EUCLIDEAN GEOMETRY: COGNITIVE GENDER DIFFERENCES

BY

LEFINA SUSANNA CRONJÉ

A thesis submitted in fulfilment of the
requirements for the degree of

DOCTOR OF PHILOSOPHY

in the

SCHOOL OF SCIENCE EDUCATION (MATHEMATICS)
UNIVERSITY OF THE WITWATERSRAND

SUPERVISOR: PROF. P.E. LARIDON CO-SUPERVISOR: PROF. J.J. BARNARD

PRETORIA

SEPTEMBER 1995

ABSTRACT

Cognitive gender differences in mathematics, which were noted in past research have latterly been found to be declining, but are still observed at secondary school level in the field of spatial skills and geometry. However, due to divergent geometry curricula and the multidimensional nature of spatial skills, results from gender studies have led to broad conclusions being drawn from investigations which were affected by random elements. Research results have been influenced by imbalanced representation of the two gender groups in samples and also by differential gender exposure to previous mathematical experience. Both these limitations in research have produced outcomes reflecting negatively on females. Internationally divergent geometry curricula are a result of the traditional synthetic approach to the teaching of Euclidean geometry falling into disfavour in western countries. Thus, Euclidean geometry has been largely overlooked in current educational research and the cognitive processes such as deductive reasoning, proof writing and hypothesis testing, have received scant attention within geometry teaching over the past decade.

The standardized system in most South African secondary schools coupled with a uniform core curriculum for mathematics lends itself to an advantageous environment for empirical studies in mathematics performance. The present gender investigation involved almost an equal number of males and females from 5 selected secondary schools near Johannesburg in South Africa. Performance in Euclidean geometry was evaluated for possible gender differences by means of multiple choice and open ended questions. In keeping with most existing research, mean differences were compared. No statistically significant gender difference in mean performance on Euclidean geometry was found.

Further analysis at item level was undertaken for the multiple choice component of the test instrument by means of the one dimensional three parameter logistic model in Item Response Theory. This model is considered to be very sensitive in identifying differences at different ability levels - a shortcoming in existing research. Four out of a total of twenty Euclidean geometry items, were identified as producing differential gender performance. Similar gender tendencies were observed for all four items. The performance was in favour of most males for each of these items. Females of lower ability (as measured by the item) did not perform as well as males of similar ability level. In all four cases females of higher ability surpassed males of a corresponding ability level. Perceptual difficulties in females were more pronounced than in males for some of the items.

DECLARATION

I declare that this thesis is my own unaided work. It is submitted for the degree of DOCTOR OF PHILOSOPHY in the University of the Witwatersrand, Johannesburg. It has not been submitted before for any other degree or examination in any other university.

L. S. Ciunje
.....

25 ..... day of September 1995.

This thesis is dedicated to my husband, *FREDERIK*.

ACKNOWLEDGEMENTS

I wish to express my sincere gratitude to the following persons and institution:

* *Prof. P.E. Laridon*, my supervisor, for his expert guidance and encouragement at all times during the completion of this thesis;

* *Prof. J.J. Barnard*, my co-supervisor, for his enthusiastic support and for his able supervision of the statistical analysis;

* *Karin and Lindsay*, for proof-reading the text;

* *Jaco and Stefanie*, for assistance with computer programs;

* My *mother*, for moral support;

* The *Centre for Science Development* (Human Sciences Research Council, South Africa) for financial assistance and for the use of their computer facilities. Opinions expressed and conclusions reached in this thesis should, however, not be regarded as those of the HSRC.

Above all, thank you to my *Heavenly Father*.

CONTENTS

ABSTRACT		ii
DECLARATION		iii
DEDICATION		iv
ACKNOWLEDGEMENTS		v
CONTENTS		vi
LIST OF FIGURES		xi
LIST OF TABLES		xii

CHAPTER 1: INTRODUCTION

1.1	EUCLIDEAN GEOMETRY	1
1.2	GENDER DIFFERENCES - RESEARCH	4
	1.2.1 Mathematics	4
	1.2.2 Spatial ability and geometry	5
	1.2.3 Euclidean geometry	7
	1.2.4 South Africa	8
	1.2.5 Summary	8
1.3	THE AIM OF THE STUDY	10
1.4	STATEMENT OF THE RESEARCH PROBLEM	10
1.5	TERMINOLOGY	10
	1.5.1 Sex differences / gender differences	11
	1.5.2 Geometry and Euclidean geometry	11
	1.5.3 Spatial ability, spatial visualization, and spatial orientation	11
	1.5.4 Cognitive	12
	1.5.5 Item	12
1.6	RESEARCH METHOD	13
	1.6.1 Measuring instrument	13
	1.6.2 Sample	13
	1.6.3 Analysis	13
1.7	A SYNOPSIS OF THE CONTENTS OF THIS THESIS	14

CHAPTER 2: EUCLIDEAN GEOMETRY

2.1	INTRODUCTION	16
2.2	WHAT IS EUCLIDEAN GEOMETRY ?	16
	2.2.1 Terminology	17
	2.2.2 South Africa	18
2.3	HISTORY OF EUCLIDEAN GEOMETRY	18
	2.3.1 Euclid's Elements	18
	2.3.2 Crisis - non-Euclidean geometries	19
	2.3.3 Number theory	21
	2.3.4 Modern view	22
	2.3.5 Motivation	24
2.4	THE STRUCTURE OF EUCLIDEAN GEOMETRY	24
	2.4.1 Axioms	24
	2.4.2 Definitions	25
	2.4.3 Theorems	25
	2.4.4 Inductive and deductive reasoning	25
	2.4.5 Motivation	26
2.5	SPATIAL AND VISUAL ELEMENTS	26
	2.5.1 Spatial visualization	28
	2.5.2 Spatial orientation	28
2.6	THE TEACHING OF EUCLIDEAN GEOMETRY	30
	2.6.1 Dimensions in the teaching of Euclidean geometry	31
	2.6.2 Historical overview of the teaching of Euclidean geometry	33
	2.6.3 Current situation	34
	2.6.4 Teaching problems	42
	2.6.5 Research on the teaching of Euclidean geometry	53
	2.6.6 Merits of teaching Euclidean geometry	58
2.7	SUMMARY	59

CHAPTER 3: GENDER DIFFERENCES

3.1 INTRODUCTION 61

3.2 MATHEMATICS:
GENDER DIFFERENCES IN YOUNG CHILDREN 64

3.2.1 Cognitive domain 64

3.2.2 Affective domain 65

3.3 MATHEMATICS:
GENDER DIFFERENCES FROM EARLY ADOLESCENCE 65

3.3.1 Cognitive domain 65

3.3.2 Affective domain 86

3.3.3 Educational factors 88

3.3.4 Biological factors 89

3.4 PROBLEMS AND LIMITATIONS IN RECORDED RESEARCH 90

3.5 SUMMARY 96

CHAPTER 4: MEASUREMENT THEORIES

4.1 INTRODUCTION 98

4.2 CLASSICAL TEST THEORY 98

4.2.1 History of Classical Test Theory 98

4.2.2 Assumptions of Classical Test Theory 99

4.2.3 Difficulty and discrimination value 100

4.2.4 Validity 101

4.2.5 Reliability 102

4.2.6 Some of the most important shortcomings of CTT 104

4.2.7 Summary 107

4.3 ITEM RESPONSE THEORY 107

4.3.1 Introduction 107

4.3.2 History of IRT 109

4.3.3 Assumptions of IRT 109

4.3.4 Definition of IRT 111

4.3.5 Item Response Function 112

4.3.6 Examples of IRFs 118
4.3.7 Estimation of parameters 119
4.3.8 Standard error of estimated ability 121
4.3.9 Test Characteristic Curve 121
4.4 DIFFERENCES, RESEMBLANCES AND ADVANTAGES - CTT AND IRT 122
4.5 OPEN QUESTIONS AND MULTIPLE CHOICE QUESTIONS 125
4.6 DIFFERENTIAL PERFORMANCE - BIAS 127
4.6.1 Differential item functioning 129
4.6.2 The detection of differential item functioning 130
4.6.3 DIF - Item Response Theory and Classical Test Theory 131
4.6.4 Distractor response analysis 133
4.6.5 Differential gender item functioning 134
4.7 SUMMARY 135

CHAPTER 5: EMPIRICAL INVESTIGATION

5.1 INTRODUCTION 136
5.2 SAMPLE 137
5.3 TEST INSTRUMENT 138
5.4 PROCEDURE 139
5.5 RESULTS OF EMPIRICAL RESEARCH ON TEST LEVEL 141
5.5.1 Mean scores 141
5.5.2 Discussion 145
5.6 RESULTS OF EMPIRICAL RESEARCH ON ITEM LEVEL 148
5.6.1 Item Response Theory: Item parameter estimates 148
5.6.2 Item Response Theory: Numerical values of parameters 151
5.6.3 Item Response Theory: Differential item performance 152
5.7 DIAGNOSTIC ANALYSIS OF OPEN ENDED QUESTIONS 169
5.7.1 Analysis of subdivisions of questions 169
5.7.2 Diagnostic evaluation of open ended questions 171
5.8 SUMMARY 174

CHAPTER 6: MAJOR OUTCOMES, CONCLUSIONS, LIMITATIONS AND RECOMMENDATIONS

6.1	INTRODUCTION	176
	6.1.1 Major aim of study	176
	6.1.2 Euclidean geometry	177
	6.1.3 Gender differences	178
	6.1.4 Favourable conditions in South Africa for the research done	178
6.2	MAJOR OUTCOMES AND CONCLUSIONS	179
	6.2.1 Comparison at mean level	180
	6.2.2 Comparison at item level	181
	6.2.3 Diagnostic analysis of open ended questions	184
	6.2.4 Conclusions	184
6.3	LIMITATIONS OF THE STUDY	185
6.4	RECOMMENDATIONS	185
BIBLIOGRAPHY		187
APPENDICES A - I		218

LIST OF FIGURES

2.1	Geometry diagram	30
2.2	Geometry diagram	30
4.1	Item Response Function (three parameter logistic model)	113
4.2	An example of the IRFs of four items	118
5.1	Item Response Function - Differential item performance for males and females	154
5.2	Item Response Functions for males and females - item 8	158
5.3	Item Response Functions for males and females - item 4	161
5.4	Item Response Functions for males and females - item 12	165
5.5	Item Response Functions for males and females - item 11	168

LIST OF TABLES

5.1: Multiple choice test: Mean scores 142
5.2: Open ended test: Mean scores 143
5.3: Correlations: Multiple choice / open ended questions / examination marks 144
5.4: Item parameter estimates (from Appendix B) 150
5.5: Item Response Theory: Means of parameters 152
5.6: Values of item parameters for item 8 156
5.7: Frequencies of key and distractors for item 8 158
5.8: Values of item parameters for item 4 160
5.9: Frequencies of key and distractors for item 4 161
5.10: Values of item parameters for item 12 163
5.11: Frequencies of key and distractors for item 12 165
5.12: Values of item parameters for item 11 166
5.13: Frequencies of key and distractors for item 11 168
5.14: Mean scores on open ended questions 170

CHAPTER 1

INTRODUCTION

"Let no one ignorant of Geometry enter my door"

PLATO

1.1 EUCLIDEAN GEOMETRY

The only branch of mathematics that had been formulated and reduced to a deductive system, in Plato's time, was Euclidean geometry. However, he was not referring to geometry only when he put the inscription above the portal of his academy. He argued that a man whose mind had not undergone a rigorous training in systematic thinking and in the art of drawing legitimate conclusions from premises, was unable to discuss the deeper problems of life - albeit moral, political or social issues.

Modern educators still pursue similar goals. Whether training engineers, doctors, teachers, accountants or secretaries, a vital part of their education lies in their acquiring that ability which enables them to correlate facts - whether concrete or abstract - and reach a logical conclusion therefrom.

It is hard to imagine a subject better suited to this purpose than mathematics in general, and geometry in particular (Baart, 1992).

> "Without doubt, geometry is important. It offers us a way to interpret and reflect on our physical environment. It can serve as a tool for the study of other topics in mathematics and science. More importantly, however, spatial thinking, which obviously undergirds geometry, has been suggested by famous mathematicians such as Hadamard and Einstein to be essential to creative

thought in all high level mathematics. Given their importance, therefore, it is essential that geometry and spatial reasoning receive greater attention in instruction and in research." (Clements and Battista, 1992: 457)

Euclidean geometry, as documented in the famous 'Elements' of Euclid around 300 BC, dominated the mathematical world for almost 2000 years. During the time of Newton (1642-1727), the geometry of Euclid was downgraded from 'made by God' to 'made by man', (Hansen, 1993), but it was still thought by many to be a perfect mirror of cosmic geometry.

It was only in the eighteenth century that the Hyperbolic and Spherical geometries, which will be discussed in chapter 2, started developing. Mathematicians had questioned Euclid's fifth postulate, the parallel postulate, for hundreds of years. They tried to show that the fifth postulate was independent of the other postulates, or that it could be proved from them. Early in the nineteenth century, Euclid's parallel postulate was replaced by two alternatives, leading to the non-Euclidean geometries, but the view that Euclidean geometry was the only possible geometry was so firmly rooted that it took a long time for the other geometries to gain recognition. Whilst Euclidean geometry was still being taught in schools to develop deductive reasoning plus visual and spatial skills, the invention of the other geometries resulted in attempts to axiomatize all branches of mathematics. This development gave rise to a rigidity of approach to the teaching of mathematics and geometry, and the inevitable decline in interest in Euclidean geometry.

During the 1960s there was a flood of criticism directed at the traditional Euclidean geometry courses, mainly because of problems with the teaching thereof. Dieudonné's now famous pronouncement, 'Euclid must go', was made in 1959 at the Royaumont Conference with far-reaching consequences. In most western countries, Euclidean geometry was replaced by other geometries, such as transformation geometry, analytical geometry, vectors, etc. At the International Conferences on Mathematics Education, ICME 4 (1984) and ICME 5 (1988), Bloom (1984; 1988) mentioned the sharp decline of interest in the teaching of Euclidean geometry.

In the meantime there was increasing concern about the over-emphasis of the non-spatial, and about the lack of an organized geometry curriculum. The decision to replace Euclid was questioned by leading mathematicians, such as Rene Thom and Michael Atiyah, at the international conferences of 1972 and 1976 (Kapadia, 1985). Also at ICME 5 there was evidence that some countries regretted their decision to 'let Euclid go'.

> "But Euclidean geometry provides a good model of our space and it is important in the further study of mathematics. Students should certainly have some exposure to Euclidean geometry at secondary school level." (Egsgard and Fletcher, 1984: 87).

Kapadia (1985) wrote in his thesis on the pedagogical importance of formal geometry that, from a mathematical and historical point of view, Euclidean geometry does justify a place in the curriculum.

> "It is a landmark in man's development with an influence far beyond mathematics. It is the first and the foremost example of a deductive system. It is an essential part of the cultural heritage which should be transmitted in school mathematics." (Kapadia, 1985: 54).

Also in England, Love (1989) objected to the fragmentation of the geometry curriculum into disjointed bits and the lack of an organized geometry system. In their preview of mathematics in the 1990s, Howson and Kahane (1986) said that formal geometry had vanished from many school systems. But few, if any, replacements give indications that they are more successful - either in being accessible to students, motivating them, or providing better preparation for future mathematical work - than the Euclid they replaced.

Euclidean geometry undoubtedly develops two essential aspects of mathematics, namely the spatial and the visual, and both seem to be of importance in the study of mathematics in general (Brown and Presmeg, 1993; Presmeg, 1993; Clements and Battista 1992; Battista, 1990; Tartre, 1990; Bishop, 1989; Ben-Chaim et al., 1988; Johnson and Meade, 1987; Fennema and Tartre, 1985; McGee, 1979; Sherman and Fennema, 1977). These spatial

and visual factors are responsible for much of the controversy with regard to differential gender achievement in mathematics.

1.2 GENDER DIFFERENCES - RESEARCH

1.2.1 Mathematics

Gender differences in mathematics learning and performance continue to attract much research attention, and Fennema, one of the leaders in this field (Fennema and Hart 1994; Fennema, 1981), draws attention to the social implications of publications on gender differences. Generalizations of research results encourage misconceptions about mathematics performance and often these misconceptions, to the disfavour of women, are given wide press coverage (Hyde et al., 1990; Tartre, 1990; Kimball, 1989; Walkerdine, 1989; Leder, 1986).

Extensive research has been done over the past two decades in countries such as the United States, England, Canada and Australia. A variety of factors have been investigated and reviewed, some of which are indicated below:

Cognitive	Feingold, 1994; Friedman, 1994; Frost et al., 1994; Feingold, 1993; Swanepoel, 1993; Clements and Battista, 1992; Tartre, 1990; Battista, 1990; Baenninger and Newcombe, 1989; Bishop, 1989; Usiskin, 1987; Fennema and Tartre, 1985;
Affective	Frost et al., 1994; Ethington, 1992; Leder, 1992; McLeod, 1992; Hembree, 1990; Visser, 1987; 1985; Fennema and Sherman, 1977;
Social and educational	Bohlin, 1994; Hanna, 1994; Ethington, 1992; Baya'a, 1990; Fennema, 1990; Kimball, 1989; Robitaille and Garden, 1989; Visser, 1987; 1985;

Biological	Hensel, 1989; Rudisill and Morrison, 1989; Benbow and Stanley, 1983; Halpern, 1986; McGee, 1979.

Some results have been **contradictory** (Bohlin, 1994; Friedman, 1994; Leder, 1992; Wainer and Steinberg, 1992; Leder and Fennema, 1990; Feingold, 1988) due to a number of factors which will be discussed, while some studies (Hanna, 1994; Ethington, 1992; Hyde et al., 1990; Armstrong, 1981) found no substantial gender differences for several areas of mathematics. The popular belief is that, where differences exist, they occur from adolescence onwards and favour males in achievement and participation in mathematics. The differences tend to increase at higher cognitive levels. Recent meta-analytical studies (Frost et al., 1994; Hyde et al., 1990) show a slight female superiority in computation with no gender differences in understanding of concepts and a slight male advantage in problem solving.

Referring to the Second International Mathematics Study (SIMS), Hanna et al. (1990) said that the issue of gender differences in mathematics is very complex and should be explored from many perspectives. The analysis of achievement results shows that gender differences in mathematics vary from country to country. An interesting and confounding factor is that Japanese pupils, who generally outperform their counterparts from almost all other participating countries, also have the most negative attitude towards mathematics. Their sample, based on enrolment, was also unrepresentative (22%) of females. The samples in England and the USA contained 34% and 44% females respectively. Robitaille and Garden (1989) caution specifically against conclusions drawn from unrepresentative samples in international studies. (See section 3.1 for the favourable empirical conditions for the present study.)

1.2.2 Spatial ability and geometry

The role that spatial skills play in mathematics achievement has attracted a considerable amount of research attention by those concerned with gender differences (Friedman, 1994; Tartre, 1990; Battista, 1990; Fennema and Tartre, 1985). According to Clements and Battista (1992), there is reason to believe that spatial ability is important in students'

construction and use of mathematical concepts - even non-geometric ones. The relative lack of an organized geometry curriculum in many western countries is thus a matter of concern.

The diversity of geometry curricula across countries is reflected in the results of SIMS. The performance of pupils of about 18 years old in fifteen countries, including the USA, England, Scotland, Belgium, Japan and Thailand, show large cross-cultural differences. The geometry test used by SIMS consisted of 28 multiple choice questions. On the whole, the achievement was low with a mean of 38% and a standard deviation of 22% (Hanna et al., 1990). Five countries, including England and Scotland, showed no significant gender differences in geometry performance. South Africa was not involved in SIMS but will take part in TIMSS (Third International Mathematics and Science Study) which starts in 1995. The geometry content of TIMSS covers no deductive Euclidean geometry, probably because it is taught in few western countries.

In analysing data of the National Educational Assessment Program (USA), Fennema and Carpenter (1981) found that males significantly outperformed females in the areas of geometry for all cognitive levels at ages 9, 13 and 17 with the difference increasing with age. Ben-Chaim et al. (1988) report gender differences favouring males in spatial visualization from adolescence onwards, but these skills are teachable and can be learned. The results of an experiment done by Battista (1990) suggest that, whereas males and females differed in spatial visualization and in their performance in high school geometry, they did not differ in logical reasoning ability or in their use of geometric problem solving strategies.

Tartre (1990) performed extensive research dividing males and females into high and low spatial ability groups. Her results do **not** support the argument that spatial skill is involved in helping males, more so than females, to achieve well in mathematics. She concluded that spatial skills contribute differently to the mathematics performance of males and females. Females with poor spatial skills experienced difficulty in accomplishing many tasks involved in solving mathematical problems.

A meta-analysis of more than 3 million pupils done by Hyde et al. (1990) provides evidence that the magnitude of gender differences has **declined** over the past three decades and that

the male superiority in geometry is relatively small. Feingold (1988) also reports a decline in cognitive gender differences, although Halpern (1989) challenges his views with regard to spatial abilities.

Feingold (1992) cautions that contemporary research on gender differences in intellectual abilities assumes homogeneity of variance. He showed that American males were consistently more variable than females in spatial visualization but concludes later (Feingold, 1994) that variability in mathematical and spatial abilities are not homogeneous across cultures.

1.2.3 Euclidean geometry

Relatively little recent research on differential gender performance in pure Euclidean geometry could be traced, probably due to the unpopularity of the subject during the past two decades in western countries. If gender studies include some geometrical content, it is usually only calculations or basic Euclidean geometry. The lack of research on the presenting of proofs in geometry is mentioned by Senk (1985). A study that was done on proof writing (Senk and Usiskin, 1983) eliminated differential gender exposure to previous geometry courses statistically and concluded that males and females perform similarly on proof writing in Euclidean geometry. SIMS stimulated a few articles on gender differences in geometry performance involving Asian pupils, but the geometrical content, however, was not of secondary school level and was referred to as visualization (Lummis and Stevenson, 1990).

Research on teaching methods in Euclidean geometry includes the research of van Hiele-Geldof (1957) and van Hiele (1973) in Holland and that of the Russian researchers Yakimanskaya (1991), Zykova (1954) and Kabanova-Meller (1950). These studies, however, do not generally report on gender differences.

1.2.4 South Africa

Visser (1985) looked at gender differences in mathematics achievement in South Africa but the research results were complicated by the low mean scores (35,5% for grade 11) on the achievement tests. She found **no** gender differences in mathematics achievement for grades 7, 9 and 11 (standards 5, 7 and 9), except for the 100 top performers in grade 11. The tests did not contain enough items to permit comparison for different branches of mathematics. Research of Nyangeni (1993) was hampered by problems in Transkei schools at the time. Grade 12 males performed significantly better than females on Euclidean geometry, but due to unforseen problems with the research, 61% of males and 85% of females failed the geometry test. Swanepoel (1990) ascertained that grade 9 boys mastered the basic concepts of Euclidean geometry better than girls.

Visser (1987; 1985) also considered affective factors and found that, for females, several attitudinal variables correlate highly with intended participation in mathematics, and that parental encouragement seems to be a major influence.

In a study carried out at the University of Potchefstroom (du Plessis et al., 1993), it was found that female first year students outperform male students in mathematics, although the male students tested higher for formal thinking. The better performance of the females was ascribed partly to better study and learning strategies.

1.2.5 Summary

The issue of gender differences in mathematics achievement is complex, with research outcomes being quite variable across countries. An important theme of recent research is to ascertain whether differences in mathematics achievement, where they exist, can be explained by cognitive or non-cognitive factors. Both have been researched in South Africa with somewhat conflicting results. Visser (1985) could not give judgement on cognitive factors except for top scorers but found considerable differences with regard to the influence of attitudinal and social factors. Swanepoel's findings (1990) link up with the school of thought that gender differences in geometry achievement do not manifest themselves

exclusively as a result of non-cognitive factors. The latter research was however done on a very basic level.

Differences in the mathematical background of students have an effect on most research done on gender differences because all the branches of mathematics e.g. algebra, geometry and trigonometry are not compulsory in most countries. What is more, international studies do not usually cater for differences in exposure to mathematics. According to Ethington and Wolfe (1984) and also Blithe et al., (1994) the amount of mathematics studied is the **most** important factor influencing the results of research on gender differences. Senk and Usiskin (1983) also support this view, and Baenninger and Newcombe (1989) suggest that further research is necessary where participation in spatial activity, such as geometry, is controlled.

In South Africa all students with mathematics as a subject at senior secondary level have theoretically the same mathematical background. At grade 11 (standard 9) level this includes 2 to 3 years of exposure to informal geometry and 3 years of exposure to formal deductive Euclidean geometry. For research purposes this is of great value because Euclidean geometry is only prevalent in a few western countries, one being South Africa.

Research on differential gender performance in mathematics, and especially geometry, is often handicapped by samples that are not comparably representative of both genders. (See section 3.1 for the suitable circumstances with regard to the present investigation.)

Possible gender differences in geometry and/or mathematics achievement can have far-reaching effects for a developing country like South Africa. The present and projected manpower shortages in South Africa in the scientific, professional, managerial and technical fields have been well publicised. Today, knowledge of mathematics is called the 'critical job filter', which facilitates entry into a wide variety of careers. Insufficient knowledge of mathematics, of which Euclidean geometry forms an integral part, may well limit a person's job opportunities and thus, indirectly, lower the quality of life possible to that person.

1.3 THE AIM OF THIS STUDY

The main aims of this study are to endeavour to contribute to knowledge concerning the teaching and learning of Euclidean geometry by

* appraising the literature on Euclidean geometry from a historical and global perspective, and how it reflects on the individual, specifically in relation to **gender;**

* establishing the contribution that Euclidean geometry makes to the mathematical education of high school pupils; and

* setting up a test situation where performance in Euclidean geometry can be isolated and analyzed for possible **differential gender performance** utilizing a homogeneous sample such as exists in some South African secondary schools;

1.4 STATEMENT OF THE RESEARCH PROBLEM

The investigation will focus on the following questions:

* Is there any gender difference with regard to achievement in Euclidean Geometry at senior secondary level in selected South African schools ?

* If gender differences in geometry achievement exist, which gender do they favour and what is the magnitude and nature of the differences ?

1.5 TERMINOLOGY

One of the reasons for conflicting results in research concerning gender differences in mathematics and geometry achievement is poorly defined concepts. It is therefore essential to define the terms involved very carefully.

1.5.1 Sex differences / gender differences

The terms 'sex' and 'gender' are used in an inconsistent way in the literature. 'Sex-difference' is lately being used to refer to biological differences between females and males; 'gender difference" to non-biological characteristics such as psychological or social aspects (Leder, 1992). Since this research will not include pure biological data, the term 'gender' is preferred. It is important to stress the close link between the two; sometimes it is difficult to decide whether differences found are based on the one or the other. (Leder, 1992).

1.5.2 Geometry and Euclidean geometry

In general, the term "geometry" is loosely used in the literature to mean anything related to spatial concepts. The geometries taught in South African schools are called either Analytical geometry or just geometry, meaning 'Euclidean geometry'. Therefore, the term Euclidean geometry will be used to indicate pure synthetic Euclidean geometry. The term 'geometry' may mean geometrical calculations, transformation geometry and sometimes even vectors as used in international literature. (See section 2.2 for further definitions with regard to geometry.)

1.5.3 Spatial ability, spatial visualization and spatial orientation

An explanation for the inconsistency in the literature on the expected relationship between spatial ability and mathematics performance has been lack of precise definitions regarding spatial abilities (Leder, 1992).

