVISED VERSION)

Name _____ Class _____

Gender (Please circle): Male Female

A number of statements which people have used to describe themselves are given below. Read each of the statements and circle the number which describes how you feel about the test questions **you have just completed.**

1 = Almost Never, 2 = Sometimes, 3 = Often, 4 = Almost Always

There are no right or wrong answers. Do not spend too much time on one statement but give the answer which seems to describe how you feel. Please answer every statement.

- | | |
|---|---------|
| 1 Thinking about my marks in a subject interferes with my work on tests | 1 2 3 4 |
| 2 I do as well, or better than others on tests | 1 2 3 4 |
| 3 My mind freezes up on important exams | 1 2 3 4 |
| 4 I don't have to rush through questions | 1 2 3 4 |
| 5 During exams I find myself thinking about whether I'll ever
get through school | 1 2 3 4 |
| 6 I am in control of my thoughts during tests | 1 2 3 4 |
| 7 The harder I work at taking a test, the more confused I get | 1 2 3 4 |
| 8 I am confident in my abilities when I work on tests | 1 2 3 4 |
| 9 Thoughts of doing poorly interfere with my concentration on tests | 1 2 3 4 |
| 10 I find it easy to understand what the questions ask | 1 2 3 4 |
| 11 I seem to defeat myself while working on important tests | 1 2 3 4 |
| 12 I am able to keep a clear mind during tests | 1 2 3 4 |
| 13 During tests I find myself thinking about the consequences of failing | 1 2 3 4 |
| 14 I do all right on tests | 1 2 3 4 |
| 15 During examinations I get so nervous that I forget facts I really know | 1 2 3 4 |

APPENDIX C
TEST BOOKLETS AND MARKING SCHEDULES FOR
THE *EARTH SCIENCE* TESTS

EARTH SCIENCE TEST BOOKLET 1A

- N.B. i) Test booklet 1B is identical to test booklet 1A except that the order of the multiple-choice and the constructed-response sections are reversed.
- ii) This test booklet has been photoreduced in size.

Time: 45 minutes

Total marks: 40 (20 Content & 20 Process)

Instructions:

1. This paper consists of two sections.
2. Answer **all** questions in the order they are presented.
3. Please rate how confident you are that you can answer each question correctly **before** you start answering the question. Instructions for answering the confidence scale are provided on a separate sheet of paper.
4. Please do not move onto section 2 until you have completed the *Attitudes Toward Tests Survey*.
5. You have been provided with adequate time to complete the test and the survey questions.

Multiple Choice Questions (20 marks)

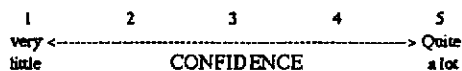
All questions are worth one mark

Circle the best answer

Question 1 (C)

The Earth's tectonic plates are supported by ...

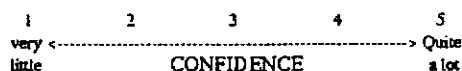
- A. convection currents.
- B. the mantle.
- C. the crust.
- D. the moho.



Question 2 (C)

The part of the Earth which is a solid that flows very slowly is the ...

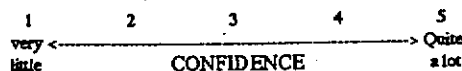
- A. crust.
- B. mantle.
- C. outer core.
- D. inner core.



Question 3 (C)

Basalt is usually a ...

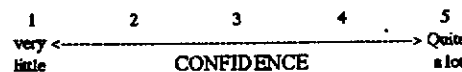
- A. dark coloured rock with large crystals.
- B. dark coloured rock with small crystals.
- C. light coloured rock with small crystals.
- D. light coloured rock with large crystals.



Question 4 (C)

Granite is usually a ...

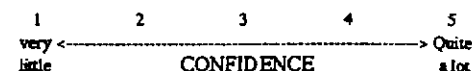
- A. dark coloured rock with large crystals.
- B. dark coloured rock with small crystals.
- C. light coloured rock with small crystals.
- D. light coloured rock with large crystals.



Question 5 (C)

Igneous intrusive rocks form from magma that cools ...

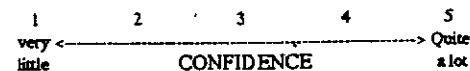
- A. rapidly at the Earth's surface.
- B. rapidly below the Earth's surface.
- C. slowly below the Earth's surface.
- D. slowly at the Earth's surface.



Question 6 (C)

Metamorphic rocks usually form under conditions of ...

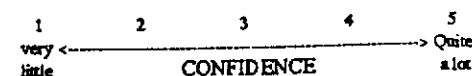
- A. increased pressure.
- B. increased temperature.
- C. increased pressure or increased temperature.
- D. increased temperature and increased pressure.



Question 7 (C)

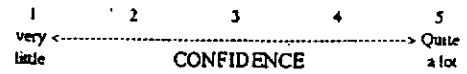
Metamorphic rocks often have ...

- A. fossils, a layered appearance and crumbly edges.
- B. intergrown crystals and crumbly edges.
- C. hard edges, crystals and a layered appearance.
- D. rounded grains, very hard edges and fossils.



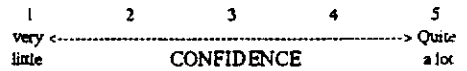
Question 8 (C)

- Sedimentary rocks often have ...
- A. fossils, a layered appearance and crumbly edges.
 - B. intergrown crystals and crumbly edges.
 - C. hard edges, crystals and a layered appearance.
 - D. rounded grains, very hard edges and fossils.



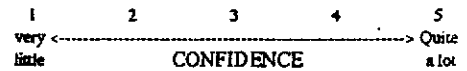
Question 9 (C)

- To say that the relative humidity is 70 per cent means that the air ...
- A. has the same amount of water vapour that it would have at 70°C.
 - B. is composed of 70 per cent water vapour.
 - C. has 70 parts of water vapour to 100 parts of dry air.
 - D. holds 70 per cent of its water vapour capacity.



Question 10 (C)

- When an air mass rises its temperature ...
- A. decreases and its relative humidity decreases.
 - B. decreases and its relative humidity increases.
 - C. increases and its relative humidity increases.
 - D. increases and its relative humidity decreases.



Questions 11 & 12 refer to the following information:

Imagine that a new plate boundary, shown on the map as - - - - -, divides the Australian plate in two. Arrows show the directions in which the new plates move.

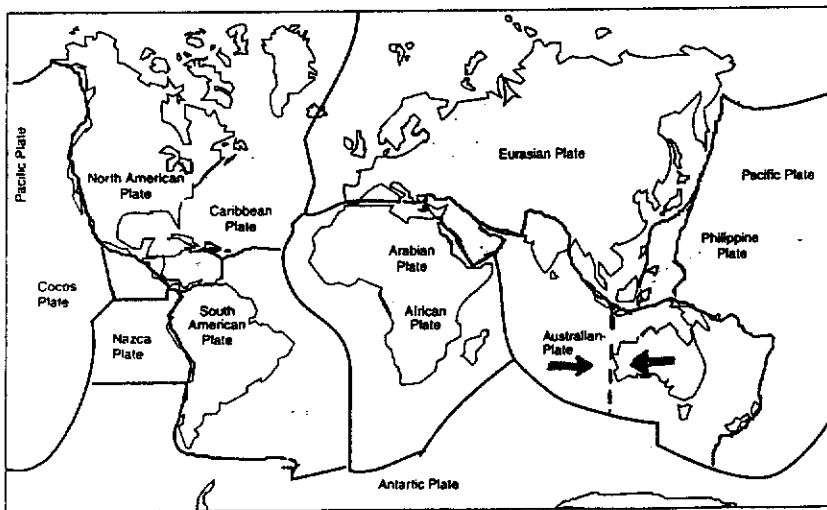
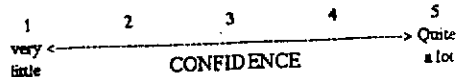


Fig. 1: The Earth's tectonic plates

Question 11 (P)

- The most likely change at the new plate boundary would be the ...
- A. formation of deep sea ridges off the Australian coast.
 - B. formation of a geosyncline on the Australian continent.
 - C. slow consumption of the Australian coastline.
 - D. formation of a deep trench off the Australian coast.



Question 12 (P)

The most likely change at the edge of the new plates would be caused by the ...

- A. build up of stresses in the crust.
 B. differing densities of the plates.
 C. production of magma.
 D. differing speeds of the two plates.

1 2 3 4 5
 very <-----> Quite
 little CONFIDENCE a lot

Questions 13 & 14 refer to the following table:

Mineral	Colour	Hardness	Lustre	Streak
Apatite	W, Gr, B, Y, P	4.5-5	R to V	W
Asbestos	W, Gr	5-6	P to V	W
Augite	Bl, G, Br, Gr	6-6.5	R to V	G-W
Biotite	Bl, Br, Gr	2.7-3.1	V	C
Calcite	C, W, B	3	D to V	W
Cassiterite	Br, Bl, R, G, W, Y	6-7	B	W, Br
Chalcopyrite	Y	3.5-4	M	Gr, Bl
Chlorite	G	2-2.5	P	W, Gr
Feldspar	W, Pink	6	P to V	C
Fluorite	W, Y, Gr, R, P, B	4	V	W
Galena	G	2.5	M	G
Garnet	R, Br, Y, W, Gr	6.5-7.5	R to V	W
Gypsum	W, C, G, Y, R	1.5-2	P to V	W
Haematite	G	5-6	M	R, Br
Halite	C, R, B, P, W	2.5	V	W
Hornblende	Bl, Gr, Bl	5-6	V	Br, Bl
Jasper	W, G, Br, R, B	7	D to V	W
Limonite	Br, Y, Bl	4.5-5	D to V	Br, Y, W
Magnetite	Bl	5.5-6.5	M	Bl
Muscovite	Gr, Br, Y, C	2.7-3.0	V	C
Olivine	Gr, Br	6.5-7	V	C
Pyrite	Y	6-6.5	M	Gr
Quartz	C, P, R, Y	7	V	W
Sphalerite	Y, Br, R, Bl, W	3.5-4	R to B	W
Talc	W, Gr	1	P	W

Abbreviation Code: Colour and streak: B = blue, Bl = black, Br = brown,
 C = colourless, G = grey, Gr = green,
 P = purple, R = red, W = white, Y = yellow
 Lustre: R = resinous, V = vitreous, D = dull,
 M = metallic, P = pearly, B = brilliant

Table 1: Properties of selected minerals

Question 13 (P)

Quartz and olivine can be distinguished from each other by their ...

- A. lustre and their colour.
 B. streak and their hardness.
 C. hardness and their lustre.
 D. colour and their streak.

1 2 3 4 5
 very <-----> Quite
 little CONFIDENCE a lot

Question 14 (P)

The properties of quartz and olivine which are the least useful for distinguishing between the two are their ...

- A. lustre and their colour.
 B. streak and their hardness.
 C. hardness and their lustre.
 D. colour and their streak.

1 2 3 4 5
 very <-----> Quite
 little CONFIDENCE a lot

Question 15 (P)

The following rock structure was most likely formed by ...

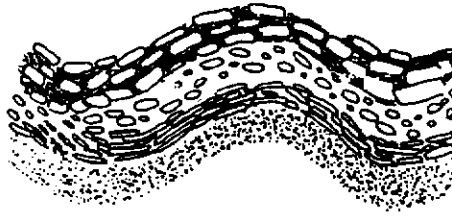
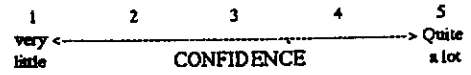


Fig. 2

- A. the action of horizontal forces.
- B. a series of lava flows.
- C. the movement of magma into a crack.
- D. the action of vertical forces.



Question 16 (P)

Structure "X" was most likely formed by ...

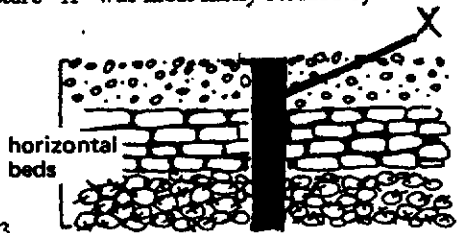
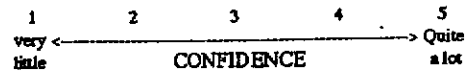


Fig. 3

- A. the action of horizontal forces.
- B. a series of lava flows.
- C. the movement of magma into a crack.
- D. the action of vertical forces.



Questions 17 - 20 refer to the following weather map.

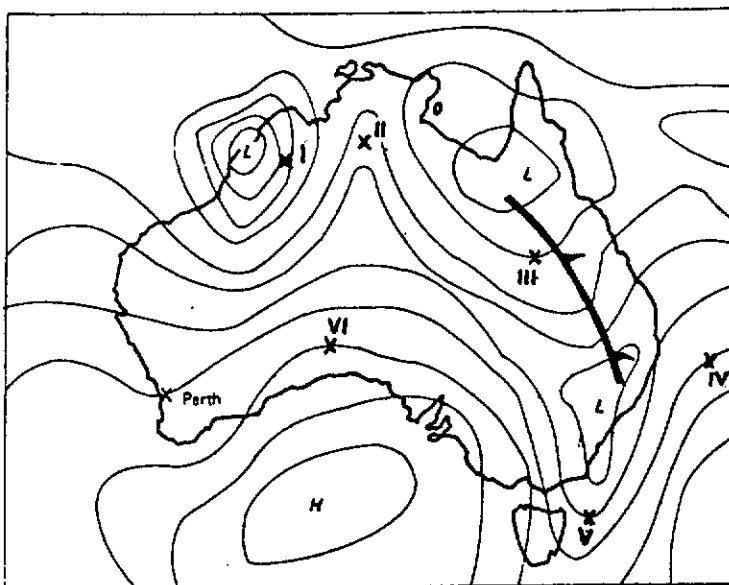
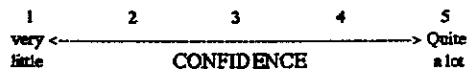


Fig. 4: Australian weather map, 6th February 1996.

Question 17 (P)

The atmospheric pressure at Perth is the same as that at ...

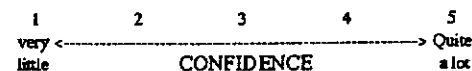
- A. III.
- B. IV.
- C. V.
- D. VI.



Question 18 (P)

The winds would be the strongest at location ...

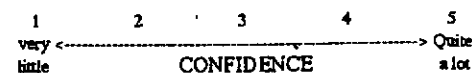
- A. I.
- B. II.
- C. III.
- D. IV.



Question 19 (P)

The wind direction at VI would be from the ...

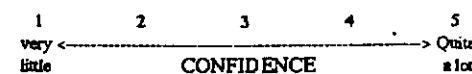
- A. north.
- B. east.
- C. west.
- D. south.



Question 20 (P)

The weather in the vicinity of (I) would probably be ...

- A. warm, windy and wet.
- B. cool, calm and dry.
- C. hot, calm and dry.
- D. cool, windy and dry.



ATTITUDES TOWARD TESTS SURVEY

A number of statements which people have used to describe themselves are given below. Read each of the statements and circle the number which describes how you feel about the test questions you have just completed.

1 = Almost Never, 2 = Sometimes, 3 = Often, 4 = Almost Always

There are no right or wrong answers. Do not spend too much time on one statement but give the answer which seems to describe how you feel. Please answer every statement.

1. Thinking about my marks in a subject interferes with my work on tests... 1 2 3 4
2. I do as well, or better than others on tests 1 2 3 4
3. My mind freezes up on important exams 1 2 3 4
4. I don't have to rush through questions 1 2 3 4
5. During exams I find myself thinking about whether I'll ever
get through school 1 2 3 4
6. I am in control of my thoughts during tests..... 1 2 3 4
7. The harder I work at taking a test, the more confused I get 1 2 3 4
8. I am confident in my abilities when I work on tests 1 2 3 4
9. Thoughts of doing poorly interfere with my concentration on tests 1 2 3 4
10. I find it easy to understand what the questions ask..... 1 2 3 4
11. I seem to defeat myself while working on important tests 1 2 3 4
12. I am able to keep a clear mind during tests 1 2 3 4
13. During tests I find myself thinking about the consequences of failing 1 2 3 4
14. I do all right on tests..... 1 2 3 4
15. During examinations I get so nervous that I forget facts I really know..... 1 2 3 4

Constructed Response Questions (20 marks)

All questions are worth two marks

Write your answers in the spaces that are provided.

Question 1 (C)

Describe a piece of evidence for each of these ideas;

- the Earth's continents were once a single land mass, and
- the Earth's plates are still in motion

1 2 3 4 5
 very ← → Quite
 little CONFIDENCE a lot

Question 2 (C)

Describe the direction that the Earth's plates move relative to each other at

- midocean ridges, and
- subduction zones.

1 2 3 4 5
 very ← → Quite
 little CONFIDENCE a lot

Question 3 (C)

Describe how shield and cone volcanoes differ from each other in their

- shapes (1 difference), and
- lava (2 differences).

1 2 3 4 5
 very ← → Quite
 little CONFIDENCE a lot

Question 4 (C)

Describe the main difference you would expect to observe between a piece of igneous intrusive rock and a piece of igneous extrusive rock. Name three types of geological structures that are formed by igneous intrusive rocks.

1 2 3 4 5
 very ←————→ Quite
 little CONFIDENCE a lot

Question 5 (C)

Describe the way cold air moves in a cold front and the conditions that are required for the formation of rain clouds.

1 2 3 4 5
 very ←————→ Quite
 little CONFIDENCE a lot

Question 6 (P)

Explain one difference and one similarity between the conditions that are required for the formation of fog and dew.

1 2 3 4 5
 very ←————→ Quite
 little CONFIDENCE a lot

Question 7 (P)

Locate point "X" on the following map and explain whether or not you would expect it to be an earthquake zone.

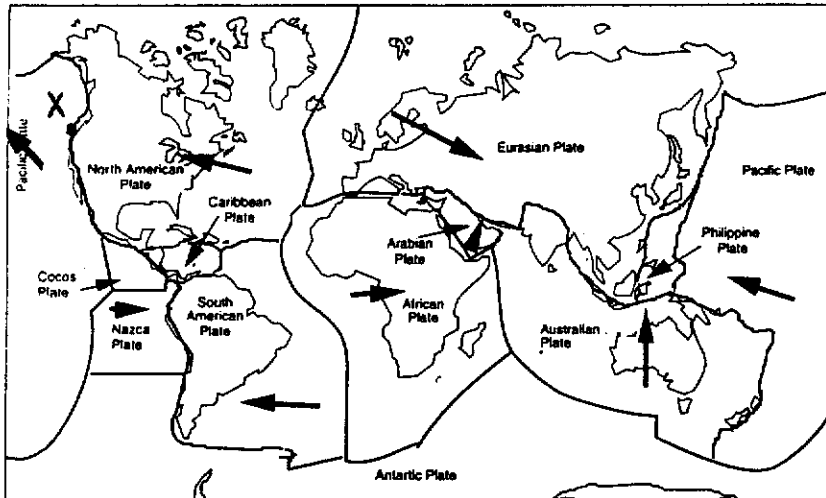
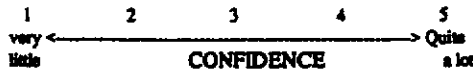


Fig. 5: The Earth's tectonic plates.



Question 8 (P)

A rock (shown below) has a light-grey colour and is made from very small crystals. It has numerous air holes that are about 2 mm in diameter. Explain, with your reasons, whether the rock is likely to be volcanic or sedimentary.

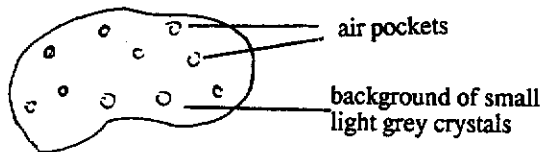
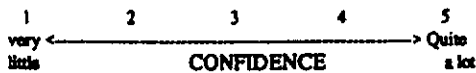


Fig. 6: Rock



Question 9 (P)
Study the following diagram.

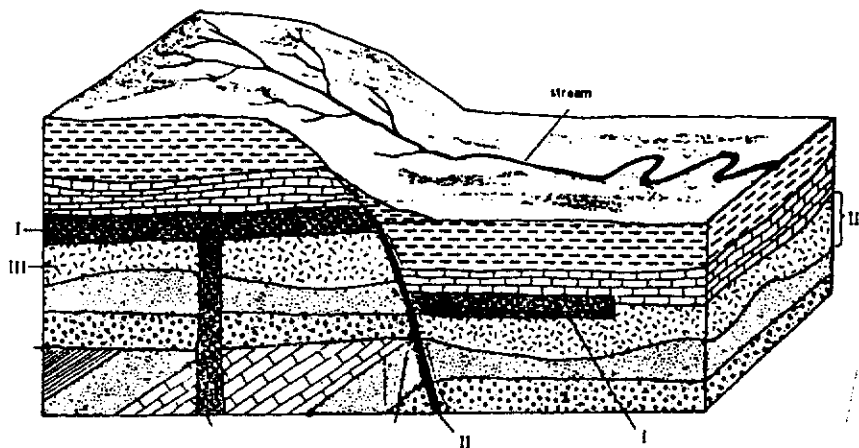


Fig. 7: A land surface and its underlying rocks.

Provide a name for structure II and describe the order in which structures I, II and III were formed.

1 2 3 4 5
very < > Quite
little CONFIDENCE a lot

Question 10 (P)

Study the following table.

Mineral	Hardness
Talc	1.0
Gypsum	2.0
Calcite	3.0
Fluorite	4.0
Apatite	5.0
Feldspar	6.0
Quartz	7.0

Table 2: Hardness of selected minerals

You have a mineral which you suspect is either chalcedony (hardness = 5.5) or olivine (hardness = 6.5 - 7.0). Explain how you could conduct a test (using minerals from the hardness table) to determine which mineral you have.

1 2 3 4 5
 very ← → Quite
 little CONFIDENCE a lot

MARKING SCHEDULE FOR *EARTH SCIENCE* TEST BOOKLETS

1A & 1B

Multiple-Choice Answers (20 marks)

1.	B	11.	D
2.	B	12.	B
3.	B	13.	D
4.	D	14.	C
5.	C	15.	A
6.	D	16.	C
7.	C	17.	B
8.	A	18.	A
9.	D	19.	B
10.	B	20.	A

Constructed-Response Answers (20 marks)

Question 1 (C)

- Any *one* of: outlines of continents appear to fit together, or different continents have matching deposits of fossils/ coal/ rock types/ areas of glaciation/ lines of magnetism, etc. (1)
- Any *one* of: seafloor is observed to be spreading/ moving apart, earthquakes are produced at the plate boundaries, mountain building occurs where the plates collide, consumption of crust at plate boundaries (1)

Question 2 (C)

- plates move apart from each other at mid-ocean ridges (1)
- plates move toward each other at subduction zones (1)

Question 3 (C)

- shield volcanoes have gently sloping sides and cone volcanoes have steep sides (0.5); shield volcanoes have low viscosity lava and cone volcanoes have high viscosity lava (0.5)

- lava differs in the minerals which are present (or amount of silica) (1)

Question 4 (C)

- intrusive rocks have large crystals and extrusive rocks have small crystals (1)
- Any three of: plutons, batholiths, sills, dykes (1)

Question 5 (C)

- cold air moves underneath the warmer air (1)
- the warm air must be moist (or relatively humid) (1)

Question 6 (P)

- similarity: decrease in temperature/ low temperature (1)
- difference: dew forms under still conditions and fog when there is a slight wind (1)

Question 7 (P)

- it would be an earthquake zone (1)
- plates must slide past each other for earthquakes to occur (1)

Question 8 (P)

- volcanic because of its properties e.g., crystals (1)
- it isn't sedimentary because it is not made up of grains (or similar) (1)

Question 9 (P)

- structure II is a fault (1)
- order: structures III, I, II (1)

Question 10 (P)

- select feldspar (1)
- if feldspar scratches the mineral then it is chalcedony, otherwise it is olivine (1)

EARTH SCIENCE TEST BOOKLET 2A

- N.B. i) Test booklet 2B is identical to test booklet 2A except that the order of the multiple-choice and the constructed-response sections are reversed.
- ii) This test booklet has been photoreduced in size.

Time: 45 minutes

Total marks: 40 (20 Content & 20 Process)

Instructions:

1. This paper consists of two sections.
2. Answer **all** questions in the order they are presented.
3. Please rate how confident you are that you can answer each question correctly **before** you start answering the question. Instructions for answering the confidence scale are provided on a separate sheet of paper.
4. Please do not move onto section 2 until you have completed the *Attitudes Toward Tests Survey*.
5. You have been provided with adequate time to complete the test and the survey questions.