Tartre (1990) pointed out that researchers often identify a spatial factor, distinct from numerical and verbal factors, as a component of general intelligence. The way in which spatial ability contributes to performance in mathematics, is however vague. McGee (1979: 897) distinguished two components of spatial skills - 'spatial visualization and spatial orientation'.

* 'Spatial visualization' means that the viewer has to mentally move an object (usually a pictorial representation) or the perspective of the person viewing the object must move. **Movement**, either mental rotation or mental transformation, distinguishes this skill from the next skill.

* 'Spatial orientation' requires that the subject understands a visual representation or some change that has taken place between two representations. Spatial orientation can be further divided into different components (Tartre, 1990), which will be discussed in section 2.5, because this specific skill is, according to the researcher, related to the visual component of Euclidean geometry.

1.5.4 Cognitive

Bloom et al., (1956) concluded that the cognitive domain includes those facets which deal with the recall or recognition of knowledge and the development of intellectual abilities and skills.

1.5.5 Item

Consistent with the literature, the term 'item' will often be used instead of 'multiple choice question' but it will not necessarily replace the term 'multiple choice question'. An item is a single unit in a test or subtest which constitutes the smallest subdivision into which a test can be divided. It consists of:

* A stem - the part in which the question is asked, a problem is posed or an instruction is given;
* Options - the part which contains the key or correct answer as well as several distractors which are incorrect answers that are attractive to ill-informed testees (Barnard, 1993a).

1.6 RESEARCH METHOD

This study is an empirical investigation done in the field of pure Euclidean geometry and the teaching thereof at secondary level.

1.6.1 Measuring instrument

The researcher compiled two geometry tests covering those circle theorems of Euclid prescribed by the syllabus of the Transvaal Education Department (South Africa) for grade 11 (Standard 9) pupils. The first test is made up of items drawn mostly from the standardized itembank of the South African Human Sciences Research Council (HSRC). The second test consists of open ended problems or geometry riders, as they are sometimes called. The latter required a fair amount of proof writing. The rationale of the tests is that they be valid instruments to test pupils' achievement and understanding of Euclidean geometry.

1.6.2 Sample

Approximately 820 pupils from traditional Afrikaans and English high schools near Johannesburg, South Africa (now in the province of Gauteng), participated in the experiment. Efforts to include a sample of black students have not succeeded due to difficulties in the schooling system at the time of the empirical investigation.

1.6.3 Analysis

A detailed analysis, using both classical and modern test theories was carried out. For the multiple choice test, a T-test for unrelated groups was done to compare the mean scores of males and females. The difficulty value and discrimination value of each item were calculated and the reliability of the test was determined using Kuder-Richardson formulas. The frequency of each distractor was calculated separately for each gender group. The responses on the multiple choice questions were also analyzed using the one dimensional, three parameter logistic model of the Item Response Theory (IRT). The IRT is a statistical

theory based on mathematical models expressing the probability of a particular response to a question as a function of the ability of the testee and of certain characteristics of the question (Barnard, 1993). The three parameter logistic model is considered to be sensitive in determining possible differences between groups, in this case, gender groups (Barnard, 1991; Hambleton et al., 1991; Bleistein, 1986).

The open ended questions (riders) of a subsample (n = 203) were scrutinized and scored by hand, the respective means compared for possible gender differences and eventually correlated (Pearson product moment) with the multiple choice test. The open ended questions (riders) were also analyzed diagnostically to determine possible differences between the two groups. The scores on both these geometry tests were also correlated with the pupils' end of the year score for mathematics in general, which includes a combination of algebra, trigonometry and Euclidean geometry.

1.7 A SYNOPSIS OF THE CONTENTS OF THIS THESIS

In chapter 2, a historical overview of the teaching of Euclidean geometry in South African schools and elsewhere is given, including the teaching problems usually encountered and recent research in this field. The nature of Euclidean Geometry is looked at briefly and the merits of including it in the secondary school curriculum are motivated.

Chapter 3 comprises a literature study on gender differences in mathematics and geometry achievement. Because of their close interrelationship, spatial ability, spatial visualization and spatial orientation are also scrutinized with regard to gender related differences. Research on cognitive gender differences is appraised and non-cognitive factors are included where they are of importance for this study.

Chapter 4 is a discussion of measurement theories, viz. Classical Test Theory and the Item Response Theory. Differential item functioning is considered within both frameworks. Items and open ended questions are also compared as means of testing.

Chapter 5 contains the details of the empirical investigation, the results of the study, the statistical analysis of the data and also a discussion thereof.

Chapter 6 is a summary of the outcome of the research, emphasizing the major conclusions. Recommendations for future research as well as the limitations of the study, are included.

CHAPTER 2

EUCLIDEAN GEOMETRY

"Geometrical truths are in a way asymptotes to physical truths, that is to say the latter approach the former indefinitely near without ever reaching them exactly."

D'ALEMBERT,
(Paris, 1898)

2.1 INTRODUCTION

Geometry is one of the most controversial sections of secondary school curricula in most countries (Masingila, 1993; Toumasis, 1991; Love, 1989; Human et al., 1989; Usiskin, 1987; Howson and Kahane, 1986; Suydam, 1985). A study of the literature leads to the conclusion that, regardless of the many problems encountered in the teaching of Euclidean geometry, it remains of great educational value.

> "In spite of the unanimous agreement about the necessity and desirability of the teaching and learning of geometry at the secondary level, criticism of, and dissatisfaction with, the course has come from many sources." (Masingila, 1993: 38).

Motivation for the continued teaching or return to teaching of Euclidean geometry will be discussed as the different aspects of geometry are considered.

2.2 WHAT IS EUCLIDEAN GEOMETRY ?

In South Africa, Euclidean geometry is also known as synthetic geometry, or congruence geometry, with emphasis on congruency and similarity of triangles (Human et al., 1989). It is characterized by a formal structure, the well known axiomatic deductive structure. Strictly speaking it is only a specific **approach** to Euclidean geometry and confusion arises

because it is sometimes called Euclid's approach (Usiskin, 1987) and is thus seen as the only real Euclidean geometry. Several other approaches (Human et al., 1989; Usiskin, 1987) to Euclidean geometry can be taken, for example

* a transformation approach;
* a vector approach, or
* an analytical approach.

In a vector approach the axioms of a vectorspace are the basis of the deductive structure. This approach was propagated a few decades ago, and it eventually lead to the synthetic approach, or Euclid's approach as it is sometimes called, becoming disfavoured. Freudenthal (1973) criticized the vector approach to Euclidean geometry because of the lack of spatial orientation in this approach. A pupil can only interpret vector geometry meaningfully if he has already had some spatial experience of geometry. Analytical geometry was criticized more than a century ago as a method that presupposes technique and not thought processes (de Villiers and Oberholzer, 1986).

Vector geometry, transformation geometry and analytical geometry can **not** be contrasted with Euclidean geometry as if it is non-Euclidean. Non-Euclidean geometries include the classic geometry of Lobachevski and Bolyai, and are seldom, if ever, mentioned in school textbooks.

2.2.1 Terminology

Approaches

Because of the above-mentioned confusion with the terms Euclidean geometry and other geometries, the following will apply: As in South Africa, Euclidean geometry will refer to the synthetic approach and other approaches to Euclidean geometry will be called transformation geometry, analytical geometry, etc.

Formal and informal

Informal geometry and formal geometry can only be distinguished at the extremes of a spectrum (Usiskin, 1987; Van Zyl, 1942). At the one end, the term 'informal geometry' is used to refer to the more experimental activities such as folding, construction, drawing, recognizing, etc. At the other extreme, 'formal geometry' usually refers to higher order thinking as in generalizations and includes rigorous deductive reasoning. According to Human et al. (1989), formal geometry is characterized by activities such as systematizing and proof. These 'levels' are not distinct and not independent. What is informal for one student may be very formal for another.

2.2.2 South Africa

In South Africa, informal geometry is generally taught for two to three years, namely in the sixth, seventh and eighth school years, and is then followed by four years of formal deductive Euclidean geometry until the last year of secondary schooling. Analytical geometry is currently part of the syllabus for the last year of secondary school.

2.3 HISTORY OF EUCLIDEAN GEOMETRY

The concern of the earliest mathematicians was with measuring the earth in a practical way. The word geometry stems from 'geo' which means 'earth' and 'metrein' which means 'to measure'.

2.3.1 Euclid's Elements

Greek mathematics experienced a crisis in about 350 BC when the followers of Pythagoras realized the existence of numbers that are not rational numbers. Greek mathematics became more geometrically orientated and less arithmetical. It took almost 2000 years to change the situation. Around 300 BC Euclid undertook the task of collecting, ordering, formulating and verifying the contemporary knowledge of synthetic (without numbers) geometry in his well known 'Elements'. This work is considered to be the first explicit example of deductive science and became, to a considerable extent, the model for all future sciences.

Scarcely any other book, except the Bible, has been more widely circulated or studied. Over a thousand editions of the 'Elements' have appeared since the first printed version of 1482.

Euclid's whole system of about 500 theorems was based on a few definitions, 5 postulates of a geometrical nature and 5 axioms that were meant to hold for all mathematics. Euclid's most famous postulate (see section 2.3.2), the fifth, can be summarized as follows: If two lines are intersected by a third line (the intersecting line) in such a way that the sum of the two interior angles on one side of the intersecting line is less than two right angles, then the two lines will intersect on that same side of the intersecting line (Baart, 1992). Euclid's axioms included assumptions such as equality of objects that are equal to the same object.

If $a = b$
and $b = c$,
then $a = c$.

In modern mathematics there is no distinction between axioms and postulates. The Greeks saw mathematical concepts, like Euclid's axioms, as being independent of experience and having a reality of their own. Until the eighteenth century Euclid's axioms were regarded as basic truths which needed no justification beyond their self-evidence and the field of mathematics was largely dominated by Euclidean geometry. (Ernest, 1991; Burton, 1985; Schutte, 1983; Boyer, 1968). (See also section 2.4.1).

> "For 2000 years mathematics was viewed as a body of infallible and objective truth, far removed from the affairs and values of humanity." (Ernest, 1991: xi).

2.3.2 Crisis - non-Euclidean geometries

Until well into the nineteenth century, Euclidean geometry was regarded by most people as the firmest and the most reliable branch of knowledge. Euclid's fifth postulate, the parallel postulate, though, was questioned by mathematicians for many centuries. Was it independent of the other postulates or could it be proven from them? In his doctoral

dissertation of 1973, Klügel made the observation that the certainty with which men accepted the truth of Euclid's parallel postulate, was based on experience.

> "This observation introduced, for the first time the thought that experience, rather than self-evidence, substantiated the axioms." (Kline, 1980: 81).

The conflict around Euclid's fifth postulate was resolved in an unexpected way during the nineteenth century. Since it could neither be deduced nor contradicted, could it be replaced? This resulted in the exploration and formation of non-Euclidean geometries that eventually ended the 2000 year reign of Euclid. It became clear that mathematicians had accepted axioms for geometry that seemed correct on the basis of their limited experience and they had been deluded into believing that these axioms were self-evident truths. Many others contributed, but Lobachevski and Bolyai were the real inventors of non-Euclidean geometry in the nineteenth century (Ernest, 1991; Burton 1985; Davis and Hersh, 1981; Kline, 1980; Jones, 1969; Boyer 1968). They

> "did not themselves comprehend the revolution in mathematics that was to grow out of their invention of another geometry." (Jones, 1969: 10).

So widely was Euclidean geometry accepted at that time to be the geometry of physical space, that it took several years, but from their work there developed

* an understanding of the nature and importance of axioms;
* the idea of the possibility of different mathematical systems, and
* attempts to also axiomatize other branches of mathematics (Jones, 1969).

This 'loss of certainty' (Kline, 1980), with regard to the foundations of mathematics, was almost an unbearable philosophical loss because the model of Euclid was seen as a model of all human knowledge (Davis and Hersh, 1981).

2.3.3 Number Theory

Mainly because of the crisis in geometry, the emphasis shifted towards number theory. Towards the end of the nineteenth century, several schools of thought developed, such as logicism under Russell (England), intuitionism (later called constructivism, in a mathematical sense) under Brouwer (Holland), and formalism under Hilbert (Germany), all with the purpose of regaining a consistent foundation for mathematics and to re-establish its certainty. Russell's discovery that the axioms of set theory also lead to contradictions further emphasized the 'crisis in foundations' (Volmink, 1988; Davis and Hersh, 1981).

Hilbert, the founder of Formalism, was convinced that the contradictions in set theory were due to lack of precision in the basic concepts of mathematics and a lack of rigour in its deductive proofs (Hanna, 1983). The aim of his school was thus to show that every branch of mathematics can be constructed in the form of a consistent and complete axiomatic system.

> "It was hoped that the application of rigorous methods would yield mathematical results that were beyond any contradiction." (Volmink, 1988: 23).

Formalism dominated the philosophy of mathematics during the 1940s and 1950s and had a great effect on the Bourbaki-writers. Bourbaki is the *nom-de-plume* of a group of mostly French mathematicians who have undertaken to reorganize all mathematical knowledge in terms of basic components and in the process have also tried to establish the basic theorems of Euclidean geometry by means of linear algebra (Dieudonné, 1973). The mathematical world was flooded with publications on set theory, algebra and analysis.

This style of viewing mathematics as formal, static, fixed and unchangeable (Ernest, 1991), gradually penetrated undergraduate mathematics and eventually into schools. It lead to radical curriculum changes which loosely became known as the "new math". This new movement was characterized primarily by a more abstract approach to mathematics in schools with emphasis on the structure and the learning of rules as in set theory. There was a general attempt to dissociate mathematics from all practical applications (Hanna, 1983; Davis and Hersh, 1981).

Further factors contributing to the movement towards formal mathematics are mentioned by Hanna (1983). Technological development after the second world war resulted in increased interest in mathematics at tertiary level. At the same time, a discontinuity between secondary mathematics and the discipline taught at university was observed. To provide students of higher mathematics with a better mathematical background, school mathematics was to be more similar to mathematics as an intellectual discipline.

2.3.4 Modern View

The 'crisis of the foundations' of mathematics was partially solved by Gödel. In 1931 Gödel published his work on axioms with the new approach that any axiomatic system is **incomplete** (de la Rosa, 1988; Volmink, 1988; Davis and Hersh, 1981), and that axioms are just convenient assumptions that can be refuted (Schutte, 1983; Lakatos, 1976). Gradually mathematicians came to accept the change.

In the meantime, educationists realized that the 'new math' was not as successful as was originally thought. The trend towards formalism in the school curriculum was criticized as early as 1958 (Hanna, 1983). In 196. Wheeler wrote:

> "Mathematics is made by men and has all the fallibility and uncertainty that this implies. It does not exist outside the human mind, and it takes its qualities from the minds of men who created it. Because mathematics is made by men and exists only in their minds, it must be made or re-made in the mind of each person who learns it. In this sense mathematics can only be learnt by being created." (Wheeler in Ernest, 1991: 205).

Several contemporary philosophers and mathematicians such as Lakatos (1976), Davis and Hersh (1981) and Tymoczko (1985), think along the same lines. They all preach a more informal approach to the teaching of mathematics and more emphasis on practical experience, similar to the problem solving view of Polya (1957). Lakatos (1976) views mathematics as a quasi-empirical discipline and, by means of a case study, challenges formalism as follows:

> "Its modest aim is to elaborate the point that informal, quasi-empirical, mathematics does not grow through a monotonous increase of the number of indubitably established theorems but through the logic of proofs and refutations." (Lakatos, 1976: 5).

His purpose was to show the inadequacy of formalism by presenting an alternate picture, a picture of mathematics living and growing, and not 'fossilized in formal axioms' (Davis and Hersh, 1981). In recent years the reaction against formalism has been growing. Educationists increasingly turn to the concrete and the applicable and emphasize the mathematics student as a creative thinker.

> "... they are affirming that mathematics is fallible, changing, and like any other body of knowledge, the product of human inventiveness." (Ernest, 1991: xi).

In the modern view, the emphasis on the human mind in the development of mathematical knowledge, is quite clear. This modern view can be traced back to Intuitionism, the school of philosophy of mathematics that views all mathematical knowledge as constructible by the human mind. In the past decade it contributed to the foundations of the so-called **constructivistic** approach to the teaching of mathematics. Constructivism, in a mathematical sense, has also derived some of its ideas from Piaget's psychology of learning.

At the heart of what is commonly known as constructivism is the contention that knowledge is actively constructed by the learner. Meaning is constructed by reflection on varied experiences, by investigation and invention, by social interaction and negotiation, and sometimes involves explanation (Ernest, 1991; Stoker, 1991).

Many variations of constructivism have been developing recently, from radical constructivism (von Glasersfeld, 1990), with too much emphasis on individual experience (Goldin, 1990), to social constructivism where knowledge is a shared experience (Prawat and Floden, 1994).

2.3.5 Motivation

Euclidean geometry is inseparable from the history of mathematics because it ruled the mathematical world for almost 2000 years. The discovery of the non-Euclidean geometries brought about a turning point in mathematical reasoning as well as in philosophy. It is almost unfair to teach pupils any mathematics without them knowing about the struggle that was part of the development of the mathematics that they learn at school. Thus,

> "from a mathematical and historical point of view, Euclidean geometry does justify a place in the curriculum." (Kapadia, 1985: 54).

A similar point of view is held by Kriel (1975), Thom (1973) and Malan (1971). (See also section 2.6.6.)

2.4 THE STRUCTURE OF EUCLIDEAN GEOMETRY

Much of the criticism of Euclidean geometry revolves around the axiomatic deductive approach, so this will be discussed in more detail.

2.4.1 Axioms

Concepts like 'line' and 'point' are undefined in synthetic geometry. **Axioms** are 'hypotheses', made with regard to the undefined entities (Kriel, 1975), and they can be refuted by undesirable consequences (Schutte, 1983). The axioms form the basic premises of the mathematical system and should not contradict each other.

It is important to note that axioms are **not** self-evident truths as is still widely believed by pupils and teachers. Axioms are only convenient assumptions (Human et al.; 1989), valid statements (Kriel, 1975), arbitrary statements (Burton, 1985) or rules (Hendricks, 1991). The modern view is that axioms (postulates) are arbitrary statements, formulated abstractly with no bearing on their 'truth', but accepted without justification as a foundation for reasoning (Burton, 1985).

2.4.2 Definitions

In any mathematical system, defined concepts play an important role. Each new concept must be defined in terms of the known concepts. Definitions should be unambiguous (Kriel, 1975).

2.4.3 Theorems

Not all, but only the more important logical conclusions deduced from the axioms and the definitions become the theorems of the structure (Kriel, 1975).

2.4.4 Inductive and deductive reasoning

Both inductive and deductive reasoning are important for the development of mathematics and geometry (see section 2.6.4). O'Daffer (1990) gives the following description:

> "Inductive reasoning is a reasoning process in which information about some members of a set is used to form a generalization about other or all members of the set.
>
> Deductive reasoning is a mathematical reasoning process in which valid inference patterns are used to draw conclusions from premises." (O'Daffer, 1990: 378).

Euclid would not have had axioms and theorems to organize deductively in a system without the amount of geometrical knowledge that the Egyptians and the Babylonians obtained inductively over many years, (Kriel, 1975). While the formal structure of Euclidean geometry was preceded by practical application in land measurement, the non-Euclidean geometries were developed deductively years before applications and examples were found.

2.4.5 Motivation

Euclidean geometry is often criticized for having a rigid structure. In section 2.6.4 it is suggested that the core of the problem is the way it is taught to be regarded as a pre-existing body of knowledge polished to perfection and seemingly beyond the ability of most people. (See also section 2.6.1.)

2.5 SPATIAL AND VISUAL ELEMENTS

Clements and Battista (1992) stated in their review of spatial reasoning and geometry that these two aspects are related. Tartre (1990) pointed out that factor-analytical techniques show a spatial factor as part of general intelligence but the exact way in which spatial skills influence performance in mathematics is not yet known. The lack of satisfactory definitions for spatial skills causes problems (Friedman, 1994; Clements and Battista, 1992; Zimmerman and Cunningham, 1991; Tartre, 1990; Linn and Petersen, 1985; Bishop, 1989; 1983). The dual usage of the term 'visualization' causes confusion. Sometimes it is used to describe visual processes in mathematics (Zimmerman and Cunningham, 1991; Bishop, 1989; 1983; Hershkowitz et al., 1990; Hershkowitz, 1989) and other researchers use it as defined by Tartre (1990) in sections 2.5.1 and 2.5.2. (See also section 3.3.1.4.)

Krutetskii (1976: 317) investigated differences in performance in algebra and geometry in Russia. He referred to two different modes of thought, i.e., 'verbal-logical and visual-pictorial'. He classified students as

* **'analytic'** - when verbal-logical modes of thought are dominant in mathematical problem solving, even for problems that could be approached visually;

* **'geometric'** - when visual-pictorial schemes are dominant even for problems more easily solved with analytical methods, and

* **'harmonic'** - when the two different modes of thought are more or less in equilibrium.

Some researchers (Brown and Presmeg, 1993; Presmeg, 1993; 1986; Zimmerman and Cunningham, 1991; Bishop, 1989; 1983) moved away from the factor-analytical approaches in the study of spatial ability, and focused on visualization as a process in the learning of mathematics in general, in view of the questions raised as to what precisely spatial ability tests are measuring. Some of these tasks can be solved by analytical procedures (Bishop, 1983). According to Bishop (1989) the visualization process appears to involve the learner in constructing some kind of visualization and using it appropriately. It is obviously a very general description and includes any visual representation of mathematical data. Bishop (1989: 11; 1983: 184) suggests two different types of spatial ability constructs important for the learning of mathematics, i.e.,

* '(IFI) - the ability to interpret figural information' which involves understanding visual representations and spatial vocabulary used in geometric work, graphs, etc. and

* '(VP) - the ability for visual processing', involving manipulation and transformation of visual representations and images, and translation of abstract relationships into visual representations.

The following description links mathematical visualization with images:

> "Mathematical visualization is the process of forming images and using such images effectively for mathematical discovery and understanding." (Zimmerman and Cunningham, 1991: 3).

According to Clements (1979), students with well developed spatial skills should be capable of imagining spatial arrangements of objects and of manipulating visual images. Clements and Battista (1992) sum it up as the maintenance and manipulation of a high quality image of the stimulus, but Hershkowitz (1989) warns that a high quality image sometimes leads to inflexible geometrical thinking. The exact relation of 'imagery' to performance in mathematics or geometry is not clear (Clements and Battista, 1992) and is also characterized by large individual differences (Brown and Presmeg, 1993). Presmeg (1986) and Brown and Presmeg (1993) describe the following five types of imagery:

* 'Concrete imagery' - picture in the mind.
* 'Memory images' - memory images of formulae, etc.
* 'Kinaesthetic imagery' - muscular activity involved.
* 'Dynamic imagery' - transformation of a concrete visual image.
* 'Pattern imagery' - a highly abstract form of imagery.

Imagery has been investigated in relation to simple geometrical problems (Presmeg, 1986), but is not a popular topic in gender research. (See section 3.3.1.4.) For the purposes of this study, the analysis of McGee (1979) and of Tartre (1990) is preferred in view of the amount of related research in the field of gender differences. (See section 3.3.1.4.) In his review of spatial factors, McGee (1979: 892) concluded that available evidence conclusively demonstrates the existence of two distinct spatial factors, i.e. 'spatial visualization and spatial orientation'.

2.5.1 Spatial visualization

This category implies that a person has to mentally move an object, usually a pictorial representation. Thus, according to Tartre (1990: 29) tests of spatial visualization expect that the viewer mentally moves the representation of an object, either by

* 'mental rotation', i.e., turning the object, or
* 'mental transformation', i.e., rearranging the object.

Tartre (1990) subdivided both components further for two and three dimensional objects. Some tests only cover questions on mental rotations and transformations, thus only measuring this one component of spatial ability.

2.5.2 Spatial Orientation

> "Spatial orientation involves the comprehension of the arrangement of elements within a visual stimulus pattern, the aptitude to remain unconfused by the changing orientations in which spatial configurations may be presented,

and the ability to determine spatial orientation with respect to one's body". (McGee, 1979: 897).

Tartre (1990: 34) divides spatial orientation into two categories:

* 'Reorganised whole' - which can be further divided into two types of tasks, i.e. those involving an 'ambiguous figure' and those showing 'multiple representations'. For the researcher both of these appear to be of a geometrical nature, but not as related to Euclidean geometry as the next category.

* "**Part of field** - ... which is concerned with the relationship of part of a representation to the whole field, either presented visually or imagined. The two kinds of tasks within this category are

  * **to find part of a whole** - ... the entire representation is given visually and the subject is asked to locate a particular part of it, such as a line or simpler figure. ... Asking students to identify or work with part of a complex drawing in **geometry** also requires use of this skill.

  * **to fit part into the whole** - requires that the subject recognize how a part of a visual representation fits into the whole picture.

Although they seem to be opposites of one another, these skills operate together in some problems". (Tartre, 1990: 37 - 40). (Bold added.)

Geometry

Tartre (1990) thus suggests that 'to find part of a whole' is related to the visual aspect of geometry and that the other category 'to fit part into the whole' sometimes operates together with the first one. The researcher interprets it as follows with regard to Euclidean geometry. If YA is given as tangent to circle AFD (Fig. 2.1) a student has to **'find the bold part of the whole'** to identify equal angles. If YA has to be proved a tangent to circle

AFD, then the student has to form a mental picture of the tangent theorem (Fig. 2.2) and **'fit that picture into the whole'** in order to identify the angles that should be proven equal.

These two categories of spatial orientation, as described by Tartre (1990), appear to be partially related to the visual aspects of Euclidean geometry. Tartre (1990) warns that her classification scheme differs from some literature on spatial skills. Linn and Petersen (1985) use a completely different categorization. (See also section 3.3.1.4.)

Figure 2.1

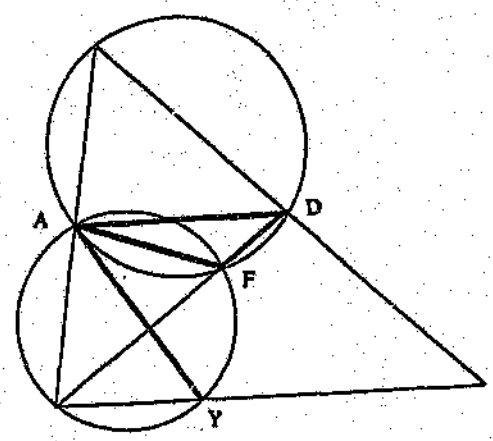
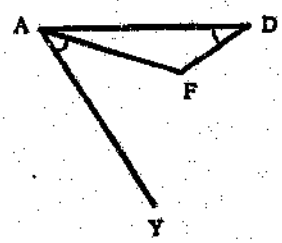


Figure 2.2



2.6 THE TEACHING OF EUCLIDEAN GEOMETRY

Seen against the background discussed in sections 2.1 - 2.5, it is conceivable that the teaching of Euclidean geometry may be problematic. Studying the literature, it is clear that in spite of all the teaching problems experienced it is considered to be of great educational value if taught in a proper way (Chazan, 1993; Schumann and de Villiers, 1993; de Villiers, 1991; Human et al., 1989; Usiskin, 1987; Dreyfus and Hadas, 1987; Kapadia, 1985; Egsgard and Fletcher, 1984; Lang and Ruane, 1981; Bell, 1978; van Hiele, 1973; Thom, 1973; 1971; Malan, 1971). Most agree, however, that the teaching methods must be less formal than in the past. Sticking to the synthetic approach may eventually be favourable for South Africa (Human et al., 1989). Asian countries like Japan, that perform

2.6.1 Dimensions in the teaching of Euclidean geometry.

Many of the complaints about geometry can be seen as a result of not having taken a **multidimensional** view of the understanding of geometrical concepts. A unidimensional approach, along with a rigorous acquaintance with axioms, theorems and proofs is often the only association with geometry pupils have.