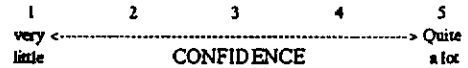
Multiple Choice Questions (20 marks)
All questions are worth one mark

Circle the **best** answer

Question 1 (C)

Which of the following does **NOT** provide evidence that the continents were once a single land mass?

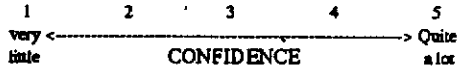
- A. Matching fossil deposits on different continents.
- B. Similar rock types in Canada and Scotland.
- C. Convection currents underneath the crust.
- D. Lines of magnetism in the rocks of the Earth.



Question 2 (C)

Evidence that the tectonic plates are still in motion is provided by ...

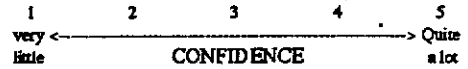
- A. the flow of heat from the Earth's core.
- B. earthquakes in the middle of plates.
- C. the formation of folded rocks.
- D. the formation of deep sea ridges.



Question 3 (C)

At the midocean ridges, the Earth's plates move ...

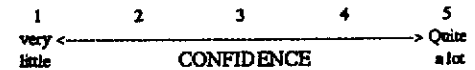
- A. toward each other.
- B. apart from each other.
- C. past each other in opposite directions.
- D. past each other at different speeds.



Question 4 (C)

At subduction zones the Earth's plates move ...

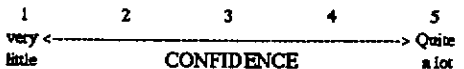
- A. toward each other.
- B. apart from each other.
- C. past each other in opposite directions.
- D. past each other at different speeds.



Question 5 (C)

Shield volcanoes differ from cone volcanoes because they have ...

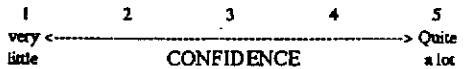
- A. less viscous lava and steep sides.
- B. less viscous lava and gently sloping sides.
- C. more viscous lava and steep sides.
- D. more viscous lava and gently sloping sides.



Question 6 (C)

A shield volcano has a different shape than a cone volcano because its lava ...

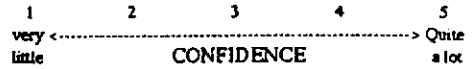
- A. has a higher temperature.
- B. produces a weaker eruption.
- C. contains different minerals.
- D. contains less gas.



Question 7 (C)

Igneous intrusive rocks differ from igneous extrusive rocks in that the former have ...

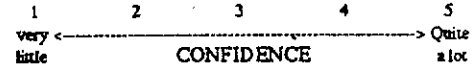
- A. a lighter colour.
- B. air pockets.
- C. larger crystals.
- D. a higher density.



Question 8 (C)

Igneous intrusive rocks are found as ...

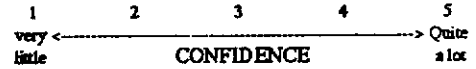
- A. volcanic ash, lava flows and bathyliths.
- B. sills, dykes and bathyliths.
- C. volcanic bombs, sills and bathyliths.
- D. sills, dykes, plutons and lava flows.



Question 9 (C)

When a cold front passes, the cold air moves ...

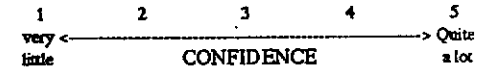
- A. past the warmer air.
- B. underneath the warmer air.
- C. through the warmer air.
- D. over the top of the warmer air.



Question 10 (C)

A cold front produces rain clouds when the ...

- A. cold air contains water vapour.
- B. warm air is less dense than the cold air.
- C. relative humidity of the warm air is high.
- D. cold air is moving faster than the warm air.

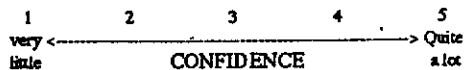


Question 11 (P)

On a clear still night, the air temperature drops from 10°C at sunset to 2°C at midnight.

These conditions favour the formation of ...

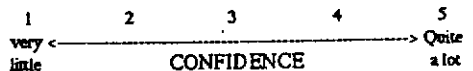
- A. frost.
- B. dew.
- C. fog.
- D. frozen dew.



Question 12 (P)

On a slightly windy night, the air temperature drops to 2°C. These conditions favour the formation of ...

- A. frost.
- B. dew.
- C. fog.
- D. frozen dew.



Questions 13 & 14 refer to the following map:

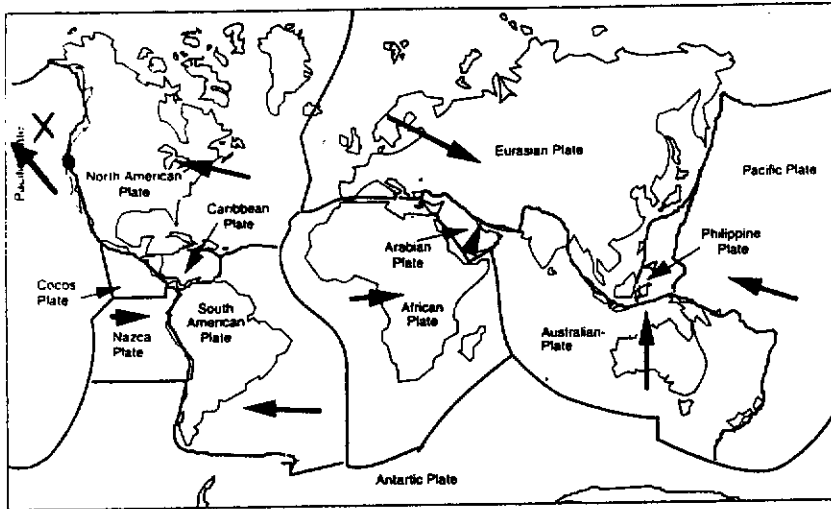
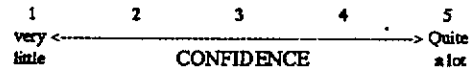


Fig. 1: The Earth's tectonic plates

Question 13 (P)

The most likely geological activity at point X is the ...

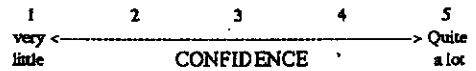
- A. formation of midocean ridges.
- B. formation of mountains.
- C. production of earthquakes.
- D. production of folded rocks.



Question 14 (P)

The geological activity at point X is caused by the ...

- A. differing densities of the plates.
- B. differing speeds of the plates.
- C. the production of magma.
- D. differing composition of each plate.



Questions 15 & 16 refer to the following information:

Rock I is a light-grey colour. It has very small interlocking crystals and numerous air holes that are about 2 mm in diameter.

Rock II has a layered appearance. It crumbles into rounded grains and it has fossils.

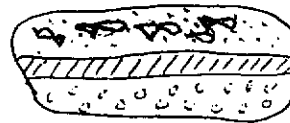
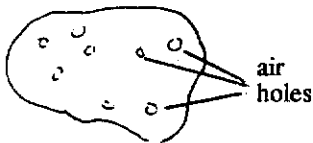


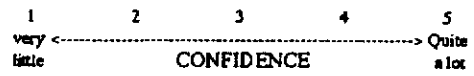
Fig. 2: Rock 1

Fig. 3: Rock 2

Question 15 (P)

Rock I is most likely ...

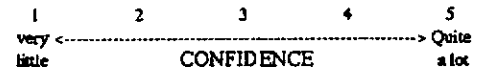
- A. plutonic.
- B. volcanic.
- C. sedimentary.
- D. metamorphic.



Question 16 (P)

Rock II is most likely ...

- A. plutonic.
- B. volcanic.
- C. sedimentary.
- D. metamorphic.



Questions 17 & 18 refer to the following diagram

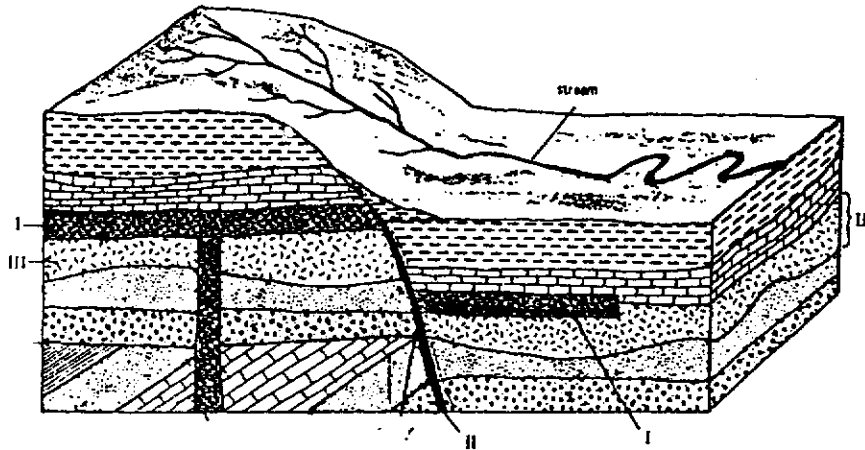
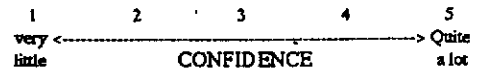


Fig. 4: A land surface and its underlying rocks.

Question 17 (P)

The structure marked "II" is ...

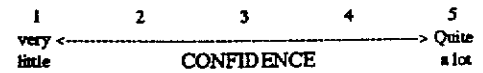
- A. a fault.
- B. an escarpment.
- C. a sill.
- D. a dyke.



Question 18 (P)

Structure I is most likely to have formed ...

- A. before structures II and III.
- B. after structures II and III.
- C. before structure II but after structure III.
- D. before structure III but after structure II.



Questions 19 & 20 refer to the following table:

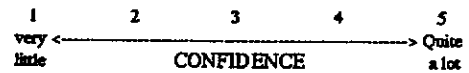
Mineral	Hardness
Talc	1.0
Gypsum	2.0
Calcite	3.0
Fluorite	4.0
Apatite	5.0
Feldspar	6.0
Quartz	7.0

Table 1: Hardness of selected minerals

Question 19 (P)

Limonite, which has a hardness of 5.3, will scratch ...

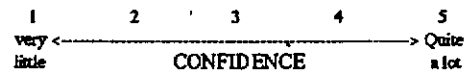
- A. fluorite but not apatite.
- B. fluorite but not feldspar.
- C. feldspar but not calcite.
- D. neither fluorite nor calcite.



Question 20 (P)

A mineral that can be used to distinguish between olivine (hardness = 6.5 - 7.0) and chalcedony (hardness = 5.5) is

- A. fluorite.
- B. apatite.
- C. quartz.
- D. feldspar.



ATTITUDES TOWARD TESTS SURVEY

A number of statements which people have used to describe themselves are given below. Read each of the statements and circle the number which describes how you feel about the test questions you have just completed.

1 = Almost Never, 2 = Sometimes, 3 = Often, 4 = Almost Always

There are no right or wrong answers. Do not spend too much time on one statement but give the answer which seems to describe how you feel. Please answer every statement.

1. Thinking about my marks in a subject interferes with my work on tests... 1 2 3 4
2. I do as well, or better than others on tests 1 2 3 4
3. My mind freezes up on important exams 1 2 3 4
4. I don't have to rush through questions 1 2 3 4
5. During exams I find myself thinking about whether I'll ever
get through school 1 2 3 4
6. I am in control of my thoughts during tests..... 1 2 3 4
7. The harder I work at taking a test, the more confused I get 1 2 3 4
8. I am confident in my abilities when I work on tests 1 2 3 4
9. Thoughts of doing poorly interfere with my concentration on tests 1 2 3 4
10. I find it easy to understand what the questions ask..... 1 2 3 4
11. I seem to defeat myself while working on important tests 1 2 3 4
12. I am able to keep a clear mind during tests 1 2 3 4
13. During tests I find myself thinking about the consequences of failing 1 2 3 4
14. I do all right on tests..... 1 2 3 4
15. During examinations I get so nervous that I forget facts I really know..... 1 2 3 4

Constructed Response Questions (20 marks)
All questions are worth two marks

Write your answers in the spaces that are provided.

Question 1 (C)

Describe two features of the Earth's mantle.

1 2 3 4 5
 very ← → Quite
 little CONFIDENCE a lot

Question 2 (C)

Provide brief descriptions for the colour of the following rocks and the size of their crystals;

- basalt, and
- granite.

1 2 3 4 5
 very ← → Quite
 little CONFIDENCE a lot

Question 3 (C)

Describe the conditions that are required for the formation of;

- igneous intrusive rocks, and
- metamorphic rocks.

1 2 3 4 5
 very ← → Quite
 little CONFIDENCE a lot

Question 4 (C)

Describe two features that you would usually expect to observe in;

- metamorphic rocks, and
- sedimentary rocks.

1 2 3 4 5
 very ← → Quite
 little CONFIDENCE a lot

Question 5 (C)

Explain what it means about an air mass if its relative humidity is measured to be 70 per cent. Describe what happens to the temperature and the relative humidity of an air mass when it ascends.

1 2 3 4 5
 very ← → Quite
 little CONFIDENCE a lot

Question 6 (P)

Study the following map.

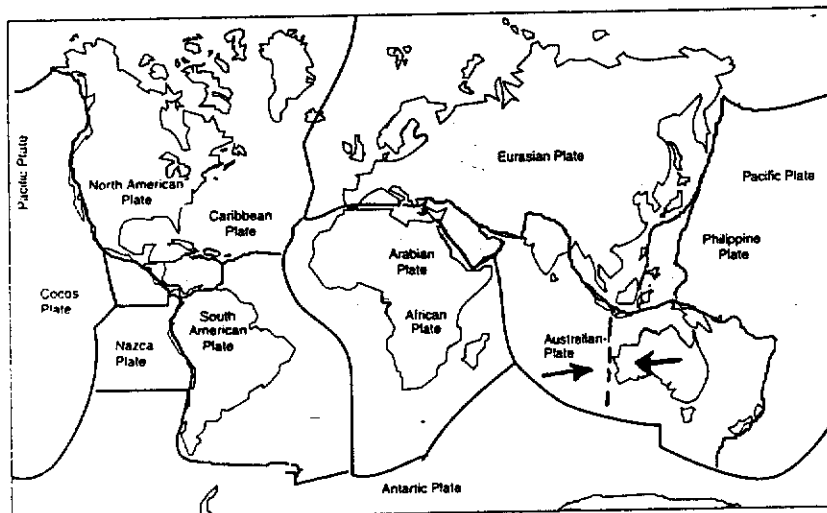


Fig. 5: The Earth's tectonic plates

Imagine that a new plate boundary, shown on the map as ----, divides the Australian plate in two. Arrows show the directions in which the new plates move.

Predict, with your reasons, the most likely change you would expect to observe at the new plate boundary.

1 2 3 4 5
 very ← → Quite
 little CONFIDENCE a lot

Question 7 (P)

Refer to the following table and explain which properties would be the most useful for telling the difference between quartz and olivine. Which properties would be the least useful?

Mineral	Colour	Hardness	Lustre	Streak
Apatite	W, Gr, B, Y, P	4.5-5	R to V	W
Asbestos	W, Gr	5-6	P to V	W
Augite	Bl, C, Br, Gr	6-6.5	R to V	G-W
Biotite	Bl, Br, Gr	2.7-3.1	V	C
Calcite	C, W, B	3	D to V	W
Cassiterite	Br, Bl, R, G, W, Y	6-7	B	W, Br
Chalcopyrite	Y	3.5-4	M	Gr, Bl
Chlorite	G	2-2.5	P	W, Gr
Feldspar	W, Pink	6	P to V	C
Fluorite	W, Y, Gr, R, P, B	4	V	W
Galenite	G	2.5	M	G
Garnet	R, Br, Y, W, Gr	6.5-7.5	R to V	W
Gypsum	W, C, G, Y, R	1.5-2	P to V	W
Haematite	G	5-6	M	R, Br
Halite	C, R, B, P, W	2.5	V	W
Hornblende	Bl, Gr-Bl	5-6	V	Br, Bl
Jasper	W, C, Br, R, B	7	D to V	W
Limonite	Br, Y, Bl	4.5-5	D to V	Br, Y, W
Magnetite	Bl	5.5-6.5	M	Bl
Muscovite	Gr, Br, Y, C	2.7-3.0	V	C
Olivine	Gr, Br	6.5-7	V	C
Pyrite	Y	6-6.5	M	Gr
Quartz	C, P, R, Y	7	V	W
Sphalerite	Y, Br, R, Bl, W	3.5-4	R to B	W
Talc	W, Gr	1	P	W

Abbreviation Code: Colour and streak: B = blue, Bl = black, Br = brown, C = colourless, G = grey, Gr = green, P = purple, R = red, W = white, Y = yellow
 Lustre: R = resinous, V = vitreous, D = dull, M = metallic, P = pearly, S = brilliant

Table 2: Properties of selected minerals

1 2 3 4 5
 very ← → Quite
 little CONFIDENCE a lot

Question 8 (P)

Study these two diagrams and explain how each of these structures may have formed.

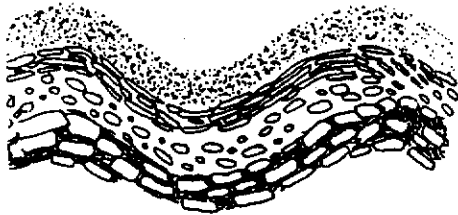


Fig. 6: Structure 1

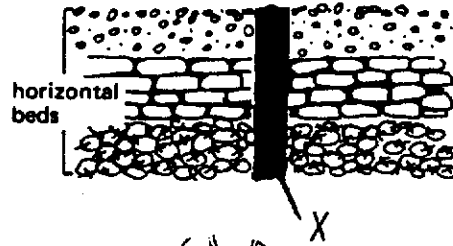


Fig. 7: Structure 2 (= "X")

1	2	3	4	5	
←				→	
very				Quite	
little		CONFIDENCE		a lot	

Question 9 & 10 Refer to the following weather map.

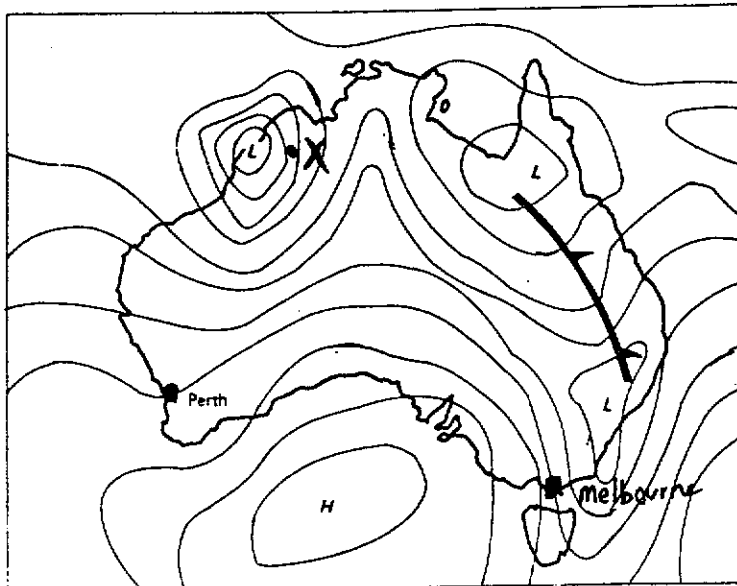


Fig. 8: Australian weather map, 6th February 1996.

Question 9 (P)

Locate Perth and Melbourne and compare their

- air pressures, and
- the strength of their winds.

1 2 3 4 5
very ←-----→ Quite
little CONFIDENCE a lot

Question 10 (P)

Locate point X on the weather map and predict

- the direction of the wind at that point, and
- the weather conditions at that point.

1 2 3 4 5
very ←-----→ Quite
little CONFIDENCE a lot

MARKING SCHEDULE FOR *EARTH SCIENCE* TEST BOOKLETS
2A & 2B

Multiple-Choice Answers (20 marks)

1. C	11. B
2. A	12. C
3. B	13. C
4. A	14. B
5. B	15. B
6. C	16. C
7. C	17. A
8. B	18. C
9. B	19. B
10. C	20. D

Constructed-Response Answers (20 marks)

Question 1 (C)

- Any *two* of; supports the crust / lies underneath the crust; has a density of about 4 gram/cm³; is solid but it flows. (2)

Question 2 (C)

- Granite has a light colour and large crystals (1)
- Basalt has a dark colour and small crystals (1)

Question 3 (C)

- magma cools slowly underground (1)
- rocks are subjected to heat and pressure (1)

Question 4 (C)

- Any *two* of; crumbly appearance, layers; has fossils, made up of grains (1)
- Any *two* of; presence of crystals, hard or sharp edges, foliation (or layering) (1)

Question 5 (C)

- The air holds 70% of its water capacity (1)

- The air mass cools and its relative humidity increases (1)

Question 6 (P)

- formation of a deep sea trench (1)
- the oceanic crust will pass under the continental crust because it is less dense (1)

Question 7 (P)

- most useful: colour and streak. (1)
- least useful: hardness and lustre. (1)

Question 8 (P)

- structure 1 formed by the action of horizontal forces. (1)
- structure 2 formed by movement of magma through a crack. (1)

Question 9 (P)

- have the same air pressure (1)
- Perth is less windy than Melbourne (1)

Question 10 (P)

- winds from the north (1)
- any *two* of: warm, windy and wet (1)

APPENDIX D
TEST BOOKLETS AND MARKING SCHEDULES
FOR THE *ORGANISMS AND FOOD TESTS*

ORGANISMS AND FOOD TEST BOOKLET 1A

- N.B. i) Test booklet 1B is identical to test booklet 1A except that the order of the multiple-choice and the constructed-response sections are reversed.
- ii) This test booklet has been photoreduced in size.

Time: 45 minutes

Total marks: 40 (20 Content & 20 Process)

Instructions:

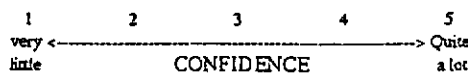
1. This paper consists of two sections.
2. Answer all questions in the order they are presented.
3. Please rate how confident you are that you can answer each question correctly **before** you start answering the question. Instructions for answering the confidence scale are provided on a separate sheet of paper.
4. Please do not move onto section 2 until you have completed the *Attitudes Toward Tests Survey*.
5. You have been provided with adequate time to complete the test and the survey questions.

Multiple Choice Questions (20 marks)
All questions are worth one mark

Circle the **best** answer

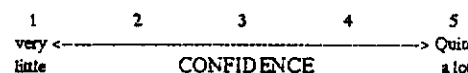
Question 1 (C)

- Chemical digestion refers to the
 A. intake of any type of food by an animal.
 B. conversion of food into simple molecules.
 C. intake, breakdown and use of food by consumers.
 D. muscular activity of the digestive system.



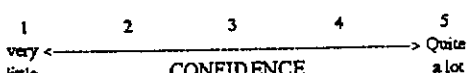
Question 2 (C)

- Chemical digestion takes place in the
 A. stomach, liver, mouth & esophagus.
 B. duodenum, stomach & mouth.
 C. esophagus, mouth, stomach & duodenum.
 D. pancreas, esophagus, duodenum & stomach.



Question 3 (C)

- Digestive enzymes are produced by the ...
 A. pancreas, liver & small intestine.
 B. salivary glands, liver & duodenum.
 C. stomach wall, salivary glands & large intestine.
 D. salivary glands, stomach wall & pancreas.

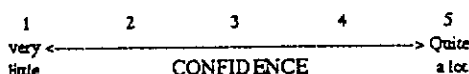


Question 4 refers to the following table from a student's workbook.

	<u>Food</u>	<u>Digested form</u>	<u>Enzyme</u>
I	starch	glycogen	sucrase
II	proteins	amino acids	aminoase
III	fats	fatty acids & glycerol	lipase
IV	cellulose	simple sugars	amylase

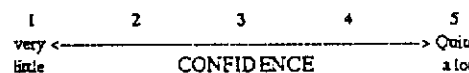
Question 4 (C)

- The row that is correct for all three columns is row ...
 A. I.
 B. II.
 C. III.
 D. IV.



Question 5 (C)

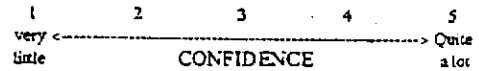
- Oxygen is transported around the body by ...
 A. the plasma.
 B. the lymph.
 C. red blood cells.
 D. white blood cells.



Question 6 (C)

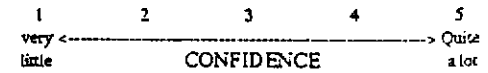
Carbon dioxide is transported around the body by

- A. the plasma.
- B. the lymph.
- C. red blood cells.
- D. white blood cells.