Usiskin (1987) argues that at least four major dimensions should be addressed in the proper teaching of geometry. These dimensions are not completely independent and cannot be ordered for curriculum purposes. Many geometrical concepts are multidimensional.

Note: The following four dimensions, as described by Usiskin (1987), must not be confused will the levels of geometrical thinking of the van Hieles. (See section 2.6.5.)

Visualization, drawing, construction

Regardless of the need for good visual and spatial perception in advanced mathematics and science and in specific occupations, we all need the ability to visualize objects in space and thus appreciate their spatial interrelationships. We also often need to be able to interpret two-dimensional representations of three-dimensional objects in everyday life. Developing visual skills can be considered a basic skill (Sherard, 1981). Insufficient practising of basic visual skills later manifests in pupils as an inability to 'read' a visual representation or geometrical diagram.

Geometry requires the drawing of simple figures and the interpretation of visual patterns. Drawing skills should be developed, also the drawing of three-dimensional objects on graph paper. This activity encourages students to analyze figures, to use ratio and proportion, and to think about similar figures. It should not be restricted to the less able students to keep them busy while the others do 'real' geometry (Usiskin, 1987; Hoffer 1981). Simple drawings should naturally be developed into more precise methods, like constructions. Construction of geometrical figures is a valuable tool for investigative work, and a good first acquaintance with theorems, which will eventually be proved.

Study of the real physical world

Usiskin (1987) reasons that although geometry developed from the physical world, connections with that world are generally not addressed in textbooks. Few textbooks in South Africa mention the numerous examples from nature, such as the shape of the honeycomb, the spider-web or seashells which enhance geometrical ideas. Geometry is apparent in nature as well as in skyscrapers, suspension bridges, Gothic cathedrals and aqueducts (Sherard, 1981). We almost daily take area or volume measurements. Hoffer (1981) suggests practical applications of geometry in the fields of architecture, astronomy, engineering and the business world.

> "Newton's genius was to realize that the physical world behaved in ways that could be expressed by mathematical formulas." (Usiskin, 1987: 27).

A vehicle for representation

'Geometric representations of non-geometric ideas' (Usiskin, 1987) are familiar in the mathematical world. In algebra the numberline is used to represent the more abstract concept of real numbers and geometric shapes are used to represent fractions; in analytical geometry an equation is represented by a line or a circle; in statistics data are represented by circles, lines or bars, and in calculus the concept of area is used to illustrate what is meant by the evaluation of an integral, etc. Currently the idea of describing something mathematically is called mathematical modelling (Hoffer, 1981).

One of the best examples of a mathematical model is Euclidean geometry. For the Greeks their geometry was a perfect model which described the universe as they experienced it. Mathematical models today are used in many fields, not restricted only to practical applications, but also in the related fields of engineering, astrology, architecture and science.

Euclidean geometry as a formal mathematical system

Geometry was the first branch of mathematics to be organized logically. Until the previous century it was the only branch of mathematics to be so organized in a structure. It is therefore not strange that Euclidean geometry is regarded as the means of showing students how to build an axiomatic system, of teaching them the deductive character of mathematics, and of training them to write formal proofs (Schuster, 1975).

Motivation

The first three of Usiskin's (1987) dimensions may probably be covered in the teaching of several approaches to geometry, but

> "secondary school geometry would fail to be true to its heritage if it did not emphasize the understanding of a mathematical system." (Farrell, 1987: 236).

Various attempts have been made to replace the synthetic approach to geometry as a means of teaching an axiomatic deductive structure, but many educationists do not agree that the alternatives have been any better than the traditional approach (Chazan, 1993; de Villiers, 1991; Human et al., 1989; Kapadia, 1985; Nicholson, 1985; Lang and Ruane, 1981; Kriel 1975; van Hiele, 1973; Thom, 1973; 1971; Shibata, 1973; Malan, 1971). Some prefer a local axiomatic approach, i.e., the systematization of small sets of related theorems, to the traditional global axiomatic approach, especially in the introductory stages of a deductive system (Human et al., 1989; Nicholson, 1985; Lang and Ruane, 1981; Schuster, 1975;). The researcher supports this view but further comment is outside the scope of this study.

2.6.2 Historical overview of the teaching of Euclidean geometry

Even before the first publication of Euclid's Elements in 1482 in Venice (Burton, 1985), it was used in European Universities. From the eighteenth century Euclidean geometry was also taught in secondary schools in Europe (Human et al., 1989) and from 1850 in the USA

also taught in secondary schools in Europe (Human et al., 1989) and from 1850 in the USA (Masingila, 1993). In 1794 Legendre published his 'Elements de Geometrie' and consequently the teaching of geometry in Europe became more flexible and less rigorous. In Germany and France a more practical and application orientated approach was taken. From the nineteenth century, Germany deviated from the strict Euclidean sequence. Legendre's influence was also visible in the USA (Human et al., 1989; van Zyl, 1942).

In Britain, geometry was introduced at secondary level during the first half of the nineteenth century. Their geometry was characterized by Euclidean strictness and in some textbooks even the strict Euclidean sequence was followed until the beginning of the twentieth century. Geometry education consisted of a study of axioms, definitions, constructions and theorems in a fixed order, without any reference to practical application and with very few exercises (Human et al., 1989). Giles (1982) reported the same narrow approach to Euclidean geometry during his school years in England in the 1940s. Without ever questioning it, theorems were learned by heart. Approximately 55% of the marks were for bookwork and those able to memorize proofs were able to pass.

South Africa

Because South African education has been under strong British influence in the past, this rigidity with regard to Euclid had far-reaching consequences for the teaching of geometry in South Africa. Most textbooks were imported from England and, with them, the very formal approach. The first Afrikaans textbook to deviate from the strict Euclidean sequence to a simpler sequence, was that of du Toit in 1921 (Human et al., 1989; van Zyl, 1942), but according to the author himself (du Toit, 1921), his book followed the English model.

2.6.3 The current situation

Over the past four decades the curricula for geometry changed drastically in most western countries mainly as result of the problems mentioned in section 2.6.4. Apart from South Africa, the USA and Greece where Euclidean geometry is still taught, it almost disappeared from curricula in western countries. In some countries, like England, it is taught at selected

schools. In Russia, Japan and other Asian countries, Euclidean geometry is still taught on a large scale.

Japan

In Japan Euclidean geometry features quite strongly (Miwa, 1992; Nohda, 1992). Geometry does not exist as an independent subject but is integrated systematically from grade 1. In grade 2 pupils are already introduced to the fundamental figures such as squares, rectangles and right-angled triangles. Various other geometrical figures and solids are studied intuitively in elementary school and mathematical inferences and proofs are introduced in grades 8 and 9. In grades 8 and 9, almost 140 hours per year are spent on mathematics. That is considerably more than the 100 hours (examinations excluded) proposed by the Transvaal Education Department (Department of Education and Culture, 1992) in South Africa for teaching grade 8 pupils mathematics. In Japan the enrolment in compulsory education was 99,99% in 1986 (Nohda, 1992). In several international comparative studies over the past decade, Japanese school children outperformed children from other advanced western countries like the USA and Britain (Lummis and Stevenson, 1990; Robitaille and Garden, 1989). (See also section 3.3.1.6.)

In order to gain insight into the problems of learning and teaching geometry, Whitman et al. (1993), investigated the attained geometry curriculum in Japan and Hawaii, relative to the van Hiele level theory. They found that, on average, in grades 1-8, American textbooks devote about 8 percent of their pages to geometry, whereas the Japanese textbooks allocate about 27 percent. The Japanese textbooks are more in line with the van Hiele levels (see section 2.6.5) than the American textbooks. The percentage of correct responses on the geometry test shows that the Japanese pupils are ahead of the Hawaiian pupils by about two years. Part of this difference can be accounted for by the geometry curriculum, which is also about two years ahead. Whitman et al. (1993) warn, however, that many other factors play a role in cross- cultural studies.

Harada et al. (1993) researched the differences in geometry performance between Japanese and French pupils in order to clarify the differences. They concluded that the French pupils

were limited in their choices with regard to 'procedures for proof'. In Japan, geometry is taught from a wider viewpoint than in France.

Russia

In Russia, the geometry program has been greatly influenced by the van Hiele levels. (See section 2.6.5.) As in Japan, Russia starts geometry in grade 1 with properties of geometric shapes and continues the study of geometry for ten years altogether (Shaughnessy and Burger, 1985).

Europe, England and the USA

Euclidean geometry is still taught in the **United States**, but it is characterized by serious problems, one of which is the fact that it is only taught for one year at secondary level. According to Usiskin (1987) the situation is as follows: At elementary school level there is no standard geometry curriculum, with the result that some students enter high school with very little knowledge of geometry. Most of them do not have enough knowledge of geometry to succeed at secondary school level. At secondary level there is a geometry curriculum but only about half of the students encounter it and only one third (Senk, 1985) of these understand it. This means that fewer than 15% of high school graduates in the United States master the writing of proofs. There is great concern (Masingila, 1993; Clements and Battista 1992; Usiskin, 1987) regarding the overall poor performance of pupils in Euclidean geometry in the USA. In 1989, the National Council of Teachers of Mathematics published recommendations for some revisions in secondary geometry in the USA. Increased attention is to be given to coordinate and transformation approaches with decreased attention to Euclidean geometry as an axiomatic system and the writing of proofs. The recommendations have had little effect because

> "the lack of specificity on the secondary geometry course seems to be an obstacle which reformers cannot work around." (Masingila, 1993: 42).

Schiddell and Ethington (1994) also mention the tendency of teachers to emphasize the more traditional instructional approaches involving a lot of memorization.

In **Europe** during the 1950s, the philosophy of mathematics and the teaching thereof was dominated by the Formalistic school of thought. (See section 2.3.3.) In France, Dieudonné, one of the Bourbaki group of writers who were under the influence of the Formalists, insisted that 'Euclid must go'.

> "For the trained mathematician of today, it is a triviality that the fundamental theorems of Euclidean geometry (in any number of dimensions, by the way, are very easily derived from the concept of a vector space ..." (Dieudonné, 1973: 18).

On the continent especially, the influence of the Bourbaki group shifted the emphasis in mathematics teaching away from a spatial approach to more algebraic methods. Euclidean geometry either disappeared from curricula or was replaced by other approaches to the teaching of geometry.

In **Britain** there has been a shift towards an experimental approach in geometry, and traditional Euclidean geometry was dropped except for the few selected schools that appear to have continued with it (Reddy, 1987). Lang and Ruane (1981) express concern about the lack of testing deductive reasoning in many O level and CSE examinations in England.

One of the problems experienced in England was the fact that there was no central control over syllabi for English schools. There are various approaches to mathematics in general and to geometry in particular. Even textbooks and syllabi are interpreted according to the teacher's knowledge and interest (Lang and Ruane, 1981). Love (1989) reported an inability of the recently introduced national curriculum in England to organize geometry after the fall of Euclid. He complained about disjointedness. Attempts to make geometry more experimental and interesting were successful, but this lead to fragmentation of the geometry curriculum.

Also Gower (1983) reports at ICME 4 that this

> "tends to weaken the structure and reduce geometry to a collection of isolated facts and procedures". (Gower, 1983: 153).

Lang and Ruane (1981) rightly concluded that the traditional school course in the geometry of Euclid had an overall 'identity' which we do not find in the geometry of modern mathematics.

In general, Euclidean geometry was replaced by transformation geometry, vector geometry or analytical geometry (Human et al., 1989; Bloom, 1984; Lang and Ruane, 1981). The approach became more experimental and informal (Brown et al., 1988; Reddy, 1987; Howson and Kahane, 1986; Suydam, 1985; Giles, 1982; Lang and Ruane, 1981). The amount of geometry taught also decreased (Bloom, 1988; Usiskin, 1987; Bloom, 1984; Lang and Ruane, 1981). Bloom stated in 1984 at the Fifth International Congress on Mathematics Education (ICME 5) in Australia:

> "There seemed to have been a general decline in interest in geometry at secondary school level. This had followed several years of attack on the traditional approach to the teaching of Euclidean geometry in favour of a more modern approach." (Bloom, 1984: 254).

At ICME 6 which was held in Hungary in 1988 Bloom concluded as follows:

> "These (meaning computers and statistics) together with a general community call to teach useful and relevant mathematics, meant that geometry in many western countries slipped from its central role in the secondary mathematics curriculum." (Bloom, 1988: 375).

It seems that there is no consensus as to exactly what should replace Euclidean geometry. Replacements were not always very well planned. Consequently, Usiskin (1987) reported that from the 40 proposed geometry items to test eight-grade students from the 23 countries

taking part in SIMS (Second International Mathematics Study - 1981/82), only 2 or 3 items were agreed to by all countries. As a result, every country had to test with some items that were totally outside the experience of its students. It was impossible to include any pure deductive Euclidean geometry. In spite of this, many general conclusions about geometry performance, also on gender differences, were made by researchers. (See chapter 3.)

The above-mentioned problems with geometry curricula in the United States and England may have influenced research results mentioned in this thesis. The fact that the USA limit their Euclidean geometry to a one-year course and that the pupils doing the course have different geometrical backgrounds in primary school implies that all the research results from the USA must be interpreted within these constraints. A similar situation prevails in England where there was no national control over syllabi.

Doubts raised due to a lack of Euclid

In the meantime there was growing concern for the lack of Euclidean knowledge. Students have a need for spatial orientation abilities (Howson and Kahane, 1986). Some other geometries, like the vector geometry that Dieudonné propagated, for instance, can only be interpreted meaningfully after the students have first become acquainted with geometry in another way (Freudenthal, 1973). Dieudonné has been strongly challenged by Rene Thom and Michael Atiyah at the International Conferences of Mathematics Education in 1972 and 1976.

> "Atiyah also argues for the re-introduction of Euclidean geometry at school level. He highlights the importance of the visual approach in mathematics which is nurtured through geometry." (Kapadia, 1985: 54).

> "... to eliminate elementary geometry to make room for calculus and linear algebra, has little to recommend it psychologically, because the algebraic objects (the symbols) are too poor semantically to permit themselves to be understood directly as is the case with a spatial figure." (Thom, 1973: 208).

At ICME 4 in 1980, the formal approach of Bourbaki and the lack of geometrical concepts were criticized by several speakers. Osserman (1983) was concerned about the primacy of the abstract, general and theoretical, over the concrete, particular, and applied. At the same congress Grunbaum said:

> "What is missing from our curricula in colleges as well as in high schools, is the accent on the many aspects of reality which are susceptible of meaningful geometric interpretation." (Grunbaum, 1983: 165-167).

Also at ICME 5 in 1984, it was mentioned that some countries regretted their decision to discontinue the teaching of Euclidean geometry. The action group for secondary schools however agreed that the approach must be less formal than in the past.

> "Euclidean geometry is a good model of our real world and it is important for the future study in mathematics. It is a good vehicle for the introduction or the extension of the concept of proof." (Nicholson, 1985: 42).

South Africa

The synthetic approach to Euclidean geometry has been included in the geometry curriculum for the last three years, sometimes four years, of secondary school since at least the turn of the century. Informal geometry has always preceded the formal deductive section for a year or more. The present geometry curriculum comprises 2 to 3 years informal geometry, i.e. experimental and intuitive geometry with emphasis on spatial and visual aspects and investigative methods of teaching. This leads to four years (inclusive of the last school year) of axiomatic deductive geometry known as the synthetic approach to Euclidean geometry.

In South Africa mathematics is compulsory until grade 9 (standard 7). Thereafter students may opt either to discontinue their mathematical studies or to continue until Grade 12 on one of three levels - higher, standard or lower grade. The last level is fairly unpopular and was not included in this study. In most schools students at grade 11 (standard 9) level are taught in the same mathematics classes, irrespective of higher or standard grade, because the content

is almost the same. Except for minor differences, which were avoided for this research, everybody does the same Euclidean geometry course in grade 11.

Curriculum problems hampering research elsewhere are thus almost non-existent in South Africa. The past mathematical experience of all the students in the sample for this research is theoretically the same. Factors influencing research done elsewhere, such as different degrees of exposure to mathematics and geometry, various approaches to the teaching of geometry, differences in geometrical content covered, and geometry that has been reduced to a few routine problems or simple calculations, could be avoided by the researcher.

Some opinions on the teaching of Euclidean geometry in South Africa

In South Africa the general opinion is that, in spite of some negative attitudes towards geometry, problems with the teaching thereof and inadequate teacher training, Euclidean geometry is still of substantial value in mathematics teaching (Baart 1992; de Villiers, 1991; 1990; Human et al., 1989; Oosthuizen et al., 1988; Malan, 1986; Nicholson, 1985; Snyman, 1984; 1982; Kriel, 1975; Malan, 1971).

> "While it is undoubtedly true that the old approach of Theorem/Proof/Problem in the teaching of Euclidean geometry can be boring in the extreme, it is equally true that removing it altogether also removes nearly all exposure to the discipline of rigorous deductive thinking." (Baart, 1992: 42).

After extensive research, Human et al., (1989) conclude that South Africa may be in an advantageous position with regard to the teaching of geometry because it held on to the traditional approach in the teaching of Euclidean geometry.

Malan (1971) says Euclidean geometry can serve as a medium to stimulate and exercise intellectual abilities and thinking skills. It provides pupils with opportunities to observe, compare, analyze and generalize, abstract and finally systematize. He says to condemn Euclidean geometry as a secondary course may be a case of 'fools rush in where angles fear to tread'.

After reporting on all the problems of countries with 'no Euclid' as mentioned at ICME 5, Nicholson (1985) has the following advice to South African teachers:

> "We must teach relationships, show the beauty of geometry and of proof wherever possible, show some of the powers of geometry, and very important, we must use geometry to teach problem solving." (Nicholson, 1985: 40).

(See also section 2.6.6.)

2.6.4 Teaching problems

King Ptolemy I once asked Euclid, 3 centuries BC, whether there was not a shorter way to learning geometry than through the 'Elements', to which Euclid replied that there is

"no royal road to geometry"

- implying thereby that geometry is no respecter of persons (Burton, 1985).

Usiskin (1987) claims that many of the current complaints in the USA can be traced to teachers taking a one-dimensional view of the understanding of geometrical concepts. The standard approach may be appropriate for understanding the mathematical system concept but it neglects the fact that geometry is related to the real world. Schuster (1975) says the overemphasis of axiomatics has resulted in underemphasizing applications, and furthermore, Farrell (1987) points out the detrimental effect of rigid teaching approaches with too much memorization. These complaints have the same underlying principle.

Another serious problem that teachers have in the USA is the lack of a systematized curriculum (Masingila, 1993). (See section 2.6.3.) It seems that the geometrical background of the teachers does not equip them sufficiently to handle the very vague guidelines of the curriculum successfully, compounded by the fact that all secondary geometry is taught in a one-year course.

Similarly, Snyman (1982) mentions the following problems encountered in the teaching of synthetic geometry in the senior secondary phase in South Africa:

* Pupils are not capable of formal deductive reasoning.
* Pupils have too little opportunity to experiment and explore.
* Poor teaching methods.
* The content is too difficult for the average pupil.

De Villiers (1991; 1990) feels strongly that teaching methods in South Africa overemphasize the verification function of proof. He claims that proof is also a means of explanation, systematization, discovery and communication. De Villiers (1988) suggests a serious misconception among teachers that experimental teaching methods are not an integral part of the teaching of formal Euclidean geometry.

Toumasis (1991) likewise thinks that the rigorous structure poses a problem in his country, Greece, where the subject of classical Euclidean geometry traditionally holds an important position. He also sees the lack of some predominant guidelines in solving geometrical problems as a dilemma, whereas other researchers tend to see this fact as an asset in view of problem solving potentialities. The main problems encountered in the literature will be discussed in more detail below.

Axiomatic deductive structure

> "The deductive structure of traditional geometry has just not been a didactical success. People today believe geometry failed because it was not deductive enough. In my opinion, the reason was rather that this deductivity was not taught as reinvention as Socrates did, but that it was imposed on the learner." (Freudenthal, 1973: 402).

Many educators agree that the structure of Euclidean geometry does not pose a problem, but that the teaching of the structure as a polished refined product does not appeal to pupils in that they are never made aware of the background and struggle that went with its development. Intuitive arguments, re-invention, methods of investigation and discovery are vital in order to get pupils actively involved in the formulation of definitions, axioms and

theorems (de Villiers, 1991; 1990; 1988; Human et al., 1989; Kriel, 1975; Schuster, 1975; van Hiele, 1973; Hull, 1969; van Hiele-Geldof, 1957).

Considering the above, it seems that Euclidean geometry would benefit from a teaching approach similar to Lakatos's quasi-empirical approach and also to the more recent constructivistic teaching approach.

It is interesting to note that the disadvantage of teaching geometry as a pre-fabricated whole, without accentuating the development there of, was already perceived at the beginning of the century.

> "It is doubtful if we have any other subject that does so much to bring to the front the danger of carelessness, of slovenly reasoning, of inaccuracy, and of forgetfulness as this science of geometry, which has been so polished and perfected as the centuries have gone on." (Smith, 1911, in Moritz, 1958: 303).

> "But what we must be aware of doing is to **begin** with the deductive development from definitions and axioms. New mathematics does not arise in this way, and existing mathematics should not be so presented. ... Let us remember that mathematics begins with questions, and only ends with axioms." (Hull, 1969: 30).

Several researchers support a **local** axiomatic approach instead of a global one (Human et al., 1989; Nicholson, 1985; Lang and Ruane, 1981; Kriel, 1975; Schuster, 1975).

> "The compromise I suggest here is that we make geometry locally axiomatic in contrast to the globally axiomatic treatment that is now given. That is, I believe that students as they grow to mathematical maturity would benefit more from smaller axiomatic systems within the subject of Euclidean geometry." (Schuster, 1975: 171).

The research reported in this thesis was based on the circle theorems in the grade 11 (standard 9) South African Mathematics core syllabus. Pupils in grade 11 should be able to reason in a global axiomatic manner. The interrelationships between the circle theorems can, however, be approached in a local axiomatic way.

Between approximately the ages of 13 and 15, pupils pass from the stage called concrete operational thinking when their thinking is almost completely related to their perceptions, to a stage of formal operational thinking, when they begin to think hypothetically and deductively (Piaget, 1970). Toumasis (1991) reckons that the age at which deductive reasoning is introduced is one of the problems encountered in Greece. In Japan, on the other hand, pupils are successfully introduced to deductive reasoning in grade 8 (Miwa, 1992). The sample for this research was drawn from pupils in their eleventh school year, i.e. at an average age of about 16 to 17 years. They started with deductive geometry two years prior and that was preceded by 2 to 3 years of informal experimental geometry.

According to de Villiers (1987) and also Human et al. (1989) some teachers in South Africa, although well qualified, have insufficient knowledge of an axiomatic structure. Many are still convinced that axioms are 'self-evident' truths. They also support the formalistic approach to the teaching of mathematics.

Proof

Proof is described by Driscoll (1982) as a process of reasoning from a set of premises through a series of connected inferences to a conclusion, in such a way that any doubt about the conclusion must be referred to the premises, rather than to the logical nature of the inferences. Less technically, it is 'an argument that convinces' (Hersh, 1993), or a 'form of discourse' (Volmink, 1988). Bell (1978) claims that teaching students how to prove a theorem is not unlike teaching them how to think.

Teachers with a rigid approach to geometry often overemphasize the role of proof. Hoffer (1981) asserts that geometry is a visual subject, but all too often the visual aspects have served only as a 'tool for proofs'.

Bell (1976: 24) distinguishes the following three functions of proof:

* 'Verification', which is concerned with establishing the truth of a proposition.
* 'Illumination', which is concerned with why a certain proposition is true. De Villiers (1987) calls this the function of explanation.
* 'Systematization', i.e., the organization of results into a deductive system of axioms.

According to de Villiers (1987) these three are actually synonymous from a strictly mathematical point of view, but from a didactical point of view the distinction is important for teaching purposes.

The crux of the problem with proof lies in the teaching thereof. Concentrating on the verification function of proof has contributed in the past to the concept of an over-rigorous structure. De Villiers (1987) claims that the majority of teachers emphasize only the verification function of proof and that such teaching methods produce negative attitudes in pupils. Hanna (1983) also supports this view. She argues that results are presented formally by means of theorems and proofs and that this rigid practice is seen mistakenly by many students as the core of mathematical practice. It seems that the approach to convince students of a mathematical truth by supplying them with a deductive proof has failed. Thom (1971) says the following about proof:

> "... each proof is, above all, the revelation of a new structure whose elements lie disconnected in man's intuition until reason joins them together. In this sense, each proof is a Socratic experience requiring the re-creation in the reader of the psychological processes necessary to elicit the implicit truth ... " (Thom, 1971: 697).

Hersh (1993) underscores this idea as follows:

> "The role of proof in the classroom is different from its role in research. In research its role is to convince. ... the educational value of proof is the value of complete explanation." (Hersh, 1993: 398).

De Villiers (1991; 1990) investigated pupils' cognitive needs for conviction and explanation within the context of Euclidean geometry in South Africa. He found that the need for personal conviction of the majority of pupils is satisfied by quasi-empirical means e.g. construction and measurement. Many of them, nevertheless, showed an independent need for explanation not satisfied by experimental means. In cases like these, where a high level of conviction already exists, it is important to stress the explanatory function of proof to make proof more meaningful to pupils. De Villiers warns that this is not a plea to ignore the verification function of proof, but only that the introduction to deductive geometry should be less formal and that proof is not a prerequisite for conviction. Other functions of proof such as discovery, communication and systematization should also receive due attention.

Ignoring the explanatory function of proof, especially in the introductory phases of Euclidean geometry where conviction is easily reached by experimental methods, has led to the criticism of 'proving the obvious' (Human et al., 1989; Kriel, 1975; Hull, 1969). De Villiers (1991) calls it 'intellectual dishonesty'. According to Bell (1978) students must have a genuine doubt of the truth of a proposition and must want to resolve this doubt in order to experience a proof as meaningful.

In geometry, empirical (inductive, experimental, investigative) and deductive methods should interact and reinforce each other (Schumann and de Villiers, 1993; Clements and Battista, 1992; de Villiers, 1991; 1990; Human et al., 1989; Schoenfeld, 1985; Driscoll, 1982; Bell, 1978; Kriel, 1975; Hull, 1969). Driscoll (1982) stresses this by saying that formal proof is usually a series of statements. Skills in proving, however, are born in the 'asking of questions'.

Care should be taken not to leave pupils with the impression that empirical and deductive methods are two separate domains, with different ways of arriving at the truth (Clements and Battista, 1992). An interaction between the empirical framework and the deductive framework of a pupil often leads to flexible geometrical thinking and creative problem solving in geometry (de Villiers, 1988; Schoenfeld, 1985). Lately, this empirical framework is also developed by computer methods.