Question 7 (C)

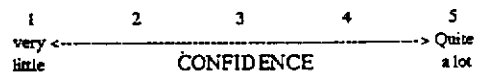
Transpiration refers to the ...

- A. loss of water from the leaves of a plant.
- B. upward movement of water in the stem of a plant.
- C. movement of water from the soil into the roots of a plant.
- D. evaporation of water from the surface of a plant.

Question 8 (C)

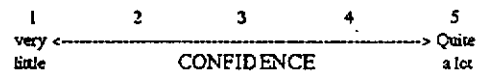
Translocation refers to the movement of ...

- A. water into all parts of a plant.
- B. food from a plant's leaves to other areas.
- C. gases into and out of a plant's leaves.
- D. starch into a plant's leaves.

Question 9 (C)

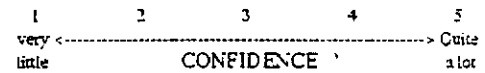
A tapeworm in an animal's intestine is an example of ...

- A. predation.
- B. mutualism.
- C. commensalism.
- D. parasitism.

Question 10 (C)

Parasitism is a type of feeding relationship between two species where ...

- A. one species benefits and the other species dies.
- B. one species benefits and the other is unaffected.
- C. both species benefit from each other.
- D. one species benefits and the other species suffers.

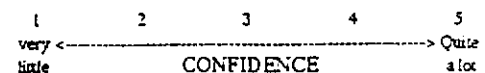
Questions 11 & 12 refer to the following facts about digestion.

- I. enzymes speed up digestion.
- II. food is converted into simpler molecules.
- III. food is broken into smaller pieces.
- IV. muscular activity mixes food with liquids.
- V. it involves physical and chemical changes.

Question 11 (P)

Facts that are TRUE for chemical digestion but NOT TRUE for mechanical digestion are ...

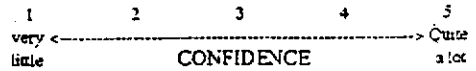
- A. I & II only.
- B. III & IV only.
- C. I, II & IV.
- D. I, III, IV.



Question 12 (C)

Facts that are **TRUE** for mechanical digestion but **NOT TRUE** for chemical digestion are ...

- A. I & II only.
- B. III & IV only.
- C. I, II & IV.
- D. I, III, IV.



Questions 13 & 14 refer to the following information.

An experiment was set up as follows. Ground up meat, starch and some gastric juices from a person's stomach were placed into a 'sausage' made out of dialysis tubing (a semi-permeable membrane). The sausage was placed into a beaker of water and the temperature was kept at 37°C.

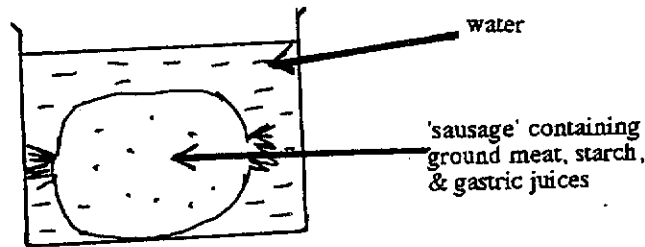
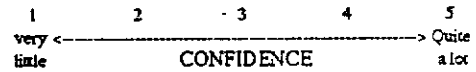


Fig.1: Experimental set up for questions 13 & 14.

Question 13 (P)

After several hours, the liquid outside the "sausage" would contain ...

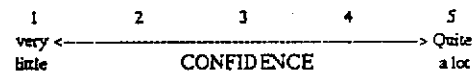
- A. amino acids.
- B. simple sugars.
- C. starch.
- D. ground up meat.



Question 14 (P)

The experiment was conducted at 37°C (body temperature) so that the

- A. pores in the semi-permeable membrane could expand.
- B. semi-permeable membrane would become soft.
- C. liquid substances in the sausage could expand.
- D. gastric juices could work most efficiently.



Questions 15 & 16 refer to the following information

Leaves from a variegated plant, that is a plant which has leaves that are green and white coloured (see diagram) were bleached and tested with iodine solution. The parts of the leaves that were originally green turned a blue-black colour. It was observed that the white parts of the leaves changed to a yellow-brown colour.

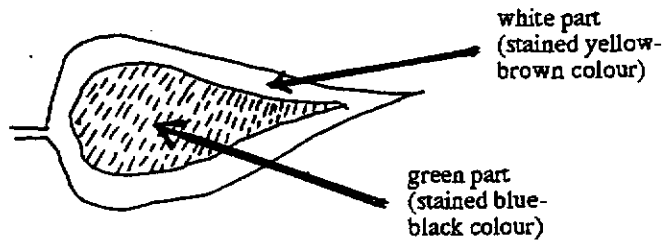
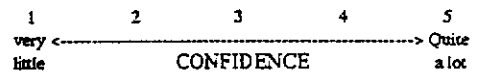


Fig. 2: Leaf from a variegated plant treated with iodine.

Question 15 (P)

The results from this experiment suggest that the green part of the leaf is involved in the ...

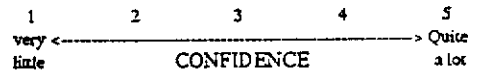
- A. storage of fat.
- B. storage of starch.
- C. manufacture of glucose.
- D. manufacture of chlorophyll.



Question 16 (P)

The results from this experiment for the yellow part of the leaf suggest that ...

- A. light is required for the manufacture of glucose.
- B. chlorophyll is needed for photosynthesis.
- C. there are fewer stomates in the yellow part of the leaf.
- D. iodine does not react with chlorophyll.



Questions 17 & 18 refer to the following diagram

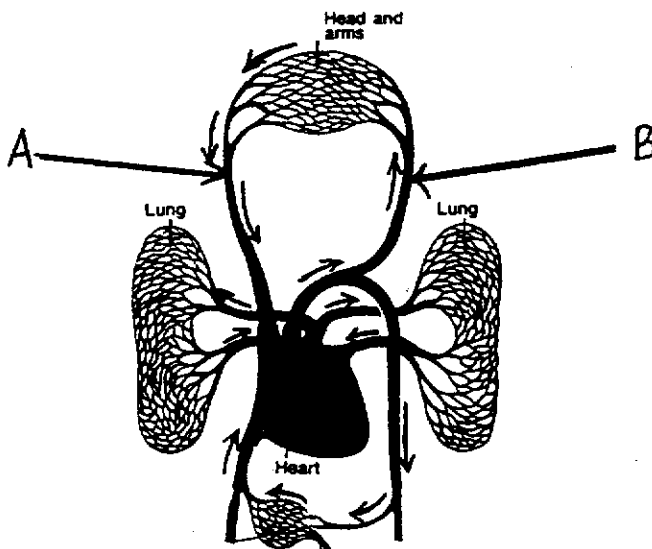
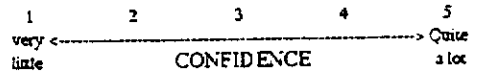


Fig. 3: The circulatory system.

Question 17 (P)

Blood at point A would be ...

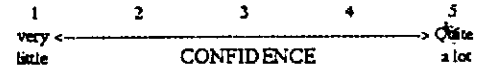
- A. oxygen rich and high in nutrients.
- B. oxygen rich & low in nutrients.
- C. oxygen depleted & low in nutrients.
- D. oxygen depleted & high in nutrients.



Question 18 (P)

Blood at point B would be ...

- A. oxygen rich and high in nutrients.
- B. oxygen rich & low in nutrients.
- C. oxygen depleted & low in nutrients.
- D. oxygen depleted & high in nutrients.



Questions 19 & 20 refer to the diagram of a cactus plant which lives in the hot, dry and harsh conditions of the desert.

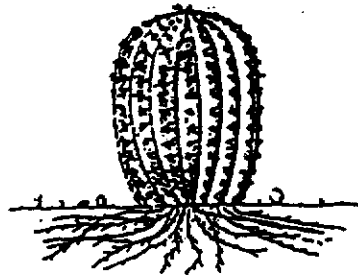
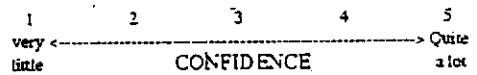


Fig. 4: Cactus plant.

Question 19 (P)

The most likely reason the cactus has a waxy coat is to ...

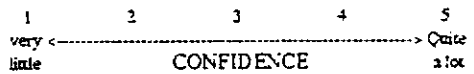
- A. keep the plant cool by reflecting sunlight.
- B. cause any moisture that falls on its surface to 'bead'.
- C. decrease water loss through transpiration.
- D. produce a tough skin for protection.



Question 20 (P)

The most likely reason the cactus has shallow spreading roots is to...

- A. provide anchorage in loose desert sand.
- B. increase the area over which water can be absorbed.
- C. enable the storage of food close to the ground.
- D. assist with the cooling of the cactus.



ATTITUDES TOWARD TESTS SURVEY

A number of statements which people have used to describe themselves are given below. Read each of the statements and circle the number which describes how you feel about the test questions you have just completed.

1 = Almost Never, 2 = Sometimes, 3 = Often, 4 = Almost Always

There are no right or wrong answers. Do not spend too much time on one statement but give the answer which seems to describe how you feel. Please answer every statement.

1. Thinking about my marks in a subject interferes with my work on tests... 1 2 3 4
2. I do as well, or better than others on tests 1 2 3 4
3. My mind freezes up on important exams 1 2 3 4
4. I don't have to rush through questions 1 2 3 4
5. During exams I find myself thinking about whether I'll ever
get through school 1 2 3 4
6. I am in control of my thoughts during tests..... 1 2 3 4
7. The harder I work at taking a test, the more confused I get 1 2 3 4
8. I am confident in my abilities when I work on tests 1 2 3 4
9. Thoughts of doing poorly interfere with my concentration on tests 1 2 3 4
10. I find it easy to understand what the questions ask..... 1 2 3 4
11. I seem to defeat myself while working on important tests 1 2 3 4
12. I am able to keep a clear mind during tests 1 2 3 4
13. During tests I find myself thinking about the consequences of failing 1 2 3 4
14. I do all right on tests..... 1 2 3 4
15. During examinations I get so nervous that I forget facts I really know..... 1 2 3 4

Constructed Response Questions (20 marks)
All questions are worth two marks

Write your answers in the space that is provided.

Question 1 (C)

Describe

- the importance of food to the human body, and
- the role of carbohydrates in human diet.

1 2 3 4 5
very <-----> Quite
little CONFIDENCE a lot

Question 2 (C)

Describe the pathway that

- food takes from the time it enters the mouth and leaves as wastes at the anus.
- the blood takes through the circulatory system using the terms body, lungs and the names for the four chambers of the heart.

1 2 3 4 5
very <-----> Quite
little CONFIDENCE a lot

Question 3 (C)

- Write out the word equation for respiration, and
- state two facts about respiration.

1 2 3 4 5
very <-----> Quite
little CONFIDENCE a lot

Question 4 (C)

- What is the original source of energy for all food chains in an ecosystem?
- Describe the direction that energy flows through an ecosystem which contains herbivores, carnivores and producers.

1 2 3 4 5
very <-----> -----> Quite
little CONFIDENCE a lot

Question 5 (C)

- Explain the meaning of the word 'ecological community' with reference to an example.

1 2 3 4 5
very <-----> -----> Quite
little CONFIDENCE a lot

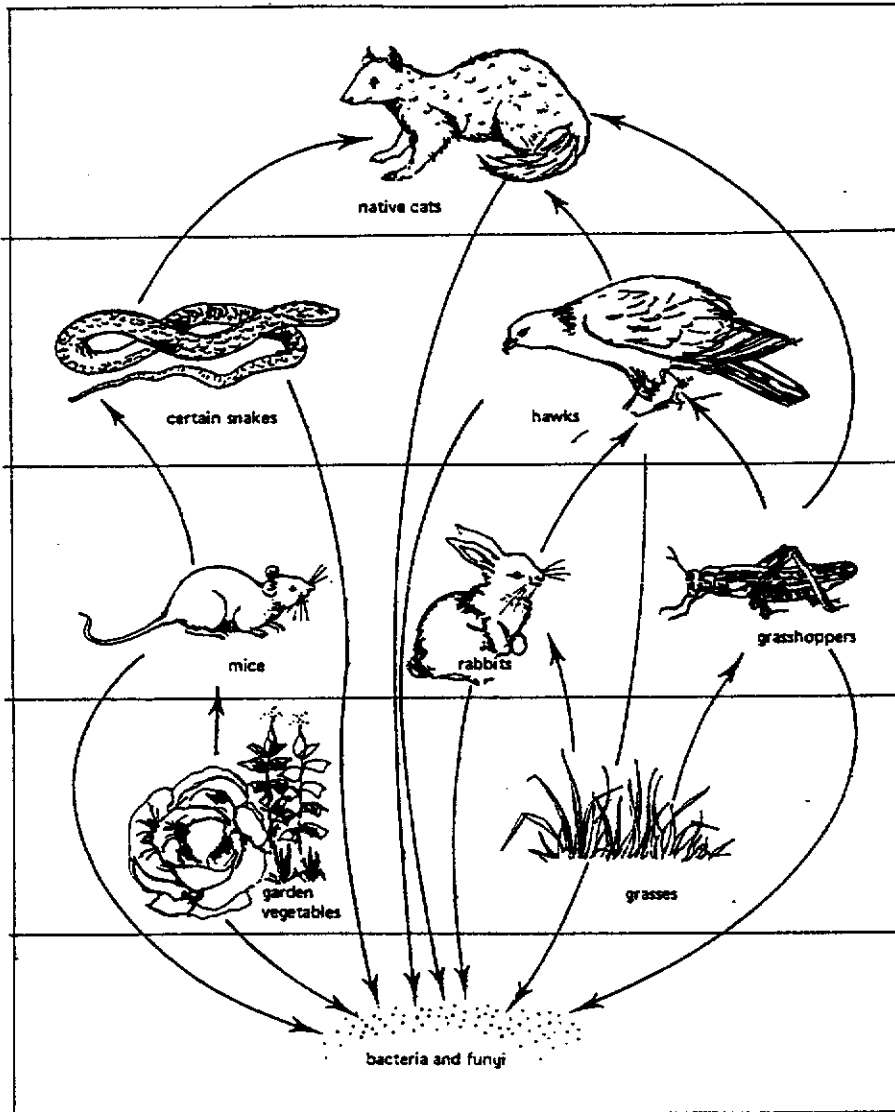
Question 6 (P)

- Describe two differences between
- phloem vessels and xylem vessels.
 - arteries and veins.

1 2 3 4 5
very <-----> -----> Quite
little CONFIDENCE a lot

Question 9 (P)

Study this food web.

**Fig. 6:** Food web for an Australian locality.

Name _____

- two organisms that compete with each other for their food.
- the decomposers in the food web.

1 2 3 4 5
very <-----> Quite
little CONFIDENCE a lot

Question 10 (P)

Refer to the food web in the previous question.

Imagine that hawks began to eat mice instead of grasshoppers. Predict, with your reasons, the effect that this change has on the numbers of snakes in the ecosystem.

1 2 3 4 5
very <-----> Quite
little CONFIDENCE a lot

MARKING SCHEDULE FOR *ORGANISMS AND FOOD* TEST BOOKLETS
1A & 1B

Multiple-Choice Answers (20 marks)

1.	B	11.	A
2.	B	12.	B
3.	D	13.	A
4.	C	14.	D
5.	C	15.	B
6.	A	16.	B
7.	A	17.	C
8.	B	18.	A
9.	D	19.	C
10.	D	20.	B

Constructed-Response Answers (20 marks)

Question 1 (C)

- provides energy (0.5) and growth/repair of body (0.5)
- provision of energy (1)

Question 2 (C)

- mouth, esophagus, stomach, small intestine, large intestine, rectum, anus (1, 0.5 if 1 or 2 are missing or out of order)
- body, right atrium, right ventricle, lungs, left atrium, left ventricle (1, deduct 0.5 if ventricles & atria out of order, deduct 0.5 if lungs and body are out of order)

Question 3 (C)

- oxygen + glucose ----> carbon dioxide + water (1, 0.5 if one substance wrong or missing)
- takes place in both plants and animals, takes place in plants in both day and night time, takes place in cells, is the reverse of photosynthesis, etc (0.5 for any one, maximum of 1)

Question 4 (C)

- the sun (1)
- producers ----> herbivores ----> carnivores (1)

Question 5 (C)

- all the plants and animals in an ecosystem which depend on each other (1)
- any appropriate example (1)

Question 6 (P)

- phloem transports nutrients and xylem transports water, phloem vessels travel downward and xylem vessels travel upward, etc (0.5 each, maximum of 1)
- arteries lead away from the heart and veins lead toward the heart, blood is under high pressure in the arteries and low pressure in the veins, veins have valves whereas arteries do not, etc (0.5 each, maximum of 1)

Question 7 (P)

- nutrients required for respiration, nutrients transported to individual cells, etc (1 mark for one similarity)
- plants don't have a pump, plant nutrients are not the products of digestion, etc (1 for one difference)

Question 8 (P)

- enzymes in saliva help digest the starch (1, 0.5 if reference to saliva only)
- dog saliva does not contain enzymes that help digest starch (1)

Question 9 (P)

- rabbits & grasshoppers (1)
- bacteria & fungi (1)

Question 10 (P)

- snakes decrease (1)
- less food available for snakes (1)

ORGANISMS AND FOOD TEST BOOKLET 2A

- N.B. i) Test booklet 2B is identical to test booklet 2A except that the order of the multiple-choice and the constructed-response sections are reversed.
- ii) This test booklet has been photoreduced in size.

Time: 45 minutes

Total marks: 40 (20 Content & 20 Process)

Instructions:

1. This paper consists of two sections.
2. Answer all questions in the order they are presented.
3. Please rate how confident you are that you can answer each question correctly before you start answering the question. Instructions for answering the confidence scale are provided on a separate sheet of paper.
4. Please do not move onto section 2 until you have completed the *Attitudes Toward Tests Survey*.
5. You have been provided with adequate time to complete the test and the survey questions.

Multiple Choice Questions (20 marks)
All questions are worth one mark

Circle the **best** answer

Questions 1 & 2 refer to the following information.

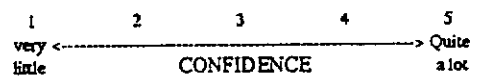
Two students wrote out this list of reasons for the importance of food.

- I. It is important for growth.
- II. It provides energy.
- III. It is needed for repairing damaged cells.
- IV. It is important for breathing.
- V. It provides materials that replace dead cells.

Question 1 (C)

The correct reasons are ...

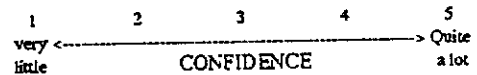
- A. I, II & IV.
- B. I, II, III & V.
- C. I & II only.
- D. I, III, IV & V.



Question 2 (C)

Carbohydrates are important because of reason ...

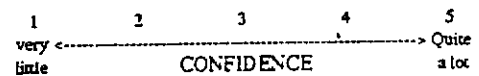
- A. I.
- B. II.
- C. IV.
- D. V.



Question 3 (C)

The correct order for the structures of the digestive system is ...

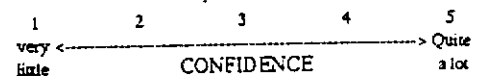
- A. mouth, stomach, esophagus, large intestine, small intestine, rectum, anus.
- B. mouth, esophagus, stomach, large intestine, small intestine, rectum, anus.
- C. mouth, stomach, esophagus, small intestine, large intestine, rectum, anus.
- D. mouth, esophagus, stomach, small intestine, large intestine, rectum, anus.



Question 4 (C)

The correct order that blood flows through the circulatory system is ...

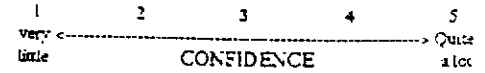
- A. lungs, right atrium, right ventricle, body, left atrium, left ventricle.
- B. lungs, left ventricle, left atrium, body, right ventricle, right atrium.
- C. lungs, left atrium, left ventricle, body, right atrium, right ventricle.
- D. lungs, right ventricle, right atrium, body, left ventricle, left atrium.



Question 5 (C)

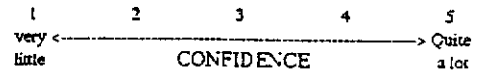
The correct equation for respiration is ...

- A. carbon dioxide + water + light ----> glucose + oxygen.
 B. carbon dioxide + oxygen + light ----> glucose + water.
 C. oxygen + glucose ----> carbon dioxide + water + energy.
 D. oxygen + carbon dioxide ----> glucose + water + energy.

Question 6 (C)

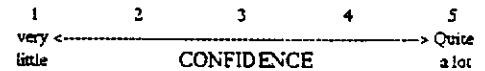
Which of the following statements is NOT TRUE. The process of respiration ...

- A. takes place in plants only during the dark.
 B. takes place in both plants and animals.
 C. takes place in cells.
 D. is the reverse of photosynthesis.

Question 7 (C)

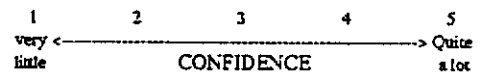
The energy for all food chains comes from ...

- A. the sun.
 B. green plants.
 C. consumers.
 D. carbon dioxide.

Question 8 (C)

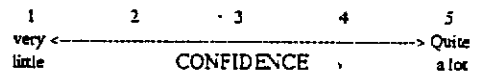
The flow of energy in a food chain is best shown by

- A. producer ----> carnivore ----> herbivore.
 B. producer ----> herbivore ----> carnivore.
 C. herbivore ----> carnivore ----> producer.
 D. carnivore ----> producer ----> herbivore.

Question 9 (C)

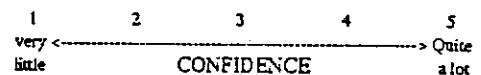
The term 'ecological community' refers to all the organisms ...

- A. of a particular species which live in an area.
 B. of a particular species which depend on each other.
 C. living in an area and their physical environment.
 D. of the different species which live in an area.

Question 10 (C)

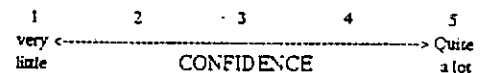
An example of a community is ...

- A. all the living organisms in a rotting log.
 B. a single bacterium in a pool of water.
 C. all the bees living a particular hive.
 D. all the organisms, air, rocks and dead leaves on a forest floor.

Question 11 (P)

Xylem and phloem vessels in plants differ from each other in that ...

- A. xylem vessels carry water upward and the phloem carries food downward.
 B. xylem vessels carry food upward and the phloem carries water downward.
 C. xylem vessels carry water downward and the phloem carries food upward.
 D. xylem vessels carry food downward and the phloem carries water upward.



Question 12 (P)

Arteries and veins differ from each other in that ...

- A. arteries carry oxygenated blood away from the heart and veins carry deoxygenated blood toward the heart.
- B. arteries carry deoxygenated blood under low pressure and veins carry oxygenated blood under high pressure.
- C. arteries have thin walls and they carry oxygen whereas veins have thick walls and they carry carbon dioxide.
- D. arteries carry blood under high pressure away from the heart and veins carry blood under low pressure toward the heart.

1 2 3 4 5
 very ←-----> Quite
 little CONFIDENCE a lot

Questions 13 & 14 refer to the following information

Here is a list of facts about nutrient (i.e. food) transport. Some of these statements apply for plants, some for animals and some for both plants and animals.

- I. only simple food chemicals are transported.
- II. nutrients are obtained from digestion.
- III. nutrients are carried to individual cells.
- IV. nutrients are pumped through the transport system.

Question 13 (P)

Facts about food transport that apply for both plants and animals are ...

- A. I & II only.
- B. I & III only.
- C. I, II & III.
- D. I, III & IV.

1 2 3 4 5
 very ←-----> Quite
 little CONFIDENCE a lot

Question 14 (P)

Which of the following facts about food transport are TRUE for animals but NOT TRUE for plants?

- A. II only
- B. II & IV.
- C. III & IV.
- D. IV only.

1 2 3 4 5
 very ←-----> Quite
 little CONFIDENCE a lot

Questions 15 & 16 refer to the following information

A student tried to dissolve some starch in boiling water and the result was a cloudy-white mixture. When she added iodine solution to some of the mixture it turned a blue-black colour. She then added some saliva from her mouth. After about 20 minutes the mixture turned a clear yellow-brown colour.

The girl repeated the experiment using saliva from her pet dog. This time she noticed that the starch mixture remained the same blue-black colour 20 minutes after adding dog saliva to it.