Chazan (1993) and also Schumann and de Villiers (1993) investigated the role of computer software in the teaching of geometry. Schumann and de Villiers (1993) emphasize the convincing power of the computer as well as the opportunity to re-discover theorems. They used two software packages, namely, 'Cabri-géomètre' and 'Geometer's sketchpad'. Chazan (1993) says his conclusions on empirical conviction of theorems, using the computer program 'geometric supposer', is valid for paper and pencil methods also. He was curious to find out whether extensive use of measurement would hinder students' appreciation of mathematical proof, but concludes that this was not the case. Comparison and contrast of deductive proof and of empirical verification, either by computer or other methods, deserves attention in the classroom.

Memorizing the proofs of theorems should be discouraged (Schiddell and Ethington, 1994; Breen, 1992b; Farrell, 1987; Driscoll, 1982; Hoffer, 1981; Giles, 1982; Bell, 1978; Kriel, 1975; Malan, 1967) because it produces negative attitudes towards geometry (Malan, 1967) and creates the impression that memorizing is a measure of mathematical attainment (Giles, 1982). If the focus in the classroom is on memorization, the students lose sight of the underlying principles (Schoenfeld, 1985), but proving theorems can be interesting if approached in a challenging way (Bell, 1978).

The performance of pupils in writing proofs is also discussed in section 3.3.1.5 where it is considered in relation to gender differences.

The core syllabus intended for implementation in 1992/93 in South Africa (at the moment frozen due to the political situation) recommends the introduction of *proof* within a

systematization context in grade 10 (standard 8). The pupils in the sample for this study had been introduced to formal deductive proof in their grade 9 year.

Diagrams

Visual processes (Hershkowitz et al., 1990) and diagrams (Kabanova-Meller, 1950) have a complex role in the forming of geometrical concepts and can contribute in both positive and negative ways. The mastery of geometrical concepts by means of a diagram in only one position, usually the textbook diagram, leads pupils to be constrained by this specific diagram. When it comes to application of concepts, pupils tend to go by the visual appearance of the diagram because they have only one visual prototype. Hoz (1981) refers to this phenomena as one form of 'geometrical rigidity'. The Russian researchers, Zykova (1954) and Kabanova-Meller (1950) both mention their pupils' inability to change a mental image of a theorem while still retaining the essential aspects of the diagram.

Some think it is the role of the teacher to encourage flexible visual images by avoiding diagrams in standard positions only (Presmeg, 1993; Zykova, 1954). Visual flexibility can also be enhanced with the aid of computer software, such as Cabri Géomètrie, in which a geometrical figure has a status similar to the status of a variable - changing but retaining all its relevant properties (Hershkowitz et al., 1990). Brown and Presmeg (1993) refer to it as 'concrete imagery' (image of a single, non-moving picture) and 'dynamic imagery' which involves the ability to transform a concrete visual image. They found that weaker students tend to rely on concrete images in mathematics, whereas the students that used dynamic imagery tend to have better relational understanding. (See also sections 5.7.2.3 and 5.7.2.4.)

Chazan (1993) questions the tendency of textbooks to make geometry problems less complicated by supplying the diagram complete with a simplified version of the 'given' and 'to prove'. He argues that such a format obscures the generality of the claim and gives students no indication that the problem presented is not an example of a single case. All the above arguments may lead to pupils' inability to generalize a single geometrical concept.

For the present research the diagrams were supplied because it is normal teaching practice and facilitated staying within the time limits of the examination.

Teachers

With appropriate guidance, secondary students can learn to think critically, reflectively, hypothetically and deductively (Driscoll, 1982). The accent here is clearly on guidance and that is one of the reasons why the role of the teacher determines the outcome in the learning of Euclidean geometry (Senk, 1989; Malan, 1971; Zykova, 1954). Kriel (1975) suggests that the attitude of the teacher towards geometry and the spirit of the teaching is crucial. According to the van Hiele theory (see section 2.6.5), the progress of pupils from one level of geometrical thinking to a higher level depends on the teaching process and the teacher's role is to guide pupils to higher levels of geometric thinking. In the Soviet Union, where they had great success in the teaching of Euclidean geometry, Zykova (1954) investigated the role of the teacher in the geometry classroom and concluded:

> "Thus, the best way of teaching geometric concepts is by organizing the instructional process in such a way that the teacher's explanations play the leading, organizing role." (Zykova, 1954: 186).

In view of the importance of the teacher in the teaching of geometry the number of problems mentioned in the literature concerning teachers of geometry, is alarming. In the report on ICME 5 that Nicholson (1985) compiled for the Mathematical Association of South Africa, he mentions the fact that many secondary teachers in other countries have no formal Euclidean geometry in their background.

In Canada a sample of 60 geometry teachers randomly selected from high schools have a completely different view of the important objectives of high school geometry compared to educators at universities. The teachers tend to feel that knowledge and comprehension objectives are important, while educators favour analysis and synthesis objectives (Pereira-Mendoza and Robbins, 1977).

In the USA, Mayberry (1983) found that 70% of a group of 19 pre-service elementary teachers were not ready for a formal deductive geometry course. In the same vein Usiskin (1987) reports that some elementary school teachers have not taken any geometry since high school, and some not even then. Even teachers qualified in mathematics may not possess enough background to handle a geometry syllabus because college mathematics departments have cut back on the teaching of geometry. Masingila (1993) blames the lack of teacher training for the 'lack of evolution' in the teaching of geometry in spite of recommendations to revise secondary geometry in the USA. Also Schiddell and Ethington (1994) object to the continued commitment of teachers to traditional teaching methods.

In South Africa a similar situation exists. Problems with the teaching of geometry and the poor performance of pupils are sometimes ascribed to inadequate teachers' education (Human et al., 1989; de Villiers, 1987). Teacher education at colleges and universities covers no, or very little, deductive geometry. According to de Villiers (1987), 56% of a group of prospective mathematics teachers had the narrow, classic concept of axioms as self-evident truths as opposed to the more modern view. About 43% of the same group of students were not successful in hierarchical thinking. This kind of thinking is a prerequisite to a meaningful conception of an axiomatic system (van Hiele, 1973).

Some of the problems in the teaching of Euclidean geometry in South Africa go hand in hand with curriculum changes. Presmeg (1992) argues that a curriculum is more than just the syllabus for that subject.

> "Curriculum encompasses aims, content, methods of teaching and assessment procedures." (Presmeg, 1992: 79).

In South Africa the geometry curriculum has undergone quite a few changes during the past two decades. Analytical geometry was replaced by vector geometry for a while, but since 1987 it has again been taught. With regard to Euclidean geometry, formal deductive reasoning was started in grade 10, then shifted to grade 9, and now back to grade 10 in some schools. The informal part of the geometry curriculum, in particular, has changed quite dramatically.

Teachers are often presented with the finished product and are expected to teach the new syllabus, sometimes without attending any in-service courses, or having training spread over one or two afternoons. As a result, there is inevitable controversy as to the merits and reasons for the inclusion or exclusion of a particular topic (Breen, 1992a). As regards Euclidean geometry, many of the present teachers have a fairly negative attitude based on their own experience of traditional, rigid teaching methods. Their training probably did not expose them to any form of informal geometry and they therefore show resistance to teaching it that way. Without sufficient exposure to informal activities that serve as a 'substructure' (Breen, 1992b) on which the formalization can be based, pupils often experience problems when first introduced to deductive reasoning.

The teachers who participated in the empirical research for this thesis, were not evaluated, but they function in departments acknowledged for sound mathematical teaching methods, generally. They agree, however, that limited time does not allow for sufficient informal experimental activities while teaching formal secondary Euclidean geometry. De Villiers (1988) regrets this situation.

Affective domain

According to the definition of the affective domain (see section 1.5.4), attitude is one of the subsets of the affective domain. Attitude is defined by Aiken as:

> "a learned predisposition or tendency on the part of the individual to respond positively or negatively to some object, situation, concept or other person." (Aiken, 1970: 551).

Attitude towards mathematics, and perhaps geometry also, is influenced by a wide range of factors, including maths-anxiety, fear of success, motivation, self-concept and confidence. Negative attitudes towards geometry are widely mentioned in the literature (Masingila, 1993; Clements and Battista, 1992; Senk, 1989; Farrell, 1987; Reddy, 1987; Driscoll, 1982; Hoffer, 1981), but little research could be found on Euclidean geometry specifically. Nyangeni (1993) found a positive correlation between Transkei pupils' attitude towards

Euclidean geometry and their geometry scores. In 1991, Cronjé showed that negative attitudes towards Euclidean geometry can be improved by using discovery methods and by also integrating the history of geometry into the teaching. In the USA a study on attitude towards geometry was done by Earle (1972), comparing traditional teaching methods with methods using computers. He found no difference in the attitudes of the two groups experiencing different teaching methods.

As far as mathematics in general is concerned, affective aspects play a role in performance of pupils (Frost et al., 1994; McLeod, 1992; Garafolo, 1989; Pretorius, 1989; McLeod, 1988; Schoenfeld, 1985; Reyes, 1984). See section 3.1 for interesting results concerning Japanese pupils. The present research concentrated only on cognitive factors. No affective evaluation was done. Gender differences in the affective domain related to this study, are discussed in section 3.3.2.

2.6.5 Research on the teaching of Euclidean geometry

Due to the unpopularity of the **synthetic** approach to Euclidean geometry during the past few decades, there is less research available in this field than in other fields of mathematics education. In Holland a long term project was undertaken by the van Hiele family (van Hiele, 1973; van Hiele-Geldof, 1957), and the translation of their work into English inspired more research, mainly in the USA. Their findings also had an influence on the teaching of geometry in Russia. In South Africa important research was done by Human et al. (1989).

The van Hiele research

The van Hiele model of geometric thought evolved from the doctoral works of Pierce van Hiele and Dina van Hiele-Geldof (1957) which were completed at the same time in Holland. With the exception of the Soviet Union, where the geometry curriculum was revised in the 1960s to conform to the van Hiele model, the work was slow in gaining international recognition. Izaak Wirszup began writing about the model in 1976 (Crowley, 1987). Some of their major works were translated into English in 1984, and since then interest has spread through the USA, parts of Europe and to South Africa (Clements and Battista, 1992; Teppo,

1991; Gutiérrez, 1991; Toumasis, 1991; Wilson, 1990; Usiskin and Senk, 1990; Smith and de Villiers, 1990; Crowley, 1990; 1987; Human et al., 1989; Senk, 1989; de Villiers, 1986; Burger and Shaughnessy, 1986; Mayberry, 1983). Despite minor shortcomings, the above writers used the model for research or acknowledged the value of it. Volmink (1988) considers it as probably the most pervasive model for research into geometry learning, teaching and curriculum.

The theory suggests that students progress through five levels of understanding geometry. These levels are sequential and hierarchical and, according to van Hiele, also discrete, although the latter has been challenged by some researchers (Burger and Shaughnessy, 1986; Gutiérrez et al., 1991). The levels are:

Level 1 - Recognition (visual)

The student can recognize a shape as a whole, e.g. a square or parallel lines, but is not aware of the properties of the shapes.

Level 2 - Analysis

The student analyzes properties of figures, but they are isolated and unrelated. At this level a student may realize that the opposite sides of a rectangle are parallel and equal but will not notice how rectangles relate to squares. Smith and de Villiers (1990) mention that students on this level are partition-thinkers because they still concentrate on the distinctive properties of geometrical figures. To reach this level of thinking it is important for pupils to involve in informal activities such as cutting, measuring and drawing (Clements and Battista, 1992; Hoffer, 1981).

Level 3 - Ordering

Students can logically order geometric figures and understand the interrelationships between figures. They know that all squares are rectangles but not all rectangles are squares. Students understand the importance of correct definitions. Deductive thinking skills are not fully developed at this level, but begin to take shape (Driscoll, 1982). Burger and Shaughnessy (1986) suggest that a student may be able to form correct informal deductive arguments, such as using the chain rule, on this level.

Level 4 - Deduction

Students understand the significance of deduction and the role of axioms, theorems and proof and can also think hypothetically.

Level 5 - Rigour

Students understand the importance of precision in dealing with foundations and can compare the different axiomatic systems of Euclidean and non-Euclidean geometry. The secondary school syllabus in South Africa does not make provision for reaching this level. Usiskin (1982) concluded that this level is either not testable or it does not exist.

The van Hiele theory has important implications for the teaching of Euclidean geometry. The theory criticizes the direct teaching of a formal system as a pre-fabricated, polished product without pupils being actively involved in developing the system by gradually going through the different levels of understanding (van Hiele, 1973). For students to function adequately at a higher level they must have mastered large portions of the lower levels. Progress from one level to the next level is highly dependent on instruction (Teppo, 1991), more so than on age or biological maturation (Driscoll, 1982). Senk (1989) thus concludes that much of a student's achievement in writing geometry proofs, i.e., reaching level 4, is due to factors within the direct control of the teacher and the curriculum. This view supports the ideas of Kriel (1975) and Malan (1971).

Students pass from one level to another at different rates. Thus, van Hiele asserts that in any given classroom one might find the teacher, text and students functioning at different levels of thinking (Senk, 1989). Mayberry (1983) also found that students were on different levels for different concepts in geometry. Exposure to formal geometry does not ensure progress beyond levels 1 and 2 (De Villiers, 1987; Burger and Shaughnessy, 1986; Mayberry, 1983). These are factors that definitely challenges the teachers' skills.

The van Hiele levels appear to be dynamic and continuous rather than static (Burger and Shaughnessy, 1986), and continuity means that the acquisition of a specific level can take several months (Gutiérrez et al., 1991). Senk (1989) found that students' achievement in

writing geometry proofs is positively related to van Hiele levels of geometric thought, which supports Usiskin and Senk's view (1990) that the original test, compiled in 1982 by the Cognitive Development and Achievement in Secondary School Geometry Project (CDASSG) is a good predictor of ability to write proofs. Crowley (1990) and Wilson (1990) question the reliability of the CDASSG test developed to 'assign students to a van Hiele-level' , but Usiskin and Senk (1990) explain that the purpose of the original test was only to test the van Hiele theory. To Usiskin's knowledge, the only other test to measure van Hiele levels was developed by the research unit of the University of Stellenbosch (RUMEUS). The RUMEUS test outperformed the CDASSG test, but it is much longer and takes longer to assess because of open ended questions.

In South Africa the van Hiele theory is criticized for having a rather narrow and unitary concept of proof in geometry (Human et al., 1989; Volmink, 1988; de Villiers, 1986). See the three functions of proof described in section 2.6.4, whereas van Hiele only emphasizes the one function of systematizing. Malan (1986) has doubts concerning the third level and Smith and de Villiers (1990) question the hierarchical nature of the levels of thinking.

In Russia, great success was obtained in geometry teaching after the curriculum was revised in 1960 to a model based on the van Hiele theory. Pupils are exposed to inductive geometry from grades 1 to 4 whereafter they start with deductive geometry (Human et al., 1989; Malan 1986; Shaughnessy and Burger, 1985). Krutetskii (1976) says that with special instruction younger children acquire much greater ability for abstraction and reasoning than is generally thought.

Van Hiele currently describes his model in terms of three (Clements and Battista, 1992; Teppo, 1991; Hershkowitz et al., 1990) levels of thought, but the basic principles remain the same.

South Africa

In 1989, extensive research on the teaching of geometry was undertaken by the research unit of the University of Stellenbosch (RUMEUS). They made a thorough study of the development of geometry teaching in Western countries as well as the Soviet Union. They argue for a reconstructive approach to Euclidean geometry which give priority to active participation of students in the **development** of definitions, axioms and proofs in order to better understand the modern view of an axiomatic system.

They suggest an extensive informal introduction, based on the van Hiele model, before commencing the deductive system during the second half of the grade ten year. Proof during the informal phase should be restricted to situations where pupils show a definite need for it, and then rather in an explanatory sense, than in the sense of formal proof. RUMEUS also recommends a training program for teachers that includes geometrical deductive reasoning (Human et al., 1989). The RUMEUS study did not look at gender differences in the performance of geometry.

Other Research

The most important research for the purpose of this study, from the USA and elsewhere, has been mentioned in section 2.6.4. Research on Euclidean geometry in the USA is influenced by curriculum problems referred to section 2.6.3, and to a great extent by the fact that all formal geometry is squashed into a one year course which is not preceded by sufficient informal geometry. Since the Second International Mathematics Study (SIMS) where the Japanese pupils outperformed most other pupils from western countries, a few articles (see section 2.6.3) involving Japanese pupils, have been published.

(See section 3.3.1.5 for research related to geometry and gender differences.)

2.6.6 Merits of teaching Euclidean geometry

The educational merits of the teaching of Euclidean geometry were indicated earlier in the chapter. It is advisable to consider them in more detail.

Historical motivation

No other branch of human knowledge has played such a dominant role in the development of civilization as Euclidean geometry. The model of Euclidean geometry was seen as the model of all human knowledge for more than two thousand years. For historical reasons pupils should be introduced to the subject that remained uniquely unchanged for so many centuries.

> "We study it because we derive pleasure from contact with a great and ancient body of learning that has occupied the attention of master minds during the thousands of years in which it has been perfected, and we are uplifted by it." (Smith, 1911 in Moritz, 1958: 305).

Pedagogical value of structure

> "Deductive argument is the central feature of mathematics and those who have never encountered it can barely claim to have encountered mathematics. Geometry is a particular good context for deductive proof since many results are by no means intuitively obvious but rarely do they contradict intuition." (Lang and Ruane, 1981: 128).

Polya (1957) in answer to the question, 'why proofs ?' said that

> "If the student failed to get acquainted with this or that particular geometric fact, he did not miss so much; he may have little use for such facts in later life. But if he failed to get acquainted with geometric proofs, he misses the

best opportunity to acquire the idea of strict reasoning." (Polya, 1957: 216-217).

Problem Solving

Euclidean geometry can be used to develop problem solving strategies, since the student can easily be confronted with an unknown situation for which a variety of ideas and strategies are necessary (Nicholson, 1985). Farrell (1987) agrees that Euclidean geometry is a useful source of problems.

> "Geometry, with its rich tradition of classical problems and contemporary usefulness in terms of mathematical models, seems to be especially suited to problem solving activities. It appears that the understanding of geometry is deepened as college students interact with each other in analysing constructions, devising proofs, or struggling to find a best-fit geometric model for a problem situation." (Farrell, 1987: 241).

No algorithm exists that will solve all geometry problems (Senk and Usiskin, 1983; Bell, 1978) so it lends itself to the evaluation of creative thinking and problem solving abilities (Farrell, 1987; Schoenfeld, 1985; Snyman, 1984; 1982; Bell, 1978). The solving of geometry riders can also facilitate intellectual development (Bell, 1978; Thom, 1971; Malan, 1971). Solving geometry problems is a very individualized activity and many students obtain a considerable amount of satisfaction from producing a piece of original work. The vast potential of Euclidean geometry as an intrinsic motivator is emphasised by Posamentier (1989). Thom (1971) says in order to judge the capabilities of a student fully, it is necessary to place him in an active role and to call on his individual initiative and enterprising spirit.

2.7 SUMMARY

The philosophy of the Formalists dominated the mathematical world during the first few decades of this century. Their view of mathematics as a fixed and unchangeable discipline

influenced mathematics education and resulted in teaching practices with similar aims. Euclidean geometry was an obvious victim of such practice, because if it is taught in a rigid way, divorced from the physical world, it easily looses its appeal and educational value. Euclidean geometry disappeared from many countries' curricula, and the lack thereof is often seen as a shortcoming in the mathematical background of students. South Africa is one of the few countries that still follows the synthetic or Euclidean approach to the teaching of geometry, although not without criticism from educators and students. Most of the criticism is aimed at a too formal and rigid teaching practice.

New doors may open to the teaching of Euclidean geometry with the current constructivistic approach with emphasis on investigative work, re-creation of principles, active participation by pupils and social negotiation. In such a milieu, Euclidean geometry will flourish, because in Thom's (1973) words, 'Euclidean geometry is that terrain which is an ideal apprenticeship for investigation and an inexhaustible mine of exercises'.

Bell concludes along more general lines:

> "We also want our students to develop respect and appreciation for logical deductive arguments and to value and apply these argument forms after they finish their high school education. People who know how to use valid arguments are better able to participate in democratic institutions, can avoid being deceived by invalid arguments, and are better able to assist in solving some of the difficult problems of society." (Bell, 1978: 304).

CHAPTER 3

GENDER DIFFERENCES

"It seems to me that the thing that is wanting in the education of women is not the acquaintance with any facts, but accurate and scientific habits of thought, and the courage to think that true which appears unlikely. And for supplying this want there is a special advantage in geometry, namely that it does not require study of a physically laborious kind but rather that rapid intuition which women possess; so that it is fit to become a scientific pursuit for them." (Clifford, 1901, in Moritz, 1958: 302).

3.1 INTRODUCTION

The above quote shows that, at the beginning of this century, the lack of geometry in women's education was linked to a lack of accurate and scientific habits of thought. Today, mathematics, of which geometry forms an integral part, is still associated with accurate and scientific thought processes, as well as with the extent to which it has become a gateway to job and career opportunities. Differences in schooling, training and careers do exist between males and females across the world. There is widespread concern about this matter in industrialized countries as well as in developing ones.

Any signs of under-achievement or under-participation of females in the field of mathematics is thus of importance. Figures from international studies are alarming. Garden (1989) reports that in the Second International Mathematics Study (SIMS), the boys enrolled for mathematics outnumbered the girls by a significant margin in all countries except two. Most countries have samples (based on enrolment) representative of both genders at the age of 13, but at the age of 18 the samples are generally highly selective. In Belgium (France) only 34% of the sample of students in the last year of secondary mathematics, was female; in

Hong Kong and Japan the figures were as low as 21% and 22% females respectively. In the United States the percentage female students was 44%. Garden (1989) warns that these results can lead to misinterpretation of international studies with regard to gender differences in mathematics achievement because in most countries a substantial degree of selection by gender has occurred by the end of secondary school.

In secondary schools (including the 5 schools that participated in the present research) of the Transvaal Education Department (TED), South Africa, less selection by gender has taken place at school leaving age. The TED population (a total of 35 000 pupils) that wrote their school leaving certificate examination (grade 12) in November 1994 consisted of 49,5% males and 50,5% females, while pupils doing mathematics on higher or standard grade (see section 5.3) made up 63% of the total number of students. This mathematics group comprised 52% males and 48% females. The situation is thus favourable for gender research in the field of mathematics, and especially for Euclidean geometry which is a compulsory component of mathematics up to grade 12 level. The sample for the present study consists of almost an equal number of grade 11 males and females.

The concern regarding differential gender performance in mathematics is reflected in the amount of literature covering the topic. As early as a century ago, differential performance in favour of males were observed in examination papers on Euclid and algebra (Leder, 1992). Over the past three decades there has been extensive research on gender differences in mathematics achievement. A variety of factors have been investigated. It is interesting to note that the popular belief, that positive attitudes towards mathematics often result in better performance, is tempered by the results of the Second International Study of Mathematics (SIMS - 1981/2). Japanese pupils who generally came first or second out of 15 countries on all the achievement tests in mathematics, also had the most **negative** attitudes towards mathematics. They also considered mathematics to be more difficult than any other participating country (Kifer and Robitaille, 1989). Although cultural factors must be taken into account, it seems that at least for Japan, neither attitude nor achievement is dependent on the other. They rather interact with each other in a complex way.

Conflicting results on gender differences in mathematics have been found due to a variety of factors that will be discussed. International studies have also shown a considerable amount of variance between the results of different countries. A great deal of the existing research is of a very general nature, combining divergent content areas of mathematics, using samples with inherent limitations and usually comparing only average performance. Quite a few researchers in the field have suggested research of a more specific nature. A further problem encountered when analysing the literature, is that often no clear distinction is made between achievement tests and ability tests in the field of mathematics and of spatial skills. Sometimes the results of the two types are compared and sometimes they are combined, although different in nature.

The popular belief is that, where differences exist, they occur from adolescence onwards and favour boys in achievement and participation in mathematics. Differences in favour of males, tend to be more pronounced in the field of spatial ability than in other areas of mathematics. Sometimes girls outperform boys in computational skills.

The matter of gender differences in mathematics achievement is complex, and can be explored from more than one perspective. The present research concentrates on **cognitive** gender differences in the field of Euclidean geometry. Gender differences related to mathematics, spatial ability and Euclidean geometry are considered, because gender research in the field of pure Euclidean geometry is rare in western countries. (See also sections 2.6.3 and 2.6.5.) Other factors influencing achievement in mathematics, such as the affective domain, educational and biological factors will be discussed briefly. It is important to keep in mind that the variety of 'other' factors can only be identified once the cognitive structures have been established. Some literature emphasizes the 'other' factors to such an extent that the cognitive domain, and especially the evaluation thereof, is given little attention.

The present study gives special attention to the comparative evaluation of boys' and girls' ability and achievement in Euclidean geometry. A variety of open ended questions are used as well as an in-depth analysis of the data obtained from multiple choice items. (See section 3.3.1.3 on aptitude and achievement and section 5.4 on the test material.)

3.2 MATHEMATICS - GENDER DIFFERENCES IN YOUNG CHILDREN

3.2.1 Cognitive domain

Cognitive gender differences in mathematics have been researched at all age levels for school pupils. In primary school and sometimes middle school, i.e. up to approximately the age of 12 to 13 years, gender differences are rare. The most common finding is that young boys and girls achieve equally well (Hyde et al., 1990; Ethington, 1990; Noss, 1987; Curio, 1987; Hanna, 1986; Visser, 1987; 1985; Armstrong, 1981; Fennema and Sherman, 1978). Fennema and Sherman (1978) compared mathematical tasks at several levels of cognitive complexity, e.g. computational skill, knowledge of concepts and problem solving as well as two other cognitive factors, verbal ability and spatial visualization. They reported no sex-related differences for their sample of grades 6-8 pupils. Noss (1987) and Curio (1987) could not pick up any gender differences in problem solving or graph comprehension for seventh graders.

If gender differences are found for young children, they tend to favour girls in computation (Hyde et al., 1990; Armstrong, 1981) knowledge and skills (Fennema and Carpenter, 1981) and sometimes spatial visualization (Armstrong, 1981). When the differences favour boys, it is in the areas of understanding and application (Fennema and Carpenter, 1981) and problem solving (Fennema and Tartre, 1985). The latter case used only a small sample. Using a national sample in the USA (Armstrong, 1981), and using meta-analytical procedures (Hyde et al., 1990), it was established that boys and girls can solve problems equally well at the age of about 13 years. Recently (Zambo and Follman, 1994) it was found that young girls of grades 6 and 8 significantly outperformed boys on problem solving. The study showed no evidence of gender related differences in the processes of solving a problem. Due to the high verbal content of this study, it is difficult to compare the results to those of other studies which usually use problems of both spatial and verbal content.

Visser (1987; 1985) found no gender differences in the mathematics achievement of South African grade 7 pupils.

From the studies mentioned above, it is seems that even though contradictions sometimes appear, gender differences in mathematics performance of young children are seldom found and, when they are found, they are relatively small.

3.2.2 Affective domain

The **affective** domain shows few gender-related differences at an early age. Ernest (in Meece et al., 1982) says that during the elementary school years, members of each sex believe that children of their own sex do better in mathematics. There is however a great deal of evidence that by grade 6, girls express less confidence in their ability to do mathematics (Meece et al., 1982) before there is any evidence that they are actually doing mathematics any more poorly than males (McCarthy, 1992; Fennema and Sherman, 1978).