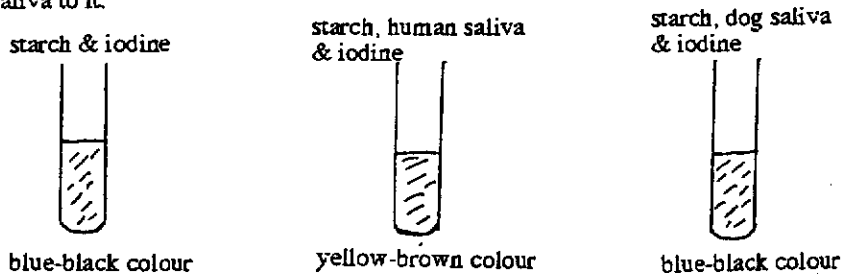
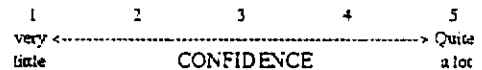


Fig.1: Results to saliva experiment.

Question 15 (P)

The original starch mixture became clear because the ...

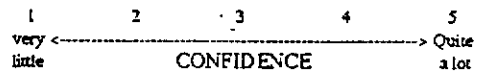
- saliva changed the starch into another substance.
- saliva was acidic and it dissolved the starch.
- iodine was neutralised by the saliva.
- iodine changed the starch into another substance.



Question 16 (P)

Results from this experiment suggest that dog saliva ...

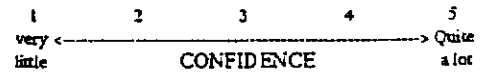
- is very similar to human saliva.
- lacks a starch-digesting enzyme.
- contains an iodine digesting enzyme.
- contains large quantities of starch.



Question 18 (P)

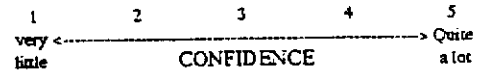
The decomposer organisms are ...

- A. bacteria and fungi.
- B. native cats and bacteria.
- C. grasshoppers and mice.
- D. fungi and grasses.

Question 19 (P)

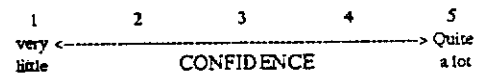
If hawks started to eat mice instead of grasshoppers, the number of snakes would probably ...

- A. decrease temporarily.
- B. increase temporarily.
- C. decrease.
- D. increase.

Question 20 (P)

The most likely reason for the change in the number of snakes when hawks start eating mice is due to ...

- A. the removal of a predator from a feeding relationship.
- B. a decrease in the amount of food available to snakes.
- C. an increase in the amount of food available for snakes.
- D. the introduction to a new predator into the ecosystem.



ATTITUDES TOWARD TESTS SURVEY

A number of statements which people have used to describe themselves are given below. Read each of the statements and circle the number which describes how you feel about the test questions you have just completed.

1 = Almost Never, 2 = Sometimes, 3 = Often, 4 = Almost Always

There are no right or wrong answers. Do not spend too much time on one statement but give the answer which seems to describe how you feel. Please answer every statement.

1. Thinking about my marks in a subject interferes with my work on tests... 1 2 3 4
2. I do as well, or better than others on tests 1 2 3 4
3. My mind freezes up on important exams 1 2 3 4
4. I don't have to rush through questions 1 2 3 4
5. During exams I find myself thinking about whether I'll ever
get through school 1 2 3 4
6. I am in control of my thoughts during tests..... 1 2 3 4
7. The harder I work at taking a test, the more confused I get 1 2 3 4
8. I am confident in my abilities when I work on tests 1 2 3 4
9. Thoughts of doing poorly interfere with my concentration on tests 1 2 3 4
10. I find it easy to understand what the questions ask..... 1 2 3 4
11. I seem to defeat myself while working on important tests 1 2 3 4
12. I am able to keep a clear mind during tests 1 2 3 4
13. During tests I find myself thinking about the consequences of failing 1 2 3 4
14. I do all right on tests..... 1 2 3 4
15. During examinations I get so nervous that I forget facts I really know..... 1 2 3 4

Constructed Response Questions (20 marks)
 All questions are worth two marks

Write your answers in the space that is provided.

Question 1 (C)

Describe what is meant by

- the term 'chemical digestion', and
- name three places in the body where chemical digestion takes place.

1 2 3 4 5
 very <-----> Quite
 little CONFIDENCE a lot

Question 2 (C)

Name

- three places in the body where digestive enzymes are produced.
- the endproducts of fat digestion.

1 2 3 4 5
 very <-----> Quite
 little CONFIDENCE a lot

Question 3 (C)

Describe the way in which

- oxygen is transported around the circulatory system.
- carbon dioxide is transported around the circulatory system.

1 2 3 4 5
 very <-----> Quite
 little CONFIDENCE a lot

Question 4 (C)

Describe the meaning of these words as they apply to the transport of materials in plants:

- transpiration.
- translocation.

1 2 3 4 5
very <-----> Quite
little CONFIDENCE a lot

Question 5 (C)

- Explain the meaning of the word 'parasite' with reference to an example.

1 2 3 4 5
very <-----> Quite
little CONFIDENCE a lot

Question 6 (P)

Identify and describe two differences between mechanical digestion and chemical digestion.

1 2 3 4 5
very <-----> Quite
little CONFIDENCE a lot

Question 7 (P)

Study the following information.

An experiment was set up as follows. Ground up meat, starch and some gastric juices from a person's stomach were placed into a 'sausage' made out of dialysis tubing (a semi-permeable membrane). The sausage was placed into a beaker of water and the temperature was kept at 37°C.

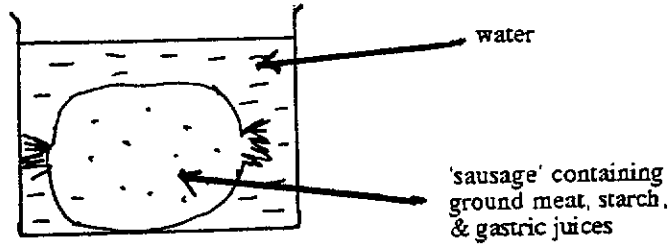


Fig. 3: Experimental set up for question 7.

Use the information you are given to

- predict the substance(s) that you would find in the water outside the sausage.
- suggest a reason why the temperature was kept at 37°C.

1 2 3 4 5
 very <-----> Quite
 little CONFIDENCE a lot

Question 8 (P)

Study the following information

Leaves from a variegated plant, that is a plant which has leaves that are green and white coloured (see diagram) were bleached and tested with iodine solution. The parts of the leaves that were originally green turned a blue-black colour. It was observed that the white parts of the leaves had turned a yellow-brown colour.

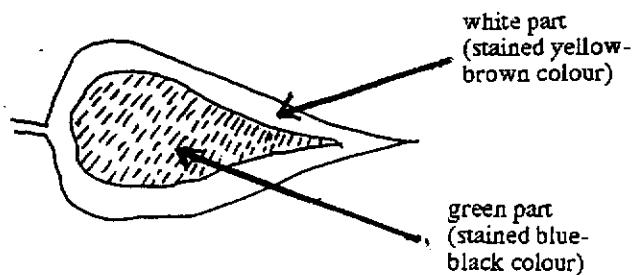


Fig. 4: Leaf from a variegated plant treated with iodine.

Explain the results that were observed for:

- the green part of the leaf.
- the yellow part of the leaf.

1 2 3 4 5
 very little CONFIDENCE Quite a lot

Question 9 (P)

Refer to the following diagram

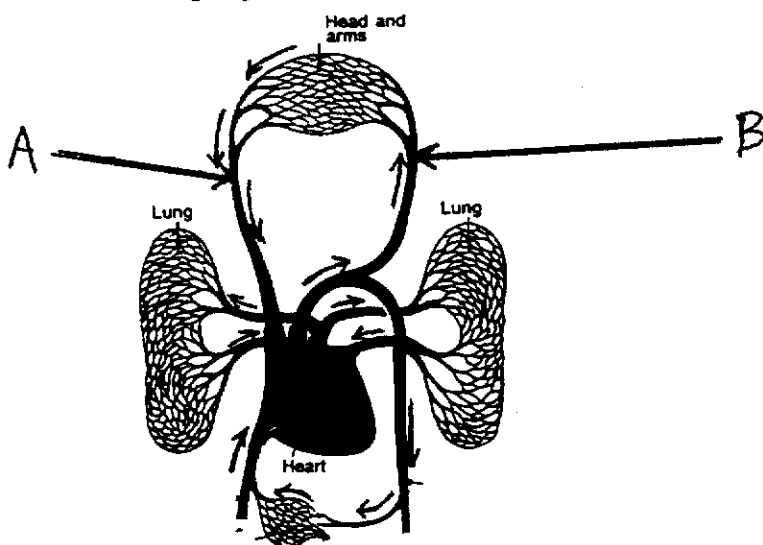


Fig. 5: The circulatory system.

- Compare the oxygen content of the blood and the level of nutrients in the blood at points A and B.

1 2 3 4 5
 very little CONFIDENCE Quite a lot

Question 10 (P)

Study this diagram of a cactus plant that lives in hot, dry and harsh desert conditions.

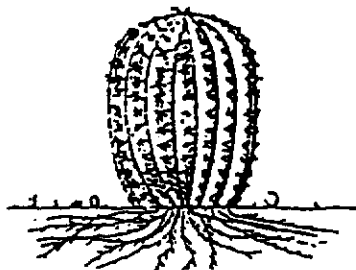


Fig. 6: Cactus plant

Suggest a reason for each of the following:

- the cactus has a waxy outer coat.
- the cactus has a shallow spreading root system.

1 2 3 4 5
very ←-----> Quite
little CONFIDENCE a lot

MARKING SCHEDULE FOR *ORGANISMS AND FOOD* TEST BOOKLETS

2A & 2B

Multiple-Choice Answers (20 marks)

1.	B	11.	A
2.	B	12.	D
3.	D	13.	B
4.	C	14.	B
5.	C	15.	A
6.	A	16.	B
7.	A	17.	C
8.	B	18.	A
9.	D	19.	C
10.	A	20.	B

Constructed-Response Answers (20 marks)

Question 1 (C)

- conversion of food substances into simple molecules (1, breakdown of food 0.5 without mention of enzymes)
- mouth, stomach, duodenum (1, 0.5 for any two)

Question 2 (C)

- salivary glands (not mouth), stomach wall, pancreas (1, 0.5 for any two)
- fatty acids (0.5) and glycerol (0.5)

Question 3 (C)

- attached to red blood cell (1)
- dissolved in plasma (1, red blood cell 0.5)

Question 4 (C)

- loss of water through a plant's leaves (1, 0.5 for movement of water in a plant)
- movement of food from a plant's leaves to other areas (1)

Question 5 (C)

- organism that feeds on another species causing it to suffer (1)
- any appropriate example (1)

Question 6 (P)

- mechanical digestion is a physical change whereas chemical digestion is a chemical change; chemical digestion involves enzymes whereas mechanical digestion does not; mechanical digestion involves muscular activity whereas chemical digestion does not (1 for each, max of 2)

Question 7 (P)

- amino acids (1)
- to enable the enzymes operate at their greatest efficiency (1, 0.5 for body temperature)

Question 8 (P)

- starch is stored in the green part of the leaf (1)
- chlorophyll is required for photosynthesis (1)

Question 9 (P)

- oxygen levels low at A and high at B (1)
- nutrient levels low at A and high at B (1)

Question 10 (P)

- to decrease the rate of water loss (1)
- to absorb as much water as possible from the soil when it rains (1)

APPENDIX E
DEVELOPMENT OF THE *SURVEY ABOUT SCIENCE TESTS*

PILOT VERSION OF THE *SURVEY ABOUT SCIENCE TESTS*

Name _____ Class _____
Age: _____ Years _____ Months Sex: Male / Female

Instructions

1. Please answer the following questions as completely as you can. Write "not applicable" to those questions which do not seem to apply to you.
2. There are no right answers or wrong answers to the questions on this survey. The purpose of this survey is to find out your thoughts and feelings about science tests so that tests can be improved in the future.
3. Your identity will be kept confidential. You are only asked to provide your name in case there are questions about some of your responses.

Survey Questions

1. Write out some of the irrelevant thoughts that pass through your mind during science tests, classifying them as negative or positive thoughts.

Negative thoughts	Positive thoughts

2. Do you think that negative thoughts effect your work during tests?

Please explain how negative thoughts effect your work
OR, why you think they don't effect your work.

3. Do you think that positive thoughts effect your work during tests?

Please explain how positive thoughts effect your work
OR, why you think they don't effect your work.

4. Does it help if you are feeling confident during a test?

Please explain your answer.