Differences in the affective domain seem to appear at an earlier age than differences in the cognitive domain with regard to mathematics. The differences are almost always to the disadvantage of girls.

3.3 MATHEMATICS - GENDER DIFFERENCES FROM EARLY ADOLESCENCE

The area of early adolescence may overlap with the age groups discussed under the heading of young children. It is **not** the purpose of this study to examine **age** related differences in mathematics performance.

3.3.1 Cognitive domain

In their Taxonomy of educational objectives, Bloom et al. (1956) state that the cognitive domain includes those objectives which deal with the recall or recognition of knowledge and the development of intellectual abilities and skills. For the purpose of investigating gender differences the cognitive domain is usually subdivided into verbal, mathematical and spatial domains. The present research is in the field of Euclidean geometry and relates to the mathematical and spatial domains. The verbal domain is not excluded but the nature of the

geometry questions in the present research is such that verbal ability is not as prominent as mathematical and spatial abilities.

3.3.1.1 Earlier research

In their influential review of literature on sex differences, Maccoby and Jacklin (1974: 351 - 352) reported

* a gender difference in mathematics achievement favouring boys from the age of 13 onwards;
* male superiority in spatial abilities from adolescence onwards;
* an increase during high school years in the male advantage on spatial abilities, and
* a gender difference in the favour of girls from the age of 11 in verbal ability.

Because of the implications of their findings, supported by earlier research (Anastasi, 1958), and wide press coverage at the time, gender differences in these abilities have been well researched and analyzed.

Fennema and Sherman (1977) noted an important shortcoming in gender research on mathematical abilities. Most studies in the USA up to that date, for example, had not controlled one of the most important variables in achievement in high school mathematics, viz. the previous study of mathematics. Fennema and Sherman (1977) found that when students with similar mathematical backgrounds were compared, differences between male and female groups in mathematics achievement were small and were significant in only two of the four schools. Where differences do occur, they suggest the influence of socio-cultural factors.

However, confirmation for Maccoby and Jacklin's findings of a male advantage in mathematical and spatial abilities was given by Hyde (1981) in the form of a meta-analysis, i.e., a statistical procedure for doing a literature review. (See section 3.1.1.3 for meta-analysis.) At that time it was realized by researchers that the problem of gender differences

in mathematics achievement is more complex than initially thought. Hyde's statistical analysis was challenged by researchers because age groups were not separated and because of the selectivity of some of the samples on the grounds of intellectual level (Feingold, 1993; 1988). (See section 3.3.1.3 for gender differences on higher intellectual levels.)

Both problems, that of differential gender exposure to mathematics and the intellectual levels of the sample, are addressed in the present research.

3.3.1.2 Main theme from the eighties

The main theme that characterizes research on gender differences in mathematics achievement reported from the 1980 onwards, is one of **contradiction.** In 1981, with the National Assessment of Educational Progress in the USA (Fennema and Carpenter, 1981) it was found that at the age of 17 years, males' average performance in mathematics exceeded that of females at every cognitive level and that the differences in achievement were greatest on the more complex tasks. These results were confirmed for problem solving by Armstrong (1981).

In the same year that Hertzog and Carter (1982) concluded that males and females have similar intellectual structures of verbal and spatial abilities, Benbow and Stanley (1982) published their controversial finding of large and consistent gender differences in mathematical reasoning (as measured by the Scholastic Aptitude Test - SAT) in favour of gifted boys. Better grades were, however, obtained by girls in the sample of Benbow and Stanley (1982), by girls in grades 9 and 11 of the New York State Regents secondary mathematics examination (Smith and Walker, 1988), by Hawaiian girls in grade 10 (Brandon et al., 1987), and by West Indian girls living in England in the school leaving examination (Driver, 1980). (See section 3.3.1.3 for aptitude tests and classroom tests.)

For college students, a male advantage was reported in problem solving (Johnson, 1984) and in mathematical reasoning (Doolittle and Cleary, 1987). Females found algorithmic computation orientated items (Doolittle and Cleary, 1987) and calculus (Ferrini-Mundy, 1987) easier than males.

Analysing data up to 1980 for grades 8 to 12 on the Differential Aptitude Test, Feingold (1988) reported no notable gender differences in the subtests of abstract reasoning and numerical ability, but a male advantage in space relations. Lynn (1992) mentions a close degree of similarity between gender differences in these abilities in Britain and in the USA.

In South Africa, one large scale study undertaken during the eighties on gender differences in mathematics achievement, relates to the present research. Visser (1985) investigated the mathematics achievement of boys and girls both on standardized aptitude tests and achievement tests. At grade 11 level, boys showed a significant advantage in spatial ability but not in mathematics achievement. When the top 100 performers were considered, boys achieved better than girls. The reliability of the achievement test was rather low (0,61), probably due to a too high level of difficulty.

Analysis according to mathematics **content** of tests is not always successful because so many studies fail to report the mathematics content or use tests with a mixture of content. Sometimes the content of tests is kept confidential. A common finding is that girls tend to do better on tests of computation and boys do better on mathematical reasoning and problem solving (Kimball, 1989; Feingold, 1988; Fennema and Sherman, 1978; Doolittle and Cleary, 1987; Fennema and Carpenter, 1981; Armstrong, 1981). The test material that was used in the South African gender study (Visser, 1985) did not contain enough items to allow for analysis with regard to content.

(Content of a spatial nature is discussed in section 3.3.1.4 and geometrical content in section 3.3.1.5.)

3.3.1.3 Recent research

The continued pattern of **contradictory** results observed by researchers and reviewers (Bohlin, 1994; Friedman, 1994; Zambo and Follman, 1994; Leder, 1992; Wainer and Steinberg, 1992; Leder and Fennema, 1990; Lummis and Stevenson, 1990; Kimball, 1989; Feingold, 1988) in research on gender differences in mathematics achievement has encouraged researchers to investigate a number of possible factors influencing mathematics

performance and also start questioning existing research on gender differences in mathematics performance.

Aptitude, standardized achievement and classroom tests

A different pattern in mathematics performance seems to develop for males and females on different types of mathematics tests, e.g. aptitude tests, standardized achievement tests and classroom tests or grades (Bohlin, 1994; Benbow, 1992; Wainer and Steinberg, 1992; Young, 1991; Kimball, 1989; Benbow and Stanley, 1982). Some researchers (Fennema and Sherman, 1977) argue that aptitude and achievement tests in mathematics are not easily distinguishable, but others (Benbow, 1992; Flaugher, 1978) see them as different measuring instruments. Aptitude tests on spatial ability often include the mental rotation of three dimensional objects which may have little resemblance to an achievement test on the identification of alternate angles. Classroom tests, or grades, overlap with standardized achievement tests with regard to content, but give different results in gender research (Wainer and Steinberg, 1992; Kimball, 1989). Also aptitude tests and achievement tests frequently overlap in content because current aptitude cannot be measured in isolation from previous achievement (Benbow, 1992; Flaugher, 1978). Du Plessis et al. (1993) point out that although male students at tertiary level function at a higher cognitive level (Lawson's standardized test) than females, the females perform better academically. The researchers ascribe this to the more effective learning and study strategies of female students. The contrasting relationships between aptitude and achievement seem extremely complex and raise some questions.

Is it possible, as is sometimes claimed, to improve spatial visualization as measured by aptitude tests, and is achievement in secondary school Euclidean geometry related to the ability of rotating solids mentally? More questions of a similar nature can be asked. Again it seems that great care must be taken in generalization of research results. A problem that often confronts researchers on gender differences in mathematics achievement is that test material is not well described or included in articles. The content of ability tests is usually treated as confidential.

The following inconsistencies have been reported. After primary school girls often achieve better than boys in tests that are classroom and curriculum orientated (Du Plessis et al., 1993; Benbow, 1992; Wainer and Steinberg, 1992; Ferrini-Mundy, 1987; Pattison and Grieve, 1984; Benbow and Stanley, 1982). In the case of Benbow and Stanley (1982) and of Benbow (1992), the grades were self-reported and may lead to subjective results. Standardized achievement tests and aptitude tests often give mathematics results in favour of boys (Benbow, 1992; Becker, 1990; Benbow and Stanley, 1982; Fennema and Carpenter, 1981; Maccoby and Jacklin, 1974).

The significance of the above discussion for the present research is that results on aptitude tests and achievement tests are often not separated in gender studies and reviews on gender differences in mathematics performance. According to Kimball (1989), female superiority in mathematics grades is often ignored or 'downplayed' in the literature on gender differences in mathematics achievement and research involving grades has shortcomings.

> "... researchers will be able to answer a number of important questions that cannot be answered at this time because of the unsystematic inclusion of grades as an achievement measure in the existing literature." (Kimball, 1989: 203).

On the other hand research involving the widely used Scholastic Aptitude Test is accused of being biased (Wainer and Steinberg, 1992; Young, 1991).

> "There is evidence of differential validity by sex disfavouring women on the SAT-M; women score lower, on average, than men of comparable academic performance." (Wainer and Steinberg, 1992: 330).

Young (1991) developed a procedure, based on the Item Response Theory, to eliminate under prediction for one gender. (See chapter 4 for details of the Item Response Theory and the three parameter logistic model that was used to analyze the data of the present research.)

The test material used for the present study is described in section 5.4. The geometry tests are curriculum based, but probably measure more than classroom performance because there is an element of novelty involved. The teachers could not specifically prepare the pupils to answer the open ended questions and almost all the items of the multiple choice test come from the standardized itembank of the Human Sciences Research Council.

Recent meta-analytical studies

Hyde et al. (1990) point out that statistical procedures to perform meta-analysis have advanced considerably since the time of Hyde's (1981) initial meta-analysis. (See section 3.3.1.1.) Some researchers see meta-analysis as a more objective and systematic method of doing literature reviews than the more traditional procedures (Frost et al., 1994). Friedman (1994) considers meta-analysis to be a way of handling an overwhelming amount of information consisting of non-uniform and sometimes contradicting data.

Care must be taken not to over-emphasize the results of meta-analysis as is often done in the literature on gender differences. Although the selection of research studies are sometimes done with care as in the case of Frost et al. (1994), many restricting factors remain inherent to the approach. A few of them are the following:

* Usually only mean differences are compared.
* A considerable variety of mathematical content is included.
* Not all studies give details about specific content.
* Samples are often divergent as well as self-selective.
* Aptitude tests and achievement tests do not always have similar outcomes for the same population.
* Different research methodologies are used in the studies involved.
* Tests of different format are included.

Recent reviews done by meta-analytical techniques (Frost et al., 1994; Hyde et al., 1990), including pupils of different ages and ethnic groups, showed a slight female superiority in computation, no gender difference in understanding of concepts, and a slight male superiority

in problem solving. The male advantage on problem solving increased to a moderate advantage in high school years. The effect size of 0,29 for problem solving in high school translates into 43% of females and 57% of males falling above the mean of the overall distribution. The sample included studies from 1967 to 1987 comprising more than 3 million pupils (Hyde et al., 1990).

Both the above-mentioned meta-analytical studies included studies using college entrance examinations. Such samples have undergone a fair amount of self selection, as is mentioned in section 3.4.

Decrease in gender differences

The meta-analytical analysis of Hyde (. (1990) provides evidence that the magnitude of gender differences in mathematics achievement has declined over the past three decades. These findings are consistent with those of Feingold (1993; 1988), who documented a reasonable decline in the magnitude of gender differences in abilities as measured by several standardized tests. Hanna (1994) also supports this view. A different opinion is held by Halpern (1989) who argues that a decrease in gender differences in mathematics achievement is an over simplification of the situation because, for example, the changing composition of the high school population in the USA over the last few years.

High performances

Gender differences in mathematics are especially observed when above-average performance is considered, with boys achieving at the higher levels (Benbow, 1992; Becker, 1990; Benbow and Stanley, 1982; Fennema and Carpenter, 1981). Feingold (1988) claims that, although many previously found gender differences are disappearing, the gap between high-school male and female achievement in mathematics has remained constant from 1960 to 1983 for high achievers. With an analysis of a longitudinal study of the top 1% performers, Benbow (1992) confirms the earlier findings (Benbow and Stanley, 1982) of a large male advantage in average mathematical ability. Benbow (1992) finds no gender differences in the relationship between ability as measured by SAT-M (Scholastic Aptitude Test -

Mathematics) and subsequent academic achievement. This is in sharp contrast with the criticism of Wainer and Steinberg (1992) of the gender bias in measurements of the SAT-M. (See section 3.4 for a discussion of criticism on SAT-M.)

It must be kept in mind that the above discussion applies to high-achievers only. Hyde et al. (1990) warns:

> "Certainly it is important to study gifted populations, but it is essential to remember that results from studies like Benbow and Stanley's do not generalize to the rest of the population." (Hyde et al., 1990: 150).

Several researchers (Hyde et al., 1990; Tartre, 1990; Walkerdine, 1989; Kimball, 1989; Fennema, 1981) blame the **press** for generalizing from restricted research results which leads to misconceptions about differential gender abilities.

In a study, performed in Hawaii inclusive of the four main ethnic groups, the high achieving females outscored the high achieving males for grades 4, 6, 8 and 10 (Brandon et al., 1987).

Visser (1985) obtained somewhat conflicting results for South African secondary pupils. When the top 100 pupils in grade 11 were considered, the boys performed significantly better than girls in mathematics. When pupils doing the higher grade or more difficult course, were separated from the rest, the top 50 performers showed no gender differences in achievement. Visser (1985) acknowledged that the mathematics tests were too difficult to draw significant conclusions from.

On the other hand, although boys and girls have similar intellectual abilities, and no noteworthy gender differences in twelfth grade final mathematics examinations are found (Visser, 1988b), boys perform significantly better in the national mathematics olympiad which is held yearly in South Africa (Le Roux, 1986). The mathematical ability required is of a much higher level than that required to obtain the grade 12 school leaving certificate.

Euclidean geometry is considered to be fairly difficult and proof writing is usually classified as a high-level cognitive task (Senk and Usiskin, 1983). Sometimes it is only the high performers that are successful in doing the more difficult riders. The level of ability required to be successful in the geometry tests of the present research is believed to be fairly high. The open ended questions are as least as difficult as the Euclidean geometry in the school leaving examination. A fair amount of deductive reasoning as well as hypothesis testing is included. Gender differences among high performers should be revealed in the results, if such differences in performance in Euclidean geometry exist.

Variability

Greater male variability in mathematical abilities could explain the large male advantage in mathematical reasoning (Benbow, 1992; Benbow and Stanley, 1983; Benbow and Stanley, 1982) found for top performers. Feingold (1992) confirmed the conclusions drawn by Maccoby and Jacklin (1974) that males were more variable than females in mathematical and spatial abilities. Feingold (1994) however, warns that these findings may be valid only for the USA. Becker and Forsyth (1994) argue that differences in variability between males and females suggest that examining differences in means alone, cannot provide a clear picture of gender differences in mathematics performance. In that respect, the present study aims to look deeper than mean differences. If present, greater male variability will show up in the statistical analysis.

3.3.1.4 Spatial component of cognitive domain

Because of the spatial nature of Euclidean geometry the spatial component of the cognitive domain is considered in more detail. Soviet researchers (Yakimanskaya, 1991; Zykova, 1954) stressed the importance of spatial thinking for geometry. The lack of clear definitions with respect to spatial abilities may explain some of the **inconsistency** in the literature on spatial ability, on the different components of spatial ability such as spatial visualization, spatial orientation, spatial perception and imagery, and also the relationship with mathematics (Friedman, 1994; Clements and Battista, 1992; Zimmerman and Cunningham, 1991; Tartre, 1990; Friedman, 1989; Johnson and Meade, 1987; Halpern, 1986; Bishop, 1983).

Also Linn and Petersen (1985) mention that no agreement exists for classification of measures of spatial ability. Friedman (1994) suggests a meta-analysis to make sense of divergent reports. (See section 2.5 for definitions and classification used in the present research.)

In their review on the literature of gender differences, also referred to in section 3.3.1.1, Maccoby and Jacklin (1974) reported that one of the areas in which significant differences between males and females exists, is spatial ability. Since then there has been debate over the existence, the extent and the first appearance of these gender differences. McGee (1979) confirmed the existence of a male advantage and also distinguished two categories of skills, spatial visualization and spatial orientation. (See section 2.5 for Tartre's (1990) more detailed categorization based on that of McGee's.)

Linn and Petersen (1985: 1483/4) used a different categorization for their meta-analysis of spatial abilities. They went to great trouble in selecting only reliable studies because in their own words 'the results of a meta-analysis is only as valid as the studies that go into it'. A great variety of spatial tests were partitioned to form three broad categories:

* 'Spatial perception' - spatial perceptions with respect to one's own body; This term relates to spatial orientation as defined by Tartre (1990) in section 2.5;

* 'Mental rotation' - the ability to rotate two- or three-dimensional objects mentally. This relates to spatial visualization as defined by Tartre (1990) in section 2.5.

* 'Spatial visualization' - spatial ability tasks that involve difficult, multistep manipulations, including the above-mentioned tasks and also non-visual strategies.

Gender differences in favour of males appear from eight years onwards for spatial perception, and from any age where measurement is possible, for mental rotation. For spatial visualization there were no differences (Linn and Petersen, 1985). The relative lack of gender differences in spatial visualization is a problem not easily explained and shows some of the complexity in investigating spatial abilities. Maybe the inclusion of non-visual tasks was a confounding factor.

The completely different use of the term 'spatial visualization' by researchers in the same country (Tartre, 1990; Linn and Petersen, 1985) has caused some amount of confusion in the literature on gender differences. Johnson and Meade (1987) differ from both above-mentioned categorizations. They consider spatial ability as a unitary trait and found a large male advantage in spatial ability from grade 10.

Baenninger and Newcombe (1989) investigated the possibility that gender differences in spatial ability may be explained by gender differences in spatial experience. Their meta-analysis showed, surprisingly, that spatial activity participation is only weakly related to spatial test performance and that the relationship is similar for males and females. They also found that spatial ability test performance can be improved by training and that the improvement is similar for males and females. The last finding is also supported by the research of Ben-Chaim et al. (1988) who investigated spatial visualization (as defined in section 2.5) before and after an instruction period. They determined that, despite initial sex differences in favour of boys, the gain was similar for both genders. The question whether girls can possibly, with further training, erase the gap was not considered by Ben Chaim et al. (1988).

The most common explanations for the gender differences on spatial tasks involve differences in the exposure of males and females to environmental factors (Kimball, 1989; Senk and Usiskin, 1983) that may advance spatial ability. According to Leder (1992), support for this explanation comes from studies that have shown that spatial visualization skills can be improved by training. From this point of view, the results of the meta-analysis of Baenninger and Newcombe (1989) are somewhat contradictory, because of the small correlation ($r = 0,1$ for males and $r = 0,09$ for females) between spatial ability and spatial

experience as opposed to the improvement of spatial ability by training. Baenninger and Newcombe (1989) admit that the lack of a stronger correlational relationship may be ascribed to a possible lack of reliability of activity participation measures.

According to Feingold (1988) the decrease in cognitive gender differences that he reported, also hold for spatial ability. However, Feingold was challenged by Halpern (1989) who claimed that the decrease only applies to certain spatial tasks. Analysing additional data, Feingold (1993) maintained that gender differences in spatial abilities have decreased over the years and now show only a small male advantage. This is confirmed by Friedman's (1994) meta-analysis.

It seems that there is consensus about a male advantage on certain spatial tasks, such as mental rotation, but not on all spatial tasks. As is the case with mathematics performance, considerable contradictory evidence is found with respect to gender differences in the field of spatial ability.

Visualization in a mathematical sense and imagery are not often investigated for gender differences. Eisenberg and Dreyfus (1991) say that females are more reluctant to visualize in mathematics than males while Presmeg (1993) reported a previously found absence of gender differences in visual imagery.

The relationship between spatial ability and performance in mathematics and geometry

> "The relationship between mathematics achievement and visual-spatial skills continues to be a matter of debate." (Leder, 1992: 613).

Numerous mathematicians and mathematics educators have suggested that spatial ability plays a vital role in mathematical thinking (Clements and Battista, 1992). Exactly what the relationship is, however, has not been established. Positive correlations have been reported between spatial ability and mathematics achievement at different grade levels (Fennema and Sherman, 1977). Some researchers, especially those that think of spatial ability in terms of imagery (Brown and Presmeg, 1993; Zimmerman and Cunningham, 1991; Presmeg, 1986),

suggest that spatial ability is important even for the solution of non-geometric problems. Clements and Battista (1992: 444), agree but also warn that the role that spatial thinking plays in the construction of mathematical concepts is 'elusive and, even in geometry, multifaceted'.

A recent meta-analytical study (Friedman, 1994) reports that spatial skill does not necessarily underlie mathematical skill. An unexpected low correlation between geometry and both spatial visualization and spatial orientation was found. Hershkowitz (1989) reasons that the role of visualization (in a mathematical sense) is extremely complex within the process of geometry concept attainment, and sometimes work in two directions. The examples of the basic concepts in geometry are visual figures. But if this initial visual image is not developed sufficiently, it later contributes to inflexible, prototype geometrical thinking, which is a hindrance (section 2.3). According to Brown and Presmeg (1993), such pupils do not develop their concrete imagery into more dynamic imagery (section 2.5). In the same vein, Yakimanskaya (1991) stresses that only the creation of an image is not enough, and that manipulation of that image is a higher level of thinking.

> "However, as our data demonstrate, the acquisition of the ability to apply techniques for creating spatial images from their graphical representation, no matter how effective and refined, still does not yield the ability to use these images successfully." (Yakimanskaya, 1991: 103).

Gender differences in spatial ability linked to gender differences in mathematics achievement

Mathematics in general is considered first; geometry is discussed in the next section. Because of the positive correlation between achievement in mathematics and spatial ability (Clements and Battista, 1992; Fennema and Sherman, 1977) and the above-mentioned gender differences in both fields, often in favour of males, researchers started to look for links between the two.

Research studies report that spatial visualization skills are important for the learning of mathematics (Fennema and Sherman, 1977), that it is more important for girls than for boys (Sherman, 1980), that the opposite is also true (Connor and Serbin, 1985) and that spatial visualization plays a key role in gender differences in mathematics performance (Johnson, 1984; Fennema and Carpenter, 1981). There is, however, also evidence that spatial visualization does not account for gender differences in mathematics performance (Fennema and Tartre, 1985; Pattison and Grieve, 1984; Armstrong, 1981). Differential gender influence of spatial abilities on performance in mathematics is suggested by Tartre (1990), Battista (1990) and Ethington and Wolfe (1984). Tartre's (1990) and Battista's (1990) research is discussed in the next section.

Again, one has problems in trying to integrate these divergent findings by respected researchers. Leder and Fennema (1990) conclude:

> "The conflicting findings - exacerbated by definitional confusion - that have characterized much of the research suggest the need for carefully structured and clearly described further study." (Leder and Fennema, 1990: 188).

Friedman (1994) tries to make sense of these divergent results in a meta-analysis and concludes that combined correlational evidence does not support the theory that spatial skill is the determiner of gender differences in mathematical achievement.

A great deal of contradiction can be explained by the multi-dimensional nature of mathematics and therefore Euclidean geometry is considered on its own, for the present research. The experimental design is fairly well structured and the procedures during the empirical investigation were properly monitored.

3.3.1.5 Euclidean Geometry

There is not much research in the field of more advanced Euclidean geometry at secondary level, for reasons mentioned previously. (See sections 2.6.3 and 2.6.5.) Most gender studies reporting on Euclidean geometry only use very basic geometrical ideas and often only

include calculations and no deductive reasoning. Often research refers to geometry in general and does not mention exactly what type of Euclidean geometry is considered. (See section 2.2 for the different approaches to Euclidean geometry.) In the present research, Euclidean geometry at secondary level including deductive reasoning and proof writing, is evaluated. The content consists of circle theorems and tangents to the circle. The approach to geometry is the classical synthetic one. (See section 5.4 and Appendices H and I.)

The most recent meta-analyses (Frost et al., 1994; Hyde et al., 1990) report a slight male superiority in geometry performance. Frost et al. (1994), however, warns that the usefulness of the analysis is compromised by the fact that so many of the studies fail to report the mathematics content of the tests, or they use tests with a mixed content.

It is interesting to note that the significant male advantage in mathematical reasoning that was found by the well-quoted Benbow and Stanley (1982), and that had inspired much research, is in the field of algebraic word problems (Becker, 1990), and not in the field of spatial or geometrical content as was often hypothesized. Becker (1990) does not support Benbow and Stanley's argument because she found that boys and girls performed similarly on geometry and girls outperformed on miscellaneous items which can also be regarded as mathematical reasoning. The term 'logical geometrical reasoning' is used in this study to distinguish it from 'logical reasoning', or mathematical reasoning, which often contains high verbal content.

In the second National Assessment of Educational Progress (NAEP) in the USA (Fennema and Carpenter, 1981), it was found that boys scored higher than girls on geometry for age levels 9, 13 and 17 overall cognitive levels. At the age of 17, the boys' advantage was an alarming 16 percentage points. It is not clear what the role of under-representation of girls in geometry courses played. Problems with the geometry curriculum in the USA are discussed in section 2.6.3. None of the first three NAEP's tested achievement in presenting proofs in geometry. Senk (1985) reports how few studies have been conducted on proof writing in geometry and how small the samples were. A general lack of assessment detail is a characteristic of research in Euclidean geometry. The time factor in assessing open ended questions in geometry is a possible explanation. The fourth NAEP (Brown et al.,

1988) includes some items dealing with elementary concepts of proof, but no multi-step deductive proof writing, probably because only multiple choice questions are used in such large scale studies. Doolittle and Cleary (1987) found that the eight geometry items included in the Assessment Mathematics Usage Test were more difficult for college bound females than for males. The sampling procedures could not, however, eliminate the differences in previous instruction of the two gender groups. The present research hopes to address those shortcomings with regard to proof writing and differential instruction. Both multiple choice questions and open ended questions are included in the geometry test instrument.

The first large scale study of geometry proof writing ever conducted in the USA (Senk and Usiskin, 1983) is of importance for the present research, because so very little research on Euclidean geometry has been published lately. Their sample consisted of about 2 500 students, of whom a sub-sample of about 1 400 students had studied proof writing. They were chosen to represent a national cross-section of educational and socio-economic conditions. The content of the tests ranged from easy to difficult and covered topics such as parallel lines, congruent and similar triangles and quadrilaterals. The sub-samples, consisting of about half females and half males, were evaluated for presentation of proofs.

Except for one form of the three forms of the test, the mean score of boys was slightly higher than that of girls, but never significantly so. When the scores were adjusted (using ANCOVA) to provide for the significantly lower scores of girls on the entry test at the beginning of the year, there was a small female advantage, which was not significant. Senk and Usiskin (1983) concluded that, although girls enter the high school geometry course with generally less geometry knowledge than boys, they achieve equally well in proof writing at the end of the one year Euclidean geometry course offered in secondary schools in the USA. The same was true for the high performers. The geometry questions used in the present research are more advanced than the ones of the above research.

Senk and Usiskin (1983) reason that geometry proof is a 'unique' topic. It does not lend itself to algorithmic thinking and is of a high cognitive level, it is based on classroom experience and is unlikely to be influenced by experience outside the classroom, for example, computers, advanced reading in mathematics or competitions. They come to the conclusion

that the more an instrument directly measures students' formal educational experience, the less the likelihood of gender differences in performance.