5. What types of things happen when you lose your confidence during a test?

6. What do you do to help yourself during a test when you start feeling flustered or worried?

7. Which type of questions do you generally feel more confident answering; multiple-choice or written answer questions.

8. Write out some of the things you find easy about multiple-choice and written answer questions.

Multiple-choice questions	Written answer questions

9. Write out some of the things you find difficult about multiple-choice and written answer questions.

Multiple-choice questions	Written answer questions

10. Write out some of the things you worry about when answering multiple-choice and written answer questions.

Multiple-choice questions	Written answer questions

11. Does it make any difference to your marks whether the test questions are multiple-choice or written answer questions?

Please explain your answer.

12. What would have to change before you could achieve a higher mark on your next science test?

Thank you for your assistance. Are there any more comments you would like to make about science tests? If so, write them in this space.

REVISED VERSION OF THE *SURVEY ABOUT SCIENCE TESTS*

Name _____ Class _____
Age: _____ Years _____ Months Sex: Male / Female

Instructions

1. Please answer the following questions as completely as you can
Write "not applicable" to those questions which do not seem to apply to you.
2. There are no right answers or wrong answers to the questions on this survey.
The purpose of this survey is to find out your thoughts and feelings about science tests so that they can be improved in the future.
3. Your identity will be kept confidential. You are only asked to provide your name in case there are questions about some of your responses.

Survey Questions

1. Write out some of the irrelevant types of thoughts that pass through your mind during science tests, classifying them as negative or positive thoughts.

Negative thoughts	Positive thoughts
1.	1.
2.	2.
3.	3.
4.	4.

2. Do you think that negative thoughts effect your work during tests?

 Please explain how negative thoughts effect your work
OR, why you think they don't effect your work.

3. Do you think that positive thoughts effect your work during tests?

 Please explain how positive thoughts effect your work
OR, why you think they don't effect your work.

4. What types of things happen to you when you lose your confidence during a test?

5. What do you do to help yourself when you start feeling worried or flustered during a test?

6. Which type of questions do you generally feel more confident answering; multiple-choice or written answer questions.

7. Write out some of the things you find easy about multiple-choice and written answer questions.

Multiple-choice questions	Written answer questions
1.	1.
2.	2.
3.	3.

8. Write out some of the things you find difficult about multiple-choice and written answer questions.

Multiple-choice questions	Written answer questions
1.	1.
2.	2.
3.	3.

9. Does it make any difference to your marks whether the test questions are multiple-choice or written answer questions?

Please explain your answer.

-
-
-
-
10. What would have to change before you could achieve a higher mark on your next science test?

Thank you for your assistance. Are there any more comments you would like to make about science tests? If so, write them in this space.

APPENDIX F

INTELLIGENCE QUOTIENT SCORES

Table F.1
Comparison of Boys' and Girls' Intelligence Quotient Scores

Students	<u>n</u>	<u>M</u>	<u>SD</u>	<u>d</u>
All students				
Girls ^a	72	105.31	11.81	
Boys ^b	57	104.35	11.94	-0.08
Research Groups 1& 2				
Girls	36	106.06	12.59	
Boys	38	103.47	10.01	-0.23
Research Groups 3 & 4				
Girls	36	104.56	11.09	
Boys	19	106.11	15.00	+0.13

Note. Intelligence quotient scores were measured on the *ACER Intermediate Test G* (Australian Council for Educational Research, 1982), d = effect size. ^aData were not available for 9 girls. ^bData were not available for 16 boys.

APPENDIX G

INTERVIEWING GUIDE

RESEARCH TECHNIQUES AND THOUGHTS DURING TESTS

- What was the student's reaction toward completing the confidence scale?
- Did the student think that answering the confidence scale had any effect on them during the test?
- Did the student complete the confidence scale before or after answering the question?
- What was the student's reaction toward completing the *Attitudes Toward Tests Survey* in the middle of the test?
- Did the student think that answering the *Attitudes Toward Tests Survey* had any effect on them during the test?
- What was the student's reaction to having the order of the multiple-choice and constructed-response questions reversed in the tests?
- Which type of questions, multiple-choice or constructed-response questions, does the student prefer to start their test with? What reasons do they give?
- Is the student the type of person who is normally fairly aware of their thoughts and feelings during tests?
- Did the student think that the changes to the tests, that is, asking them to think about their thoughts during tests, cause them to become more aware of their thoughts and feelings during the tests? If so, did the student think that this increased awareness had any influence on them during the test?
- Does the student normally have time to spare at the end of science tests?
- Was the student concerned about the extra time that was required in the tests to complete confidence scale and the survey questions?

QUESTIONS ABOUT SELF-EFFICACY

- What types of things makes the student feel confident before they go into a science test?
- What type of things make the student feel as if they lack confidence before they go into a science test?
- Does the student think that the confidence they have in tests is related to the teacher they have for that subject? If so, how?
- Does the student think that the confidence they have in tests is related to the sex of their teacher? If so, how?
- What types of things causes the student to feel confident during tests?
- What types of things causes the student to feel they lack confidence during tests?
- Does the student think that having confidence is good or bad during tests? What reasons do they give for their response?
- Does the student think that overconfidence is good or bad during tests? What reasons do they give for their response?
- Does the student possess any strategies to make themselves feel more confident during science tests?
- Is the student more confident answering multiple-choice or constructed-response questions? What reasons do they give for their response?
- What suggestions does the student make about changing science tests so that tests are better for them?
- What suggestions does the student make about changing the assessment program in science so that it is better for them?

QUESTIONS ABOUT WORRISOME COGNITIONS

- What types of things does the student worry about during tests?
- How do worrisome thoughts influence the student during tests?
- What causes worrisome thoughts to stop during tests?
- Does the student have any ways of coping with worrisome thoughts during tests? If so, what are these ways?

APPENDIX H

FACTOR ANALYSIS OF THE WORRY DATA

Table H.1
Principal Components Analysis of Worry Items (All Students)

Item Number	Communality	Eigenvalue	Percent of Variance	Cumulative Percent of variance
1	1.00	3.60	45.0	45.0
3	1.00	0.96	12.0	57.1
5	1.00	0.77	9.6	66.7
7	1.00	0.75	9.4	76.1
9	1.00	0.70	8.7	84.8
11	1.00	0.55	6.9	91.7
13	1.00	0.37	4.6	96.3
15	1.00	0.30	3.7	100.0

Note. 141 cases were analysed.

Table H.2
Loading of Worry Items (All Students)

Item Number	Loading on Factor 1
1	0.69
3	0.81
5	0.51
7	0.52
9	0.80
11	0.57
13	0.72
15	0.69

Note. 141 cases were analysed.

Table H.3
Principal Components Analysis of Worry Items for Girls and Boys

Item Number	Communality	Eigenvalue	Percent of Variance	Cumulative Percent of variance
Girls^a				
1	1.00	3.93	49.0	49.0
3	1.00	1.00	12.5	61.6
5	1.00	0.78	9.8	71.3
7	1.00	0.64	8.0	79.3
9	1.00	0.59	7.3	86.6
11	1.00	0.44	5.5	92.2
13	1.00	0.37	4.6	96.8
15	1.00	0.26	3.2	100.0
Boys^b				
1	1.00	3.14	39.2	39.2
3	1.00	1.07	13.3	52.6
5	1.00	0.93	11.6	64.2
7	1.00	0.79	9.8	74.0
9	1.00	0.73	9.2	83.2
11	1.00	0.68	8.5	91.7
13	1.00	0.35	4.4	96.1
15	1.00	0.31	3.9	100.0

Note. ^aScores from 80 girls were analysed. ^bScores from 61 boys were analysed.

Table H.4
Loadings of Worry Items for Boys and Girls

Item Number	Loading on Factor 1	Loading on Factor 2
Girls ^a		
1	0.63	-0.52
3	0.82	0.02
5	0.57	0.68
7	0.66	-0.38
9	0.83	-0.08
11	0.63	0.33
13	0.79	0.11
15	0.63	-0.07
Boys ^b		
1	0.77	0.02
3	0.77	0.10
5	0.38	-0.70
7	0.25	0.63
9	0.73	-0.14
11	0.52	0.19
13	0.65	-0.24
15	0.74	0.24

Note. ^aScores from 80 girls were analysed. ^bScores from 61 boys were analysed.

Table H.5
Loadings of the Worry Items For Girls and Boys After
Rotation of the Axes

Item Number	Loading on Factor 1	Loading on Factor 2
Girls ^a		
1	0.82	0.02
3	0.61	0.54
5	-0.01	0.89
7	0.75	0.15
9	0.68	0.48
11	0.26	0.66
13	0.53	0.59
15	0.53	0.36
Boys ^b		
1	0.76	0.06
3	0.77	-0.02
5	0.31	0.74
7	0.31	-0.60
9	0.71	0.21
11	0.54	-0.14
13	0.62	0.30
15	0.76	-0.16

Note. ^aScores from 80 girls were analysed. ^bScores from 61 boys were analysed.

APPENDIX I
RESPONSE CHARACTERISTICS FOR THE *SURVEY*
ABOUT SCIENCE TESTS

Table I.1
Response Characteristics for Questions About Negative Thoughts

Survey Question	Item Type	Response Rate (%)		Response Frequency	
		Boys	Girls	Boys	Girls
1a	General	98	95	2.4	2.7
8a	Multiple-choice	46	65	1.5	1.6
8b	Constructed-response	68	80	2.0	2.2

Table I.2
Response Characteristics for Questions About Positive Thoughts

Survey Question	Item Type	Response Rate (%)		Response Frequency	
		Boys	Girls	Boys	Girls
1b	General	86	87	1.9	1.9
7a	Multiple-choice	92	95	1.8	2.0
7b	Constructed-response	48	61	1.3	1.5

APPENDIX J
CHARACTERISTICS OF THE EQUIVALENT TESTS

Table J.1
Item-Specific Self-Efficacies for Equivalent Multiple-Choice and
 Constructed-Response Items

Grouping		Mean Self-Efficacy			
		<u>MC</u> items		<u>CR</u> items	
		Set 1	Set 2	Set 1	Set 2
<i>Earth Science test</i>					
1	Boys	3.16	-	-	3.23
2	Girls	3.10	-	-	2.74
3	Boys	-	3.38	2.98	-
4	Girls	-	2.84	2.86	-
<i>Organisms and Food test</i>					
1	Boys	-	3.21	3.09	-
2	Girls	-	2.93	2.50	-
3	Boys	3.09	-	-	2.75
4	Girls	2.58	-	-	2.49

Note. Dashes indicate cells where no data were collected. Self-efficacy judgements were made on 5-point scales (1 = very little confidence, 5 = quite a lot of confidence). MC = Multiple-Choice; CR = Constructed-Response.

Table J.2
Worry Scores for the Equivalent Multiple-Choice and Constructed-Response Tests

Groups		Extent of Worrisome Cognitions			
		<u>MC</u> items		<u>CR</u> items	
		Set 1	Set 2	Set 1	Set 2
<i>Earth Science test</i>					
1	Boys	16.40	-	-	15.78
2	Girls	17.41	-	-	19.22
3	Boys	-	14.91	14.93	-
4	Girls	-	18.11	18.36	-
<i>Organisms and Food test</i>					
1	Boys	-	16.69	17.92	-
2	Girls	-	17.21	17.62	-
3	Boys	14.03	-	-	15.55
4	Girls	15.70	-	-	16.69

Note. Worry data were collected only for the questions on the first part of each achievement test; dashes indicate cells where no data were collected. Minimum score = 8; Maximum score = 32. High scores correspond to high levels of worry. Positive effect sizes indicate higher worry scores for boys than girls. MC = Multiple-Choice; CR = Constructed-Response; d = effect size.

Table J.3
Achievement Scores for Equivalent Multiple-Choice and Constructed-Response Tests

Grouping		Mean Achievement			
		<u>MC</u> items		<u>CR</u> items	
		Set 1	Set 2	Set 1	Set 2
<i>Earth Science test</i>					
1	Boys	12.10	-	-	8.32
2	Girls	12.25		-	8.00
3	Boys	-	11.40	9.74	-
4	Girls	-	10.95	10.79	-
<i>Organisms and Food test</i>					
1	Boys	-	12.41	9.73	-
2	Girls	-	12.59	9.03	-
3	Boys	11.50	-	-	9.02
4	Girls	11.86	-	-	8.97

Note. Dashes indicate cells where no data were collected. Maximum score = 20.
MC = Multiple-Choice; CR = Constructed-Response.