Battista (1990) discovered a few worthwhile relationships with regard to geometry but the study was limited to one school and a sample of only 145. He investigated the role of spatial visualization and logical reasoning in geometry performance and related it to gender differences in geometry performance. Males scored significantly higher than females on spatial visualization, geometry achievement and geometric problem solving but there was no evidence of gender differences on the test of logical reasoning. It must be mentioned that the logical reasoning test contained a high verbal content. When predicting geometry achievement, spatial visualization was the most important factor for females, whereas logical reasoning and the discrepancy between spatial visualization and logical reasoning were most important for males. Battista (1990) suggests a fundamental difference in the role that spatial visualization and logical reasoning play in males' and females' learning of geometry. For poor performers, spatial visualization correlated higher with geometry performance than logical reasoning. No details on this are given for the two genders separately. The geometry content is not described fully but seems to have been a mixture of coordinate geometry and plane sections of three-dimensional figures. No pure Euclidean geometry, in a deductive sense, was included.

Battista's (1990) results are consistent with those of Ethington and Wolfe (1984) who report that women appear to convert their spatial abilities into mathematics achievement to a greater extent than men do. Tartre (1990) also found differential gender influence of spatial orientation (SO) on performance in mathematics. (See Tartre's description in section 2.5.) No significant difference was found overall between males and females in SO skill or in mathematics achievement. However, females that have low SO scores appear to have substantial lower achievement means in mathematics than males with comparable spatial orientation scores. On the other hand the high-SO females correctly solved the visually presented problems more often than the high-SO males. Tartre (1990) suggests that the 'integration' of spatial skills and analytical skills may be of importance in the successful solution of problems that are presented visually. Tartre's (1990) problems were of a general type and not Euclidean geometry. In the present research all diagrams to the problems are

supplied. Integration of spatial and analytical skills in mathematics achievement is also stressed by Battista (1990), Eisenberg and Dreyfus (1991) and Tall (1991).

South Africa

Swanepoel (1990) compared the knowledge of grade 9 boys and girls with respect to basic concepts of Euclidean geometry. He found that boys outperformed girls, and came to the conclusion that his findings link up to the school of thought according to which gender differences in mathematics achievement do not manifest exclusively as a result of non-cognitive factors. A male advantage for geometry is also reported by Nyangeni (1993). Due to problems experienced in secondary schools in Transkei at the time of the research, the achievement scores are rather low for meaningful conclusions to be made from them. Approximately 60% of males and 80% of females failed (below 40%) the geometry test. The gender research of Visser (1985) gives no details on Euclidean geometry, only of mathematics in general.

The present study intends to contribute to research in the field of Euclidean geometry. The geometrical content used for the evaluation of Euclidean geometry goes far beyond basic concepts and includes deductive reasoning and hypotheses testing, generally associated with the fourth van Hiele level (section 2.6.5). The teaching of mathematics in general, and also of Euclidean geometry is fairly successful in the 5 schools that participated in the experiment. The failure rate will probably not restrict the research results.

3.3.1.6 International studies

The data for the Second International Mathematics Study (SIMS) was gathered during the 1981-82 school year. South Africa did not participate in SIMS but will participate in the Third International Mathematics and Science Study (TIMSS). Two groups were evaluated for SIMS, i.e. pupils approximately 13 years old and pupils in their last year of secondary school who are studying mathematics as a substantial part of their academic program.

International studies on achievement in mathematics and geometry are worth noting because of large cross-cultural differences that may question the comparability of research results for different countries. Hanna (1994: 421) reports that (for pupils in their last school year) the maximum difference **among countries** in the mean percentage is about 44 percentage points, while the differences **between genders** only differ from 1 percentage point (Thailand) to 11 (Japan).

Younger pupils

The data of SIMS shows approximately **equal** numbers of boys and girls in the samples of most of the 20 countries that participated. The following is a summary of results for 13 year old pupils:

> "... girls tended to outperform boys in the use of computational skills and in algebra. Boys tended to outperform girls in geometry and measurement." (Robitaille and Travers, 1992: 700).

Studying the original data (Robitaille, 1989), one notices that in four countries, viz. Belgium (France), Belgium (Flemish), Sweden and Thailand, young girls scored better than boys in **geometry**, though the differences were small. Robitaille (1989) warns that these results may be due to a lack-of-fit between the geometry curriculum in those systems and geometry defined by the set of items used in SIMS.

Lummis and Stevenson's (1990) more limited international study compares the mathematics achievement of young children from three countries, Japan, Taiwan and the USA. They improved on previous studies by including only those test items that were taught at approximately the same grade level in all three locations. Gender differences in favour of boys were found as early as the first grade and also in grade 5 in, inter alia, **visualization** of various transformations of geometrical forms. Even though Japanese girls attained higher average scores on these and other tasks than did American boys, the superiority of Japanese boys to Japanese girls was as great as that of American boys to girls. Gender differences were of comparable magnitude in all three countries.

SIMS (Robitaille and Garden, 1989) does not report on visualization separately, which makes comparison impossible.

Older pupils

At the age of 17-18 years mathematics has become predominantly a male domain in most of the 15 countries that took part in SIMS. Apart from Hungary and Thailand, in all other participating countries, the percentage of females based on enrolment dropped drastically, e.g. to 21% in Hong Kong, 22% in Japan, 34% in England and Wales and 44% in the USA. This means that, by the end of secondary school, a substantial degree of selection by gender has occurred in most countries (Wagemaker and Knight, 1989). This fact, along with differences between countries in rates of retention of students in mathematics, complicates the comparison of gender differences across countries (Robitaille and Travers, 1992). Pupils from Hong Kong and Japan generally outperformed pupils from other countries. (See section 3.1 for the almost equal gender representation in the present research.)

Thailand, British Columbia and England showed almost no significant differences between the achievement of males and females on all subtests. In countries like Hungary, Japan, Israel, Finland, Belgium (French) and Belgium (Flemish) there were several large differences in favour of males. For **geometry** males did better than females on all but one of the items (Hanna et al., 1990). The average for geometry was only 38% however, which is again an indication of the lack of international consensus regarding appropriate geometrical content. No Euclidean geometry, beyond elementary calculations, was included (Robitaille and Taylor, 1989).

In a smaller international study (Robitaille and Travers 1992) Japanese students outperformed their Illinois counterparts by more than two standard deviations in mathematics. Japanese males outscored Japanese females by about one standard deviation, whereas Illinois male and female scores differed very little. Japanese females far outscored both sexes in the Illinois sample (Robitaille and Travers, 1992).

If the selection by gender of high school samples, as mentioned by Wagemaker and Knight (1989) is considered, the above comparison may be questioned on grounds of the degree of selectivity of Japanese high school samples in favour of males. One, however, wonders to what extent the much larger portion of Euclidean geometry that Japan includes in the school curriculum, influences such results. Miwa (1992) and Nohda (1992) give the following details of the Japan geometry curriculum. From grade 2 until grade 6 pupils are introduced to the drawing of basic geometric forms, parallel lines, perpendicular lines, parallelograms, circles and scale drawings. Proof is introduced in grade 8 and circle theorems and tangents (as is used in this research) are taught in grade 9. This means that Japanese pupils start with geometry earlier and eventually are two years ahead of South African pupils.

3.3.2 Affective domain

Several factors from the affective domain have been researched in relation to gender differences in mathematics achievement. This study concentrates on the cognitive domain but, because of the modern view of interaction between the two domains, the affective domain will be discussed briefly where it interacts with the present study. Benbow's (1992) train of thought sums it up. Ability, as measured with aptitude tests is

> "a measure of potential, whereas educational opportunity, family characteristics and attitudes are some of the factors that determine whether childhood potential is translated into adult achievement." (Benbow, 1992: 59).

Nyangeni (1993) found a positive correlation between pupils' performance in Euclidean geometry and attitude towards geometry, while Bohlin (1994) ascertained a higher correlation between confidence and geometry as between confidence and algebra. Males' higher scores on confidence in and a positive attitude towards mathematics (Bohlin, 1994; Frost et al., 1994; McLeod, 1992) seem to be two important factors influencing participation in mathematics courses. Fennema and Sherman, (1978) found that girls show less confidence than boys in their ability to do mathematics as early as grade 6, and from grade 10 onwards, these negative attitudes contribute to increasingly differential enrolment of the two sexes in

advanced mathematics classes. Similar conclusions were made by Visser (1985) for a South African sample. Many girls drop mathematics at the end of grade 9 as a result of low confidence in the subject. The 5 large schools that participated in this experiment have approximately equal numbers of boys and girls enrolled in mathematics at school leaving age. The sample seems to be representative of both genders for these schools. (See section 5.2.)

Differential gender attributions to success and to failure are often linked to mathematics achievement (Leder, 1992; McLeod, 1992; Kimball, 1989; Walkerdine, 1989). Leder (1992) describes it a follows:

> "Work carried out within the fear of success paradigm suggests that females' lower performance in mathematics is a function of internalization of, and conforming to, the expectations of others rather than being a function of ability per se." (Leder, 1992: 615).

Dweck and Gilliard (1975) described the difference between male and female attitudes towards success and failure in mathematics. According to them, women attribute success to hard work, while men attribute success to ability. Also, women attribute failure to lack of ability while men blame lack of effort or bad luck. Walkerdine (1989) adds that many teachers also treat pupils accordingly.

Kimball (1989) uses the above arguments to explain the issue of females performing better in grades and males better in standardized tests. (See section 3.3.1.3.) Standardized mathematics tests, in relation to internal examinations, is more likely to be perceived as a measure of one's ability, and may favour boys. On the other hand, examinations may be presented by teachers as a measure of how hard one has studied, i.e., a measure of effort which will favour girls. Kimball (1989) also argues that the novelty of standardized tests may disfavour girls who generally show less confidence in mathematics than boys. The tests that are used in the present study are curriculum based and were treated as an examination. The tests, however contain an element of novelty because they are not compiled by the teachers and do not resemble the usual textbook material.

3.3.3 Educational factors

Only factors which can be linked directly to the present research, such as learning strategies and the gender of teachers, are discussed.

G. Pask (1976) mentioned two distinct learning strategies - 'serialist and holist'. Serialists proceed from certainty to certainty, learning, remembering and conceptualizing a body of knowledge in small sequential steps. They have a need for structure in the learning situation. Holists, on the other hand, prefer to start in an exploratory (often unstructured) way and have a need for an overall framework before considering the details. According to Scott-Hodgetts (1986), males tend to employ holistic learning strategies in mathematics more frequently than females do and argues that it may help explain the gender differences in mathematics performance. The most successful mathematics pupils will generally be those with versatile learning strategies. Bohlin (1994) supports the above argument as a possible explanation for gender differences in mathematics performance. She found that females have a greater need for a structured mathematics learning environment. The two modes of thinking, holist and serialist, may be related to girls' better performance in computations and boys' better performance in problem solving as discussed in section 3.3.1.2. The above reasoning can also be related to brain laterization as discussed in section 3.4. Solving problems in the field of Euclidean geometry seems to require, at least initially, a holistic learning approach rather than a serialistic one and the learning environment is usually highly unstructured. Both these factors may be to the disadvantage of females according to the above research results.

Hanna (1994) points out that the ratio of female to male mathematics teachers in high schools may be an important factor in explaining gender differences in performance, but data from SIMS (Hanna, 1994; Hanna et al., 1990) shows no consistent pattern. Very small gender differences were observed in British Columbia where 3% of the mathematics teachers are female. In Hungary, on the other hand, where 60% of the teachers are female, very large gender differences in mathematics performance emerged. The teachers involved in the present study are mainly female, not as a matter of choice, but as a result of the majority of

secondary mathematics teachers being female in the Transvaal Education Department of South Africa.

3.3.4. Biological factors

The controversial findings of Benbow and Stanley (1982) that boys outperform girls by far among the highly gifted, sparked off a renewed interest in biological factors as partial determinants of spatial ability and mathematical ability. Benbow and Stanley (1983) stated that by the age of 13 a large gender difference in mathematical reasoning ability exists and that it is especially pronounced at the high end of the distribution where boys outnumber girls at the ratio of 13 to 1. It was suggested that the difference has a biological determinant. A few years earlier McGee (1979) pointed out that sex differences in spatial skills are likely related to sex differences in hemispheric specialization. The right hemisphere is specialized for spatial thinking processes and males have greater right specialization than females.

The solving of problems in Euclidean geometry requires some spatial thinking and usually such problems are classified as tasks of high cognitive level. According to the above arguments males are likely to outperform females on such tasks. The other argument that is often offered is that the brain performs optimally when both hemispheres are well developed and when there is good intersupport between them. Tall (1991) gives the following summary of different modes of thought processes linked to the different hemispheres of the brain:

"Left hemisphere	-	verbal, gestural, logical, analytic, linear, sequential, conceptual similarity;
Right hemisphere	-	visual spatial, communication, analogical, intuitive, synthetic, gestalt, holist, multiple processing, structural similarity." (Tall, 1991: 107).

Tall (1991) warns that the breakdown is not always related precisely in the above way. He argues that the most powerful way to use the brain in mathematics, is to 'integrate both ways

of processing', relying on the right hemisphere to give 'global linkages and unifying patterns, whilst analysing the relationships and building up logical inferences between concepts with the left'. Eisenburg and Dreyfus (1991) also emphasize the integration of visual and analytical thinking in mathematical problem solving.

Problem solving skills in Euclidean geometry possibly involve spatial, analytical and logical modes of thought and obviously a large degree of interaction between the different modes. Clements and Battista (1992) underscore the idea of using both hemispheres with regard to Euclidean geometry.

The debate whether hemispheric laterization is related to gender differences in spatial ability and perhaps also mathematics achievement is not settled yet (Rudisill and Morrison, 1989; Halpern, 1986). The trainability of spatial abilities complicates the issue. Halpern (1986) reasons that the decrease in gender differences in spatial abilities cannot be ignored.

> "Spatial ability is clearly not an immutable, biologically determined trait. Like almost all of the traits that are of interest to psychologists, the extent to which an individual develops spatial ability depends on a host of interacting factors. It seems likely that sex differences in spatial ability would be reduced, possibly even eliminated, if sex role pressures and expectations became sex neutral." (Halpern, 1986: 1015).

The present research may shed some light on this issue because the previous exposure to geometry has been similar for both genders. If some components of spatial ability, as represented in performance on Euclidean geometry, are influenced by training, it might show in the results of this study.

3.4 PROBLEMS AND LIMITATIONS IN RECORDED RESEARCH

A variety of problems and limitations are encountered in literature on gender differences in mathematics achievement. Most of the literature on gender differences in mathematics achievement appear in journals published in the USA. A substantial amount of these studies

use scores obtained on the Scholastic Aptitude Test - Mathematics (SAT-M) or other college entrance examinations. The research of Benbow and Stanley (1982; 1983) and Benbow (1992) were, for example, based on SAT scores. It is thus a matter of crucial to note the limitations that arise in the SAT and similar admission tests with regard to gender differences (Wainer and Steinberg, 1992; Becker, 1990; Feingold, 1993; 1988; Young, 1991; Halpern, 1989). Also note that Hyde et al. (1990) excluded the scores of about 800 000 testees on SAT in their meta-analysis on gender differences in mathematics. Chipman et al. (1991) suggest that contradictory findings with respect to geometry may be the result of the small number of geometry items in the SAT.

Feingold (1988) states that normative data on the SAT are available **only** for self-selected high school pupils who registered for and completed the examination usually required for college admission. Such unrepresentative samples can yield biased trends. The sample for the present research is more representative of both genders. (See sections 3.1 and 5.2.)

This criticism of the SAT seems fair in the light of Wainer and Steinberg's (1992) results. (See also section 3.3.1.3.) These authors matched a large number of students according to the grades they obtained in first year tertiary mathematics. These students' scores on the SAT, however, showed a substantial advantage in favour of male students. Wainer and Steinberg (1992) conclude that their results provide evidence of 'differential validity' of the SAT by gender, disfavouring women. Halpern (1989) agrees that SAT scores lead to under-prediction for women, and Young (1991) recommends Item Response Theory to eliminate such gender bias.

Since Fennema and Sherman (1977) identified the problem of differential gender exposure to mathematics, it has been mentioned as a complicating factor in the research of high school mathematics in the USA (Leder, 1992; Kimball, 1989; Smith and Walker, 1988; Ethington and Wolfe, 1984; Senk and Usiskin, 1983; Armstrong, 1981) and also in New Zealand (Blithe et al., 1994). Sometimes it is controlled by means of statistical techniques and sometimes by matching mathematics courses, but it remains a problem that is not easily dealt with. Some researchers see it as one of the most important factors.

"Finally, the present research suggests once again that a great deal of the difference in mathematics achievement between men and women can be explained by differences in background, ability, attitudes, grades, and formal exposure to mathematics in the classroom. Of these, the variables measuring exposure to mathematics had the most influence on explaining variation in mathematics achievement." (Ethington and Wolfe, 1984: 376).

As result of differential exposure, different outcomes for the same study in the USA have been reported by Senk and Usiskin (1983). In 1964, the project TALENT study originally reported significant gender differences, favouring males, in mathematics scores of grade 12 pupils. A re-analysis of the same data in 1979, controlling for the number of years students have studied mathematics, showed no significant gender differences.

Differences in the mathematical backgrounds of pupils are not a problem for the present study because of the school system in South Africa at the time of the empirical part of the research. All pupils in the sample have theoretically the same past classroom experience in mathematics.

Problems in research with regard to the multi-dimensionality of mathematics and spatial ability have been identified by several researchers (Friedman, 1994; Feingold, 1993; Yakimanskaya, 1991; Halpern, 1989; Connor and Serbin, 1985; Linn and Petersen, 1985; Pattison and Grieve, 1984). Feingold (1993) stresses the misconceptions that can arise when a general point of view is taken on gender differences in mathematics achievement.

" ... it is inappropriate to make generalizations about gender differences in verbal, mathematical, and spatial abilities because gender differences vary across the more homogeneous processes within these domains. Thus, research on cognitive gender differences should focus on uni-dimensional processes." (Feingold, 1993: 98).

Contradictory research findings may be a result of viewing spatial ability as a unitary trait. Spatial visualization, one of the components of spatial ability, is usually evaluated by means

of tasks that involve mental rotation of objects in space. (See section 2.5 for definitions.) These tasks are often performed better by males than females (Battista, 1990; Halpern, 1989; Linn and Petersen, 1985), but these tasks are highly specialized and may not have an influence on performance in mathematics or even Euclidean geometry.

> "Processes of speed of mental rotation, most strongly implicated in spatial ability sex differences, do not correspond to processes likely to contribute to mathematics and science performance." (Linn and Petersen, 1985: 1493).

Spatial orientation, another component of spatial ability, often show no (Tartre, 1990) or small (Linn and Petersen, 1985 - different terminology) gender related differences. This component seems to be partially related to the visual aspects of Euclidean geometry, as was mentioned in section 2.5. If research on these two components are combined, i.e. spatial ability is considered as a unitary trait, contradictions are likely to appear.

Confusion caused by three different interpretations of the term 'visualization' has probably also contributed to contradictory reports. In psychometrics, spatial 'visualization' is used to describe tests of mental rotation. (See section 2.5 for Tartre's (1990) definition.) Secondly, the same term is used in a totally different sense by Linn and Petersen (1985) in their categorization of spatial abilities. Thirdly, 'visualization' in a mathematical sense cannot be isolated from the rest of mathematics. Visual thinking and graphical representations are linked to other modes of mathematical thinking and students move back and forth among the different modes of thinking (Zimmerman and Cunningham, 1991).

International studies, large national studies and meta-analytical studies may have merit in identifying general trends, but generalizations often lead to misconceptions about gender differences in mathematics achievement. There seems to be a need for more specific, in-depth research in the field of gender differences in mathematics achievement (Feingold, 1993; Leder, 1992; Lynn, 1992; Hyde et al., 1990; Linn and Petersen, 1985; Pattison and Grieve, 1984; Marshall, 1983). Leder and Fennema (1990) advise research on micro level, using more homogeneous groups (Leder, 1992). Hyde et al., (1990) mention the problems they had analysing the data of their meta-analysis because of diverse content.

"Few authors specified the content clearly, probably because the content was mixed. We must know if there are large gender gaps for certain types of content. That can be determined only when researchers construct tests and report results that assess the various kinds of mathematics content separately." (Hyde et al., 1990: 151).

Euclidean geometry is of a different nature to algebra and word problems and comparison across these different branches of mathematics is not always meaningful. Another problem that was experienced with regard to this study, involves the mixed content of tests that are used in research on problem solving. Geometry is considered a good topic to evaluate problem solving but usually some algebraic word problems are also included in tests which make comparisons with either word problems, or geometry, impossible. The solving skills seem to be different.

Some researchers also argue that comparison of mean differences is not enough when considering gender differences in mathematics performance (Becker and Forsyth, 1994; Feingold, 1994; Becker, 1990; Doolittle and Cleary, 1987). When items are used for research purposes, however, the information on the different items is often insufficient to draw meaningful conclusions (Feingold, 1994). Sometimes the confidential nature of ability tests prevent researchers from publishing details of test materials.

The present research aims to contribute in this respect. The research concentrates only on Euclidean geometry at secondary level. Unidimensionality is one of the assumptions of the Item Response Theory which is used to analyze the geometrical content of the tests used in this study.

Burton (1990) raises the issue of using only multiple choice questions in large scale gender studies. Responses on multiple choice questions can only provide certain information and the format of a test influences students' scores differentially on the basis of gender. Sometimes boys tend to do better than girls on multiple choice items (Bailey, 1993; Becker, 1990) and sometimes not (Curtis et al., 1989). (See section 4.5 for a comparison of different question types.) For the present research, multiple choice items are used for the

large sample. The responses to ordinary geometry riders of a sub-sample, consisting of about one quarter of the original sample, are also analyzed. (See also section 5.4.)

Ability levels are seldom considered separately in research on gender differences. Yet, evidence of differential gender functioning in mathematics at different ability levels exists (Becker and Forsyth, 1994). In the introduction to Becker and Forsyth's article, Brusselmans-Dehairs and Henry (1994) advise the use of Item Response models in such cases. The three parameter logistic model of the Item Response Theory which is applied in this study, is considered to be very sensitive in identifying differences. (See section 5.6.)

Psychometric tests that are usually used to measure 'logical reasoning' have a very high verbal content. To what extent results on these tests can be used to explain logical reasoning in a mathematical sense, is debatable. This research considers logical geometrical reasoning as is used for solving problems in Euclidean geometry.

The influence of the press

There is evidence that the media helps to 'shape ideas and attitudes' as well as 'reinforcing popular beliefs' (Leder, 1992). Kimball (1989) mentions the way parents were influenced by the often quoted research of Benbow and Stanley (1982). Mothers of daughters who read media reports (in favour of boys) of Benbow and Stanley's study, when compared to mothers who had not read it, felt that their daughters had less mathematical ability and therefore had to work harder to do well in mathematics. Fennema and Hart (1994) mention the responsibilities of those reporting on research that has strong social implications. Reporting of such research is often more extreme than the results of the original study. Fennema (1981) also refers to the controversial reports in the media on the research of Benbow and Stanley (1982) which were published earlier in the USA. Walkerdine (1989) quotes similar exaggerations in the USA press. The negative role that the press plays toward women when mathematics performance is considered, is also mentioned by Hyde et al. (1990) and Tartre (1990).

3.5 SUMMARY

Conflicting research results are the most outstanding feature of this chapter on gender differences in achievement with regard to mathematics and geometry. Also the attempts to explain these gender differences, where they do occur, produce contradictions. Several complicating factors are almost unavoidable in such research and one can only try to clarify some of the confusion that arises.

International research has shown that gender differences in mathematics performance within countries is much smaller than differences between the performances of different countries. Sometimes these results are generalized across countries.

Most of the recent research on differential gender performance in mathematics is performed and published in the USA. An acceptable consequence of such a situation is that research results on gender differences are compared to those of the USA. At the same time research in mathematics at secondary level in the USA is characterized by several problems. Ability tests that are used at national level for selection purposes to enter tertiary institutions seem to be biased against females with regard to mathematics. Results on these tests make up a large portion of the surveys that are usually quoted in research on gender differences. Samples used in mathematics research at secondary level are subjected to differential gender exposure to mathematics because of the school system that allows different choices of subjects and also different branches of secondary mathematics. It is almost impossible to find fairly large samples in the last year of secondary school with theoretically the same mathematical background, and without limitations of self selection. In that respect, South Africa is in a rather favourable position for research purposes because of a secondary school system that is more uniform. The approximately 800 pupils that participated in this experiment had all taken similar mathematics courses and is representative of both gender groups.

Research results on aptitude tests and classroom orientated tests tend to give different outcomes if gender performance is considered. This accounts for a fair amount of the contradictory results because distinction between different measures is not always made.

The lack of **more specific** research on gender differences in mathematics achievement is mentioned by several researchers as an obstacle in coming to meaningful conclusions. Mathematics is not uni-dimensional and research results covering divergent solution processes may explain some of the contradictions. Gender differences vary across the more homogeneous processes within the different branches and topics in mathematics. The present research is an attempt to investigate gender differences at item level. Only Euclidean geometry is considered and it is believed that the solution processes are unidimensional since the focus is on spatial abilities and their relevance to proof writing in geometry. The three parameter logistic model of the Item Response Theory that is used to do the item analysis, is considered a very efficient procedure to identify gender differences. The model also takes different ability levels into account. In addition to this in-depth item analysis, geometry riders are used to investigate trends that may not be evident when testing with multiple choice questions only.

CHAPTER 4

MEASUREMENT THEORIES

"If anything were actually 'immeasurable', it would bear no relationship of any kind to the rest of the universe, it would not affect nor be affected by anything else in any manner whatever, it would enact no causes and bear no consequences - in short it would not exist." (Rand, 1986: 289).

4.1 INTRODUCTION

Two different measurement theories, i.e., Classical Test Theory (CTT) and Item Response Theory (IRT) are considered. Both theories are discussed briefly on an intuitive level and also compared with each other. Multiple choice questions or items, as they are generally known, are used as question format in both theories. Although open ended questions are traditionally used in tests, multiple choice questions have become popular in countries such as the United States, Australia, New Zealand and also in South Africa in the past two decades. Some advantages of using items in tests will be apparent, especially in the discussion of differential item functioning within both theories.

4.2 CLASSICAL TEST THEORY

4.2.1 History of Classical Test Theory

Classical Test Theory (CTT) developed at the beginning of this century when Spearman and Thorndike applied statistical methods to construct tests, to interpret scores and make decisions about the test scores. The model has been refined over the years and served measurement specialists well for a long time (Barnard, 1992; Hambleton, 1989). It is still widely used. In CTT, test scores are analyzed within the framework of concepts like total

score (raw score), mean, standard deviation, difficulty value, discrimination value, validity, reliability and standard error of measurement.

4.2.2 Assumptions of Classical Test Theory

According to Barnard (1992; 1991) and Allen and Yen (1979), some of the problems with CTT originate in the following assumptions. The **first** assumption of CTT is that the observed score (O) of a testee on a test equals a true score (T) plus an error score (E). Mathematically this leads to the following linear relationship: $O = T + E$. The value of T is a fixed value and represents the ability of the testee. O and E are variables and the above equation shows how the observed score is influenced by error of measurement. The value of E can be positive or negative depending on overestimation or underestimation of a testee's true ability. Hambleton (1989) states that these observed and true scores are test-dependent and rise or fall with changes in test difficulty, amongst other factors.

Suppose a testee writes the same test an infinite number of times, and assume that the testings are independent in that each test has no influence on any other. The average of these observed scores is then equal to the true score of the testee, according to the **second** assumption of CTT. It is clear that the concept of true score is a theoretical one, since an infinite number of administrations of a test to the same testee is not feasible. Nonetheless it still provides useful results in practice.

The first two assumptions concern the theoretical distribution of a testee's observed scores on subsequent testings of the same test. The next assumption concerns the scores of a number of testees obtained in a test. The **third** assumption of CTT states that the true scores and error scores obtained from a number of testees in a test are uncorrelated, which means that a testee with a high ability has the same error measurement as another with low ability. It is acceptable to assume that a testee with below average ability is more likely to guess than a testee with high ability when items are used. Hambleton (1989) says that this assumption of equal errors of measurement for all examinees seems implausible and although there are corrective measures to take, it is often overlooked. Barnard (1992; 1991) and

Allen and Yen (1979) mention several other assumptions of CTT regarding error scores and parallel tests which are not important to consider here.

4.2.3 Difficulty and discrimination value

Difficulty and discrimination values are two of the key concepts that operate on question or item level in CTT. The **difficulty value** of an item is defined (Barnard, 1992; Allen and Yen, 1979) as the proportion of testees who answered an item correctly, to all the testees, i.e.

$$p_i = \frac{N_r}{N_s}$$

where

p_i is the difficulty value of the item;

N_r is the number of testees who answered the item correctly, and

N_s is the number of testees who took the test.

Note that all testees don't always answer all the items and therefore N_s is sometimes taken as the number of testees who answered the item.

An easy item will have a relatively high difficulty value (close to 1) and a very difficult item a rather low (close to 0) one. It is not desirable to prescribe limits to this index as it will differ for different purposes. Generally a value between 0,2 and 0,8 is acceptable (Barnard, 1991). Allen and Yen (1979) reason that if maximum discrimination among examinees over all levels of performance is desired, a range of difficulty levels from 0,3 to 0,7 is best.

The **discrimination value** of an item is defined as the product moment correlation (Pearson) between the total score on an item and the total score on the test (Barnard, 1992; Allen and Yen, 1979). This index indicates the degree to which the item succeeds in distinguishing between testees that perform generally well and those that perform generally poorly. The formula for the discrimination value is the following:

$$d = \frac{X_r - X_t}{S_t}\sqrt{\frac{p}{q}}$$

where

d is the point biserial correlation coefficient;

X_r is the mean score of the testees who answered the item correctly;

X_t is the mean score of all testees who took the test;

S_t is the standard deviation of the scores of the group;

p is the proportion of the testees who answered the item correctly, and

$q = 1 - p$, i.e. the proportion of testees who answered the item incorrectly.

With tests consisting of items, the proportion of right answers is inflated by an increment due to chance. Guilford (1965) gives a formula for the proportion of correct answers to an item, with chance success eliminated but Barnard (1991) concludes that corrections for guessing remain only an estimate.

4.2.4 Validity

Guilford (1965) suggests that a test is valid, in the crudest terms, when it measures what it is presumed to measure. Validity is a highly relative concept and therefore the valuation of a test's validity must always be in respect to its specific purpose and a specific situation. A test that is valid in one situation may not be valid in another. Various aspects of validity have been distinguished. The best known classification of the types of validity was developed by a joint committee of the American Psychological Association and the National Council on Measurement in Education (Ary et al., 1985). This committee distinguished three types of validity - content validity, criterion-referenced validity and construct validity. Crocker and Algina (1986) and Barnard (1992) also consider these as the major components of validity.

Content validity refers to the extent to which the instrument represents the content of interest. In order to have content validity, a measure must adequately sample both the topic and the cognitive processes included in the content universe under consideration. Content validity

is a subjective judgement and is generally not expressed in terms of a numerical index. This type of validity is often considered the most important in scholastic achievement tests.

Criterion related validity refers to the relationship between the scores on a measuring instrument and an independent external criterion. The emphasis is on what the instrument can predict rather than on the content. Empirical methods are used to study criterion related validity. Criterion-referenced validity is usually of prime importance for standardized psychological tests.

Construct validity is concerned with the extent to which a test measures a specific trait or construct. A measure of a particular construct should be as independent as possible of measures of other constructs. This type of validity is usually of greater importance in psychological tests than in scholastic tests.

4.2.5 Reliability

The reliability of a measuring instrument is the degree of consistency with which it measures what it purports to measure. As was mentioned in section 4.2.2 the observed score of a testee contains an error component. Reliability procedures are concerned with determining the degree of inconsistency in scores due to random error. In other words the reliability of a test concerns the measurement error that can cause a testee's score to fluctuate from one testing to another. The reliability of a test is expressed as a number between 0 and 1, and represents the proportion of the true score variance to the observed score variance (Guilford, 1965). For example, a reliability coefficient of 0,8 means that 80% of the observed score variance can be attributed to true variance and 20% to error variance.

Determining the reliability coefficient

There are various ways in which a reliability coefficient can be determined. Reliability can be estimated by correlating the scores obtained by the same individuals on different occasions or with different sets of equivalent items. These procedures require two administrations of a test and are generally called **test-retest reliability** or **parallel-forms reliability**. In

practice it is usually difficult to obtain data on two occasions. It is, however, also possible to get a measure of reliability from a single administration of a test by correlating two comparable halves of the test. This results in **split-half reliability**. This method obviously requires a large sample and a relative large number of items. Because of the practical problems inherent in these methods, methods not reliant on correlation procedures were developed to determine the internal consistency of a test. Probably the most widely used index of homogeneity are the Kuder-Richardson formulas, of which the KR 20 formula is perhaps the best known. It is based on the proportion of correct and incorrect responses to each of the items of the test. The Kuder-Richardson 20 formula is:

$$\rho_{xx} = \left(\frac{n}{n-1}\right)\left(\frac{\sigma_x{}^2 - \Sigma p_i q_i}{\sigma_x{}^2}\right)$$

where

ρ_{xx}	is the reliability coefficient of the test;
σ_x	is the variance of the test;
p_i	is the difficulty value of item i;
q_i	is equal to $1 - p_i$, and
n	is the number of items in the test.

Guilford (1965) suggests that the Kuder-Richardson formulas, and all the internal consistency formulas that depend upon a single administration of a test, probably underestimate the reliability of a test.

The interpretation of a reliability coefficient should be based on a number of considerations, some of which are:

* The reliability of a test is in part a function of the length of the test. The longer the test, the greater the possibility of high reliability.

* Reliability is partly a function of group heterogeneity. The more heterogeneous the sample of testees, the higher the reliability coefficient will be.

* The reliability of a test is in part a function of the ability of the individuals who take the test. A test may be reliable for high levels of ability but not for lower levels and vice versa (Crocker and Algina, 1986; Ary et al., 1985).

A concept usually linked to the reliability coefficient of a test is the **standard error of measurement**. It is calculated as follows:

$$\sigma_e = \sigma_x \sqrt{1 - r_{xx}}$$

where

σ_e is the standard error of measurement;

σ_x is the standard deviation of the test scores, and

r_{xx} is the reliability coefficient of the test.

It may be considered as an estimate of the average standard deviation of testees' individual error distributions for a large number of repeated testings. It should be noted that the standard error of measurement reflects an average of many testees' individual standard errors. It is probably not realistic (Crocker and Algina, 1986) to suppose that these standard errors are equal for all testees.

4.2.6 Some of the most important shortcomings of CTT

Although CTT is syntactically correct, problems arise with practical applications (Barnard, 1992; Vorster, 1988) because item statistics are relative to the characteristics of the test and of the testees (Hambleton et al., 1991; Hambleton, 1989; Angoff, 1982; Warm, 1978; Lord, 1977a).

Some statistics derived by means of CTT are **sample dependent** (Hambleton, 1989). The classical item difficulty, defined as the proportion of testees who answered an item correctly, is directly related to the ability rating of the group. A sample of relatively high ability testees will give higher difficulty indices than a group of lower ability testees. Hambleton

(1989), Ebel (1986) and Lord (1977a) conclude that this group-dependence reduces the usefulness of item statistics.

The discrimination value of an item was defined earlier as the correlation between the item score and the total test score. This value is a function of the general ability level of the testees in the sample. It thus depends on the homogeneity of the ability level of the testees in the sample (Barnard, 1992). The effect of group heterogeneity on correlation coefficients is also mentioned by Hambleton (1989), Crocker and Algina (1986), Ary et al. (1985) and Guilford (1965).

> "The item statistics consequently depend on the samples from which they have been obtained. If the statistics of the same set of items are calculated for testees from different samples, the CTT model offers no method according to which the different statistics can be used to obtain the item parameters. Methods falling outside the framework of this model have to be employed. When representative samples are used to calculate the item parameters, such parameters can be used only for groups of testees who are comparable with respect to their abilities and the distribution of such abilities in the group for which the results have been calculated." (Barnard, 1992: 12).

Another shortcoming of classical test models is that the observed and true scores are **test-dependent** (Hambleton, 1989). The total score of a testee is equal to the number of items answered correctly. This score, which is associated with the testee's ability, is directly dependent on how difficult or easy the test was. What actually happens is that the scores of the testees change according to the level of the test, but that their abilities remain the same. A direct consequence of this is that comparisons between testees can be made only if the same or strictly parallel tests have been administered (Barnard, 1992).

Two more sources of concern with respect to Classical Test Theory lie in the definition of **reliability** and, what may be thought of as its conceptual converse, the **standard error of measurement** (Hambleton et al., 1991). From the definition of reliability (see section 4.2.5) it is clear that a reliability coefficient is relative to the variance of the test and thus

to the heterogeneity of the sample. Barnard (1992) asserts that this classical index therefore also depends highly on the sample used to obtain the data.

Hambleton's (1989) criticism of the assumption of equal errors of measurement has been mentioned in section 4.2.2. The problem with the standard error of measurement, which is a function of the test score reliability and variance, is that it is assumed in CTT to be the same for all testees. In practice it has been found (Barnard, 1992) that its consistency varies as a function of ability. It is likely that testees with high ability may perform more consistently than testees with very low ability, since the latter more often tend to guess, especially with multiple choice questions. (See section 4.2.2).

Vorster (1988), Hambleton (1989) and Barnard (1992; 1991) mention more shortcomings of CTT with regard to parallel tests.

Another characteristic of CTT is that, although it sometimes uses item statistics, it usually operates at test level. The total score of a testee is used as basis for the statistical analysis (Barnard, 1992; Vorster, 1988; Hambleton and Swaminathan, 1985; Baker, 1985; Lord, 1977b). CTT is thus more test than item orientated. However, fluctuations within a test, like testees' responses to different items, are often important for the researcher.

The issue of bias, either cultural or gender, is often researched using CTT procedures. Several researchers strongly object to this because of the failure of CTT to handle different ability levels of subgroups.

> "It can be shown that classical parameters, e.g. the difficulty value p, will generally not be linearly related across subgroups of a population. This means that the test for cultural bias using classical parameters can lead to an artifactual detection of bias." (Warm, 1978: 16).

Hambleton and Swaminathan (1985) and Lord (1977a) are of the same opinion. They assert that classical approaches to the study of item bias have often been unsuccessful because they fail to adequately handle true ability differences among groups of interest. Also, the mere

fact of test dependent ability estimates and dependence of statistics on the group of testees from which they were obtained, questions the validity of bias studies within this framework.

4.2.7 Summary

CTT and associated methods used for interpreting test scores have been used successfully for decades and it is by no means rejected as a model to analyze test scores. Measurement specialists have devised procedures to deal with the problems mentioned. It is the practical application of the model that causes problems, especially with regard to parallel measuring instruments which are difficult to obtain in practice. Hambleton et al. (1991), Hambleton (1989), Ebel (1986) and also Lord (1977a) conclude that there exists a need for a model that can give item statistics that are not group dependent and which describes a testee's ability irrespective of the difficulty level of the test administered.

4.3 ITEM RESPONSE THEORY (IRT)

4.3.1 Introduction

Considering the problems encountered with the CTT it is strange that a theory like IRT was not used more widely earlier. Warm (1978) gives the following explanation.

> "It is widely acknowledged in the field that one of the main reasons that item response theory has been so slow to catch on among testing practitioners is the mathematical complexity of the literature." (Warm, 1978: 9).

In order to make the basics of the theory more accessible to the general practitioner, Baker (1985) developed a software package dealing with Item Response Theory for a microcomputer. He mentions that even his brief introduction to Item Response Theory would not be possible without the use of a computer due to the amount of advanced mathematical procedures involved in the theory.

Hambleton and Swaminathan (1985) also blame the mathematical complexity of the topic for the slow progress. Additionally they encountered notation problems, but conclude that in a rapidly expanding field such as IRT, uniformity of notation cannot be expected. Hambleton et al. (1991) motivate their recent publication as follows:

> "our approach focuses on the conceptual basis of item response theory and avoids discussion of mathematical derivations or complex statistical aspects of the theory." (Hambleton et al., 1991: x).

Scarcity of advanced computer programs is mentioned as a further problem by Barnard (1991). The iterative procedures involved in the estimating of parameters require sophisticated software and powerful computers. Until a few years ago such software was not readily available. Most of these problems have been solved and, especially in the USA, a considerable number of articles on successful application of IRT have been published. During the past decade the Item Response Theory has become well established in countries such as the USA, Canada, England, Western Europe and Australia (Wolfaardt, 1990). Today IRT can be considered a very comprehensive and refined model in educational and psychological measurement (Masters, 1994; Barnard, 1992; Hambleton et al., 1991; Barnard, 1991; Vorster, 1988).

IRT has not been used widely as a technique for item analysis in South Africa to date. The few articles published are mainly from researchers associated with two or three research bodies. See Barnard (1993a; 1993b; 1992; 1991), Barnard and Oosthuizen (1992), Wolfaardt (1990), Owen (1989), Vorster (1988) and Hoffman (1985).

Barnard (1991) warns that technical and other problems still remain to be solved, but it is also clear that the IRT technology is sufficient to provide a theoretical basis for the solution of a wide variety of problems.

4.3.2 History of IRT

Item Response Theory (latent trait theory or the item characteristic curve theory, as it was called earlier) originated, theoretically, in the 1920s, but it was only in 1952 with the publication of Frederick Lord's Ph.D dissertation that Item Response Theory was presented in its own right. Lord is considered the father of the Item Response Theory. He prepared the basis for the application of IRT to dichotomous items as it is used today. Progress in the fifties was painstakingly slow (Hambleton and Swaminathan, 1985) due to the non-existence of computer programs and computer hardware to handle the complex statistical calculations. Georg Rasch developed the one parameter model of the Item Response Theory during the sixties and Benjamin Wright introduced it to educationists. This model is still in use, also in South Africa. In 1968 Lord and Novick published their psychometrics textbook, in which Birnbaum, a well known statistician (now deceased), presented the mathematics of the three parameter logistic model (Barnard, 1991; Warm, 1978).

According to Baker (1985) the fields of educational measurement and psychometrics are in a transitional period. Although the majority of present practice is based on Classical Test Theory, there is a clear tendency towards models of the Item Response Theory because of its advantages (Barnard, 1991; Hambleton et al., 1991; Ebel, 1986).

> "In the last decade we have witnessed a revolution in educational and psychological measurement as the application of classical measurement theory has given way to the use of item response theory." (Hambleton et al., 1991: vii).

4.3.3 Assumptions of IRT

Because of few limitations in its assumptions, CTT is valid for the majority of tests and this makes the theory flexible. Problems experienced in practice have been discussed in section 4.2.6. IRT is less flexible because of its assumptions (Barnard, 1992). Item response

models are mathematical models (Hambleton, 1989) and therefore based on specific assumptions about the test data.

The **first assumption** concerns dimensionality (Hambleton, 1989). **Unidimensionality** is assumed in most IRT models, i.e., it is accepted that there is one predominant trait that explains the test score of a testee (Barnard, 1992). Generally this means that all the items in a test measure only one area of ability or knowledge (Warm, 1978). It is interesting to note that unidimensionality is generally accepted in CTT without mentioning it.

This unidimensionality assumption cannot be strictly met because there are always cognitive, personal and other factors that affect test performance (Hambleton, 1989). Research has shown that the unidimensionality assumption requires only that the items in a test measure the same composite of abilities, rather than only a single ability. According to Barnard (1992) and Hambleton et al. (1991) this assumption is satisfactorily met if achievement can be explained by one dominant underlying factor. This dominant component is referred to as the ability (trait) measured by the test. Typical ability tests, such as numerical and spatial perception tests are probably unidimensional (Warm, 1978).

Two matters relating to dimensionality are important. A speed test has a second dimensi involved which affects the responses on the last items, and therefore IRT is restricted power tests (tests where the majority of testees have the opportunity to finish the test) unless certain adjustments are made to accommodate speed tests. Secondly, the dimension of a test is not influenced by the distribution of the abilities of the testees (Barnard, 1992; 1991).

Models in which it is assumed that more than one ability is necessary to account for a testee's test performance, are referred to as multidimensional. They are more complex and, to date, have not been well developed (Hambleton et al., 1991).

The **second assumption** of IRT is that of **local independence** (Hambleton, 1989). For this assumption to be true, a testee's performance on one item must not affect his response on any other item of the test. One item must not provide a hint for another item and no item must be a prerequisite for another (Barnard, 1992; 1991; Hambleton et al., 1991). The

most common situation where local independence does not hold is with speed tests. A testee may get the last questions wrong simply because he did not reach them and consequently there is no possibility of the last questions being correctly answered (Warm, 1978). Computer programs usually make provision for questions not reached and questions not done.

Thirdly is assumed that a strictly increasing relationship exists between the probability of success in an item and the underlying ability that it measures. **Finally**, it is assumed that a testee will probably answer an item correctly if he/she knows the answer to the item.

4.3.4 Definition of IRT

A theory of test-item responses supposes that a testee's performance can be explained by certain abilities or traits. Barnard (1992) states that these abilities are changeable; they can be improved by training and education. However during the testing session ability is assumed to remain constant.

Barnard (1992) gives the following definition for Item Response Theory:

> "**Item Response Theory** is a statistical theory consisting of a family of **mathematical** models expressing the **probability** of a particular response to an item as a **function** of the **ability** of the testee and of certain characteristics of the item." (Barnard, 1992: 15).

The three most popular unidimensional IRT models are the one-, two- and three-parameter logistic models - so named because of the number of item parameters each incorporates. These models are appropriate for dichotomous item response data.

The three parameter model is the most sophisticated and usable model of the three for multiple choice items (Barnard, 1991; Crocker and Algina, 1986; Warm, 1978) if the sample is of sufficient size. For the three parameter model, Barnard (1991) advises a sample of at least 500 if the computer program WARMEST is used. If the program LOGIST is used, 1000 testees would be better. Hambleton (Barnard, 1991) reckons that

a minimum of 600 is sufficient. The number of items should not be less than 15 (Hambleton, 1989).

4.3.5 Item Response Function

Item Response Theory postulates two things. First, the performance of a testee on a test can be explained by an underlying dominant ability, and secondly, that the relationship between the performance of a testee (on each item) and the ability measured by the test can be described by a **monotonic increasing** function (Barnard, 1992; Hambleton, 1989). This function is referred to as the **Item Response Function (IRF)**, also called the Item Characteristic Curve (ICC), and specifies that as the level of **ability** increases, the **probability** of a correct response to an item increases (Hambleton et al., 1991).

An IRF is thus a **mathematical** function that relates the probability of success on an item to the ability measured by the item and the characteristics of the item, called the item parameters. The standard **mathematical** model for the Item Response Function is the cumulative logistic distribution function with the usual S-shape. In general S-shaped curves are described mathematically by the equation

$$y = \frac{p}{q + e^{-rx}}$$

where e is the natural base for logarithms and *p, q and r* are variables. If *p, q and r* are all equal to 1, the equation becomes

$$y = \frac{1}{1 + e^{-x}}$$

and it is clear that if $x \to 0$, then $y \to 0$ and if $x \to \infty$, then $y \to 1$. The upper asymptote of this function is 1 and the lower one is 0.

The mathematical function (IRF) for the three parameter logistic model shows some similarities:

$$P(\theta) = c + (1 - c)\frac{1}{1 + e^{-aD(\theta - b)}}$$

where

$P(\theta)$ is the probability of answering the item correctly;

$D = 1{,}702$ is a scaling constant;

a is the item discrimination parameter;

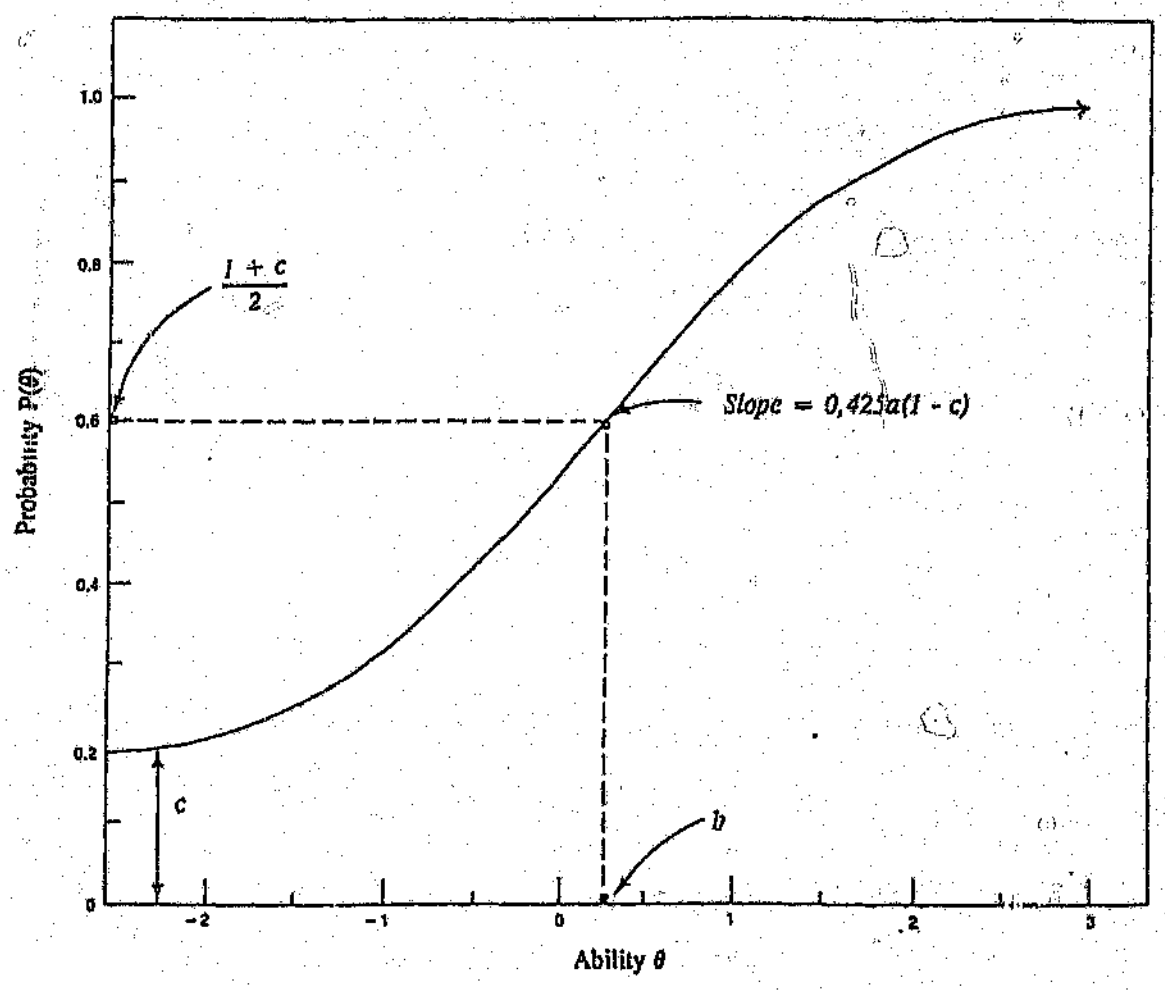
b is the item difficulty parameter;

c is the pseudo-chance level parameter, and

θ is a measure of the testees' ability.

If this function is presented graphically (Hambleton, 1989), the concepts become clearer.

Figure 4.1: Item Response Function (three parameter logistic model)



* θ represents the ability of the testees and theoretically it ranges from $-\infty$ to $+\infty$. Since the unit of measurement and the origin are arbitrary, it is common to transform the scores to have a mean ability score of 0 and a standard deviation of 1 (Crocker and Algina, 1986; Baker, 1985). That means that most of the ability values will lie between -3 and +3 as for the normal distribution.

* $P(\theta)$ is a **probability** and the scale for $P(\theta)$ ranges between 0 and 1. Note, however, that the range of $P(\theta)$ is between c and 1 for the three parameter model. This will be clear when the c-parameter is explained.

  $P(\theta)$ is a **mathematical** function of the abilities (θ) of the testees and of the characteristics of the item, called the item parameters. The three parameter model involves three parameters, denoted by **a**, **b** and **c**, and they determine the shape of the curve. The upper asymptote of the curve is 1.0 and the lower asymptote depends on the value of **c**. (See next paragraph.)

* The **c-parameter** is discussed first because it influences the value of the b-parameter. The c-parameter was incorporated into the model to take into account test scores at the lower end of the ability scale where guessing is a factor, such as in multiple choice tests. It is often called the guessing parameter.

  In the three parameter model the c-value defines the lower asymptote of the Item Response Function and it is different for each item because it depends on the responses of the testees on each item. In the one- and two-parameter models with no provision for guessing, the c-value is 0 and the lower asymptote of the curve is 0 (Lord, 1977a). (See also examples in section 4.3.6.)

* The **b-parameter** refers to the difficulty value of the item and is that point on the ability scale that corresponds to the point of inflection of the IRF. (See also section 5.6.1.) It is also the point (on the ability scale) where the probability of a correct response is halfway between the c-value and 1 (Hambleton, 1989).

The difficulty parameter of the item, namely **b**, is thus the point on the ability scale where

$$P(\theta) = \frac{1 + c}{2}$$

Note that the b-value is associated with a value halfway between c and 1 on the Y-axis, but the actual value of b is found on the θ-axis or the ability scale.

The greater the value of the b-parameter, the greater the ability that is required of a testee to succeed with the item. This parameter is a location parameter that indicates the position of the IRF in relation to the ability scale. Difficult items are thus located to the right or the higher end of the ability scale. A b-value of 0 therefore means an item of average difficulty and negative b-values represent relatively easy items. (Also see examples in section 4.3.6.)

* The **a-parameter** is referred to as the discrimination index in IRT. The a-parameter is proportional to the slope of the IRF. It is a maximum at the point where $\theta = b$ (Lord, 1977a). High values of the a-parameter result in item response functions that are steep, and low values lead to functions that increase gradually as a function of ability. The value of a determines the degree to which the item may distinguish between testees with low abilities and those with higher abilities. (Also see examples in section 4.3.6.)

Similarities between CTT and IRT do exist with regard to concepts and terms used. However, there is no direct relationship between, for instance, the difficulty index (b-parameter) of IRT and the classical interpretation of difficulty value of an item as in CTT. The basic definitions differ (Barnard, 1991).

Since the Item Response Function is a strictly increasing monotonic function, testees with high scores (abilities) have a higher expected probability of answering an item correctly than testees with lower scores. In other words, the IRF is the non-linear regression function of item score on ability measured by the item (Hambleton, 1989).

To clarify the meaning of $P(\theta)$, consider an item i. The interpretation of $P_i(\theta)$ is not the probability that a **particular** testee at ability level θ will answer item i correctly. $P_i(\theta)$ must be viewed as the probability associated with a **randomly** selected testee at ability level θ, answering item i correctly (Hambleton, 1989; Crocker and Algina, 1986). An equivalent explanation is that if $P(\theta) = 0{,}7$ it means that 70% of testees at ability level θ answered the item correctly (Barnard, 1991; Crocker and Algina, 1986).

Numerical values of parameters

Most authors are hesitant to classify numerical values for the parameters, since these values are relative to some frame of reference. The opinions of a few authors will be discussed.

The **b-parameter** gives an indication of where the item functions on the ability scale. For example, an item with b = -1 functions optimally among the lower ability group while a b-value of 1 denotes an item that functions well among the higher ability testees.

Barnard (1991) gives the following guidelines for the difficulty parameter b.

$b < -1{,}96$	- very easy items
$-1{,}64 < b < 1{,}64$	- average items
$b > 1{,}96$	- very difficult items.

Hambleton (1989) and Hambleton et al. (1991) give a similar range of values between -2 and +2. Baker (1985) considers items with $b > 1{,}35$ as difficult items.

The discrimination index is the **a-parameter** and this parameter indicates the extent to which the item succeeds in distinguishing between high and low performers. The a-parameter is proportional to the slope of the IRF. At $\theta = b$ the slope will be a maximum because it is the point of inflection of the curve. (See also section 5.6.1.) The actual slope can be calculated by differentiating the function $P(\theta)$ with respect to θ and substituting b for θ, which then gives a value of 0,425a(1-c) (Hambleton and Swaminathan, 1985).

Although the value of **a** will be highest at the point corresponding to **b** on the ability axis, it is important to note that these two values are independent of each other. The value of **a** is reflected in the slope of the IRF and cannot be determined on one of the axes, as is the case with the other two parameters.

Barnard (1991) gives the following classification for a-values:

$a < 0{,}3$	- very low discrimination
$0{,}65 < a < 1{,}34$	- average discrimination
$a > 1{,}34$	- high discrimination

Warm (1978) says experience has shown that discrimination values of typical items vary from 0,5 to 2,5.

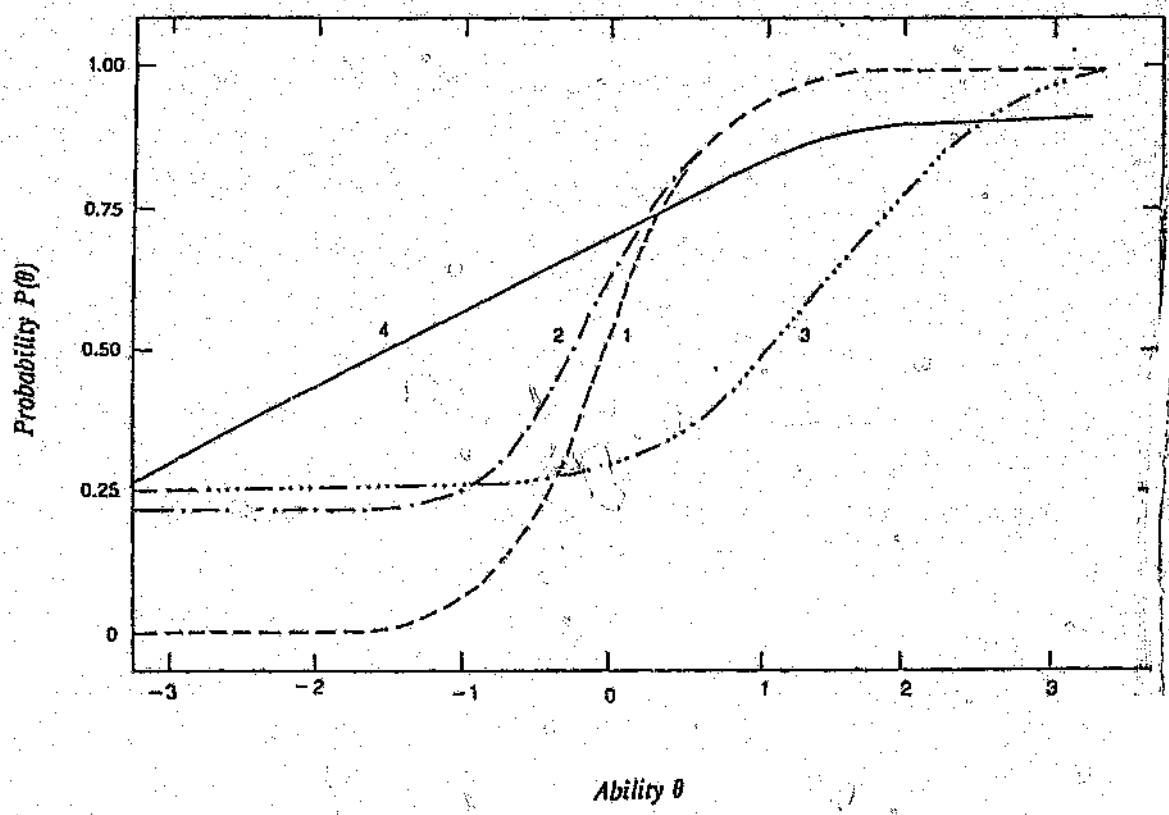
The third parameter, **c**, is a measure of the amount of guessing that has taken place. If a testee guesses randomly in a multiple choice test with four possible answers, his chances of answering an item correctly is 0,25. Typically, but not always, the value of **c** is less than the chance level that would be achieved by a testee guessing at random. The reason is that test developers are ingenious in providing attractive distractors to items, with the result that testees who do not know the answer often do less well than if they had chosen their answers randomly (Lord, 1977a). For this reason Hambleton et al. (1991) prefer referring to it as the pseudo-chance level parameter instead of the guessing parameter (Crocker and Algina, 1986), as it is often called. Note that the possibility of testees with low abilities answering an item correctly, approaches **c** as the ability θ becomes smaller (Barnard, 1991).

The c-parameter has a theoretical range between 0 and 1. In practice, c-values above 0,35 are not considered acceptable (Baker, 1985). According to Barnard (1991), values of **c** exceeding 0,3 are undesirable.

4.3.6 Examples of IRFs

For further explanation, the graphical representations (Hambleton et al., 1991) of the IRFs of four items are discussed.

Fig. 4.2: An example of the IRFs of four items



Examples of the Item Response Functions of four items is found in Figure 4.2. The numerical values of the respective parameters for the four items are the following:

Item	b	a	c
1	0	1,8	0
2	0	1,8	0,2
3	1,5	1,2	0,21
4	−1,2	0,4	0,25

Items 1 and 2 are of average difficulty and discriminate quite well for testees of average ability. Item 2 has the greater chance factor of the two. Item 3 is more difficult than items 1 and 2, therefore the shift to the right of the curve. It discriminates best at a higher ability level than the first two items, and the chance for guessing is 0,21. Item 4 is very easy and has a very low discriminating value of 0,4. Most of the testees have been relatively successful on this item.

The diagrams of the IRFs are determined by the values of the three parameters. Although the diagrams serve a good explanatory function, they are not essential for comparing items. Considerable information can be obtained by comparing the numerical values of the parameters. However, if the performance of two groups, like the gender groups for this study, are compared, differences can be illustrated very well by means of graphical representation of the Item Response Functions. (See section 5.6.3.)

For the present study, the item parameters of the multiple choice questions (see section 5.7.2) are estimated using the three parameter logistic model of the Item Response Theory. (The Item Response Functions of the items that produce differential item function, are supplied in sections 5.6.3.2 - 5.6.3.5.)

4.3.7 Estimation of parameters

Hambleton et al., (1991) say the first and most important step in applying IRT is that of estimating the parameters of the chosen model successfully.

> "In fact, the successful application of item response theory hinges on the availability of satisfactory procedures for estimating the parameters of the model." (Hambleton et al, 1991: 32).

The estimation of the parameters is part of the complex mathematical procedures that were referred to in section 4.3.1. Only the basic outline of a very simple example is discussed here.

Assume the abilities of a group of testees are known. Divide the group into many smaller groups, each with the same ability θ. Now the proportion of correct responses, which is an estimate of the probability of correct responses, i.e. $P(\theta)$, can be computed for each of the ability levels. If the values of θ and corresponding values of $P(\theta)$ are plotted on a suitable scale, the result will be a scattered pattern of dots, roughly in the expected S-shape.

The basic task now is to choose a suitable model that best fits the observed pattern of responses. Suppose that the three parameter model fits best. Now initial estimates for the three parameters are chosen and the values of $P(\theta)$ for each ability level is computed using the equation for the IRF. The agreement of the computed value of $P(\theta)$ and the observed value of $P(\theta)$ is determined for all ability groups, and the item parameters are adjusted to find a better agreement. This process of adjusting the estimates, also called calibration of parameters, continues until adjustments get so small that little improvement is possible. In other words, it is an iterative process that leads to convergence to certain parameter values. These last estimates of the parameters are used to calculate the probabilities of correct responses, $P(\theta)$, at each ability level θ and the item response function can be plotted. This curve is thus the curve that best fits the specific set of data for that item. The extent to which the resulting IRF fits the original data can be tested in several ways, one of which is the popular χ^2 – goodness-of-fit index (Baker, 1985).

> "It should be noted that Item Response Theory cannot be shown to be correct because it is simply a general framework for characterizing persons, items and relationships between them. The situation changes as soon as a model within the IRT framework is specified. The appropriateness of particular IRT models with any set of test data can be established by conducting a suitable **goodness-of-fit** investigation." (Hambleton, 1989: 149).

Merely the basic principles of estimation of parameters were given. In practice, the situation is usually more complicated when both ability and item parameters are unknown. A lot of research has been done on the best estimation procedures, specifically to speed up the process of convergence (Barnard, 1992). A variety of computer programs are available, some so efficient as to estimate ability and item parameters simultaneously. One of the most

widely used programs is LOGIST. It uses the joint maximum likelihood method of parameter estimation. To calibrate the parameters it solves (3N + n) simultaneous equations, where n is the number of items in the test and N is the number of testees (Hambleton, 1989). These laborious computations for each item can be a problem in terms of computer time.

WARMEST is a computer program developed by Thomas Warm (Warm, 1978) to calibrate parameters. WARMEST uses the weighted likelihood estimation algorithm while having both the maximum likelihood and Bayesian procedures as options. According to Barnard (1991) WARMEST compares favourably to other programs and it is more economical in terms of computer time. Warm (1985) claims that his method, 'the weighted likelihood estimation', is superior to other methods of estimation. WARMEST can only be used for the three parameter logistic model and has a built-in function to test the extent to which the data on each item fits the model. The empirical results of Barnard (1992) show that the weighted likelihood estimation method is an acceptable alternative to other methods.

WARMEST was used to estimate the item parameters for this study. (See Appendices A, B, C and D for the estimates of the parameters and the goodness-of-fit indices.)

4.3.8 Standard error of estimated ability

As in the case of CTT, there is no way to determine a testee's ability in IRT. It can only be estimated. One can, however, calculate a 'standard error of estimation' (Hambleton, 1989), that provides some indication of the precision of the estimate. The formula (Barnard, 1991; Baker, 1985) expresses standard error as a function of ability. The standard error for each ability level can thus be calculated. Minimum error of measurement is found in the vicinity of the difficulty parameter of the item (Barnard, 1991; Warm, 1978).

4.3.9 Test Characteristic Curve

Statistical analysis in IRT is based primarily on item statistics. Sometimes, in practice the raw scores of testees are of importance. The purpose of the Test Characteristic Curve

(TCC) is to provide a means of transforming ability scores to raw scores. For a specific ability, the raw score of the testees is the sum of all the possibilities of correct responses at that level of ability. In this way the raw scores can be plotted for each ability level, resulting in the TCC (Baker, 1985).

4.4 DIFFERENCES, RESEMBLANCES AND ADVANTAGES - CTT AND IRT

If the assumptions of IRT can be met, the estimates of parameters obtained have two desirable properties for which the IRT became known. They are

> **"item free ability estimates and sample free item parameter estimates."** (Hambleton, 1989: 148).

These properties of invariance of item and ability parameters are the outstanding advantages of IRT (Lord, 1977a) and its major distinction from CTT. The invariance of item parameters, i.e., item parameters that are not dependent on the sample, are considered first.

Hambleton et al. (1991) examined an example of a two-parameter model where different ability groups with known possibilities [$P(\theta)$] were compared. When two values of θ and $P(\theta)$ each (for an interval) are substituted in the IRF, it results in two simultaneous equations with two unknowns, a and b. Calculations showed that the equations produced identical solutions for a and b, irrespective of the interval of θ that was considered. The low ability group, the high ability group and the group as a whole generated the same values for the item parameters of difficulty and discrimination. This means that item parameters are independent of the sample and only dependent on the characteristics of the items.

Secondly, in IRT the role of true score is replaced by an ability parameter, and, in a similar way to that for item parameters, it can be shown that a testee's ability is independent of the test used to determine it. Just as any sub-range of the ability scale can be used in the estimation of item parameters, the corresponding segments of curves of several item response

functions can be used to estimate testees' ability. Thus the ability of a testee is invariant of the test items used.

The above assessment may appear subjective unless a few conditions are mentioned. For the invariance of ability it is important that all the items must measure the same trait and the values of all item parameters must be in the same metric. Note also that the underlying ability of a testee is not immutable; it may change as a function of educational changes as in remedial instruction.

> "The importance of the property of invariance of item and ability parameters cannot be overstated. This property is the cornerstone of IRT and makes possible such important applications as ... investigation of item bias, ... (Hambleton et al., 1991: 25).

In CTT, item statistics are dependent on the group the test was administered to, and the ability of testees (raw score) is dependent on the difficulty level of the test. This is an unfortunate result of the definitions within CTT. In the experiment mentioned previously, Hambleton et al. (1991) found that the **classical** indices for discrimination and difficulty changed considerably for different ability groups. As result of similar findings Hambleton concludes:

> "Dependence on the examinee group reduces the usefulness of these item statistics in test development work and other applications." (Hambleton, 1989: 147).

As was mentioned in section 4.3.5, there is some resemblance between the values of the a- and b-parameters in IRT and the concepts of discrimination and difficulty value in CTT, but there is no direct relationship because the definitions differ. Only in very exceptional cases, when the normal ogive is used as model, can complicated mathematical formulas be deduced to link the two (Barnard, 1991). According to Crocker and Algina (1986) these models have largely been replaced by the logistic models because of computational problems with the models based on the normal ogive. In an experiment with Industrial Psychology

undergraduate students, Wolfaardt (1990) found some similarities in results analyzed by CTT and logistic models of IRT.

Through IRT very useful conclusions can be made, but only if the fairly firm restrictions of the model can be met and the fit of the model is satisfactory.

> "The advantages claimed for item response models can be realized only when the fit between the model and the test data of interest is satisfactory. A poorly fitting model cannot yield invariant item- and ability-parameter estimates." (Hambleton, 1989: 172).

Some computer programs, like WARMEST, have built-in functions to test the degree to which a set of data fits a model. The better the fit, the smaller the index of goodness-of-fit. According to Barnard (1991), a value of less than 1 is acceptable. (See Appendices A, B, C and D for the goodness-of-fit indices of the present research.)

The concept of **reliability** is not used in IRT at all. Instead, errors of estimation can be determined. In IRT, error of measurement is a function of ability (Barnard, 1992; Vorster, 1988; Hambleton and Swaminathan, 1985). This means that the error of measurement can be calculated for each ability level, unlike CTT where

> "the classical standard error of measurement is a group statistic that is applied in the same way to all examinees in the sample." (Hambleton, 1989: 164).

A further advantage for research is the fact that the basic concepts of Item Response Theory rest upon the individual **items** of a test rather than upon a total test score. Total score is useful in some instances, but often two or more testees obtain the same total score on a test while responses to different items vary considerably. In this respect IRT supplies useful information.

4.5 OPEN QUESTIONS AND MULTIPLE CHOICE QUESTIONS

Although valuable information can be obtained using items within the framework of IRT, open (or open ended) questions are used traditionally in tests and examinations on Euclidean geometry. A combination of items and open questions is probably the most appropriate way to evaluate geometry. Both open questions and items, as means of testing, are used in this empirical study and the advantages of both will be discussed.

In South Africa the evaluation of knowledge of Euclidean geometry at senior secondary level usually consist of one or two theorems and riders based on these theorems. At present multiple choice testing in mathematics is not widely used in South Africa, but with the growing secondary school population, multiple choice testing is gaining support (Webb, 1989), and according to Van den Berg (1994) and also Ebel (1986), any aspect of cognitive educational achievement can be tested by multiple choice questions.

> "Empirical research ... has consistently indicated that multiple-choice questions can be used to a very large extent to measure the outcomes measured by traditional written examinations." (Van den Berg, 1994: 6).

A multiple choice question was referred to as an item in the preceding sections. It can, more formally, be defined as follows:

> "An item is a single unit in a test which constitutes the smallest subdivision into which a test can be divided. It consist of a stem (the part in which the question is asked, a problem is posed or an instruction is given) and options, (the part which contains the correct answer as well as several distractors which are incorrect answers)." (Barnard, 1993b: 146).

The following advantages of testing with items are often mentioned:

* Such tests are much quicker to mark and a large number of papers can be marked by one person or with the assistance of a computer.

* The scoring is objective and can be done by machine or non-specialist clerical assistants.
* A large portion of the syllabus can be covered in a relatively short time.
* Assessing time is used most efficiently.
* The costs of examinations, written on a large scale, can be reduced.

The accusation is often made that the use of multiple choice questions tends to fragment knowledge to such an extent that testees fail to appreciate knowledge in an integrated manner. This kind of argument is based on the fact that testees sometimes do better in objective tests than traditional tests. But, as Van den Berg (1994) points out, the level of difficulty of a test can be manipulated by the test compiler, and the higher levels, such as the generalization of knowledge, can be tested by means of multiple choice questions.

The amount of time that it takes to write a good item is one of the disadvantages of multiple choice testing. A more serious objection to multiple choice questions in mathematics is that incorrect thinking processes can lead to correct answers (Prins, 1992) and that erroneous thinking processes may never be picked up by the teacher. Multiple choice questions thus emphasize the product, i.e., the answer and not the process and quality of thought by which the answer was reached (Aiken, 1987). Van den Berg (1994) and Barnard (1993a) reason that wrong answers are not likely to be given accidentally and, if the distractors are chosen carefully, they can pick up misconceptions and faulty reasoning processes. Good items do not permit correct responses on the basis of simple recognition, sheer rote memory or meaningless associations.

Multiple choice questions are sometimes criticised in that the testee can arrive at the correct answer by a process of elimination. Ebel (1986) says it is unlikely that a testee who is totally ignorant of the correct answer, would have knowledge enough to eliminate the incorrect alternatives especially designed to show up common mistakes. It is true that with multiple choice questions testees are penalized equally strictly for careless errors and completely wrong answers, but on the other hand, in a subject like mathematics, accuracy is of great importance. The number of options a multiple choice question should have is debatable. Statistically there is almost no difference in using four of five options (Barnard,

1993a). Although the probability of guessing the correct answer decreases as the number of options increases, items with three options are frequently used with success (Aiken, 1987; Ebel, 1986).

Open questions are, for this study, Euclidean geometry problems (or riders as they are generally called) which consist of a diagram and some given information. Questions are in the form of statements that have to be proved. Several ways of arriving at the right answer are possible. Open questions are easier than multiple choice questions to prepare, but grading of testees' answers is time consuming and highly subjective. Even with a subject like geometry, it is sometimes hard to judge an incomplete proof that is heading towards one of many correct possibilities.

Those who argue for open ended questions usually do so on the ground that open tests provide a better indication of students' real achievement in learning. In geometry, students must have command of an ample store of knowledge which they must be able to organize in a coherent and logical way. Unfortunately, many of the traits that open ended questions are said to measure, such as critical thinking, originality and problem solving in mathematics, are not always clearly defined. The characteristics of answers that indicate these higher order thinking processes are seldom stated explicitly (Ebel, 1986).

However, both types of testing have merits. The optimum testing situation is probably where a variety of question types are used (Barnard, 1993a; Aiken, 1987). In most instances free-response and choice type measures are found to be nearly equivalent within the limits of their respective reliability (Ebel, 1986). Barnard (1993a) found that scores on certain objective tests and conventional type examinations which he compared in South Africa, correlated highly.

4.6 DIFFERENTIAL PERFORMANCE - BIAS

The purpose of this study is to investigate **differential gender performance** in the field of Euclidean Geometry. Most research on gender differences with regard to mathematics performance are done on test level (Battista, 1990; Tartre, 1990; Hanna et al., 1990;

Feingold, 1988; Fennema and Tartre, 1985; Visser, 1985), where only the average score of gender groups are compared. Aptitude tests, standardized tests and achievement tests are used without a clear distinction between the different types (Leder, 1992; Kimball, 1989; Flaugher, 1978). Recently, a few researchers have tried meta-analytical studies in which a great variety of tests from different investigations are combined (Friedman, 1994; Frost et al., 1994; Hyde et al., 1990), to come to a general conclusion about gender differences in achievement in mathematics. Feingold (1994), Hambleton and Swaminathan (1985), Reynolds (1982), Berk (1982) and also Flaugher (1978) pointed out that mean differences in themselves are not legitimate standards for identifying differential performance. The same applies to international studies (Hanna, 1994; Hanna et al., 1990), where even more over-interpretation of differences takes place. Because research often shows differences in spatial visualization (Tartre, 1990; Ben-Chaim et al,. 1988; Fennema and Carpenter, 1981), but findings are inconsistent, researchers have attempted to focus on geometrical concepts. Due to lack of geometrical content in most curricula internationally, however, research is often confined to fairly simple calculations of a spatial nature (Hyde et al., 1990; Hanna et al., 1990; Doolittle and Cleary, 1987). (See also section 3.3.1.5.)

This investigation is aimed at gender differences in achievement in the field of Euclidean geometry, but it is not only interested in a vague general comparison of possible differential average performance, it is also concerned with more specific differences, like differential item performance, or item-bias, as it is often called. Since the item is the most fundamental level of content, Berk (1982) advises that all bias studies should include analysis at item level. It is, however, seldom done in the field of mathematics (Becker, 1990). Usually a variety of questions, ranging from algebraic manipulations and number patterns to calculations of a spatial nature are grouped together to come to a general conclusion about differential gender performance. If the geometry content is separated from other mathematical content in a test, it is usually treated as a test on its own, resulting in an average score for geometry based on a few items (Doolittle and Cleary, 1987). Chipman et al. (1991) state that conclusions which are made on grounds of the few geometry items in the Scholastic Aptitude Test (SAT) are to be blamed for contradicting research results with regard to gender differences.

4.6.1 Differential item functioning

Investigation of item-bias implies gathering **empirical** evidence concerning the relative performances on test items of members of the two or more groups involved (Hambleton et al., 1991; Crocker and Algina, 1986). Empirical evidence is necessary, but not sufficient, to draw the conclusion that bias is present. Differential performance is sometimes a valid reflection of real differences that have to be recognized as such. Boys are, for example, physically stronger than girls of the same age. For these reasons, the expression **differential item functioning (DIF)** is preferred rather than item-bias. Bias and item-bias are terms that usually imply unfairness (Doolittle, 1985), but in achievement tests, DIF can exist without unfairness to any of the subgroups. Differential item functioning is gradually replacing the value-laden term, bias (Hambleton et al., 1991; Bleistein, 1986; Doolittle, 1985; Scheuneman, 1982). In this thesis the term bias will not necessarily be replaced where authors have used it as such.

Definitions

The simplest definition of item bias is the classical one that compares the index of item difficulty of the groups concerned. The problem with this definition is that this index describes not only the item, but also the group that was tested. More recent definitions of DIF or item-bias are increasingly (Owen, 1989) based on Item Response Theory. An important characteristic of these definitions is that a condition is imposed regarding the ability level of the testees. Only testees with the same level of ability are compared. According to Hambleton et al. (1991), the definition of differential item functioning (DIF) accepted by psychometricians, is that

> "an item shows DIF if individuals having the same ability, but from different groups, do not have the same probability of getting the item right." (Hambleton et al., 1991: 110).

4.6.2 The detection of differential item functioning

The following are some of the most important methods or techniques for the detection of biased items or items that cause differential functioning for different groups (Owen, 1989):

* Comparison of difficulty values of items.
* Comparison of discrimination values of items.
* Transformed item difficulty values.
* Analysis of variance.
* Analysis of responses to item distractors.
* Chi-square methods.
* Methods based on Item Response Theory.

The first five methods are based on CTT and subject to the limitations mentioned in section 4.2.6. The chi-square methods, of which Scheunemann's chi-square is popular for studying bias, have some resemblance to Item Response Theory in that ability intervals are taken into account. These are statistically less elegant (Ironson, 1982) than IRT but it offers a good practical alternative. Considerable agreement exists between Scheuneman's chi-square procedures and the methods of IRT (Burrill, 1982), and especially with the three parameter logistic model (Ironson and Subkoviak, 1979). In a comparative study (Subkoviak et al., 1984), the latter method proved to be the most favourable for studying bias. Ironson (1982) is of the same opinion. The fact that IRT is especially suitable to study differential item performance is supported by several researchers (Barnard, 1992; Hambleton et al., 1991; Barnard, 1991; Owen, 1989; Hambleton, 1989; Crocker and Algina, 1986; Bleistein, 1986; Subkoviak et al., 1984; Osterlind, 1983; Burrill, 1982; Ironson, 1982; Warm, 1978; Lord, 1977a).

Problems of high costs and sophisticated software have limited IRT procedures to being used mainly for large investigations such as the standardization of tests and the equating of different examinations on the same scale for promotion and selection purposes. Lately, a few researchers have applied IRT methods to bias studies, but mainly cultural bias (Bleistein,

1986). Studies involving gender differences in the field of mathematics, analyzed by IRT models, are the exception. (See also sections 3.3.1 and 4.6.5.)

4.6.3 DIF - Item Response Theory and Classical Test Theory

One of the desirable features of IRT is that it provides a unified framework for conceptualizing and investigating bias or differential item functioning (DIF) at item level (Hambleton et al., 1991). In most cases the methods of IRT are more sensitive to differences than other methods. Different procedures can be used to investigate DIF within Item Response Theory. The most popular of these procedures is the comparison of Item Response Functions. An item shows DIF if the Item Response Functions across different subgroups are not identical (Hambleton et al., 1991). Since the Item Response Functions are completely determined by their corresponding item parameters, two IRFs can have different shapes only when the item parameters that describe them are different. Comparison of item parameters of the different groups is considered the most direct approach. Other methods within IRT include total area between IRFs and squared or weighted differences. (See also section 5.6.3.)

The three parameter IRF procedure is the most theoretically sound method for measuring differential performance, due to its provision for guessing and its invariance property of estimated parameters (Bleistein, 1986; Crocker and Algina, 1986; Subkoviak et al., 1984; Ironson, 1982). Invariance means that the shape of the IRFs are not dependent on the particular distribution of ability in the sample. The sample invariant nature of the parameter estimates makes it less likely to artifactually label an item as biased when true differences are present (Ironson, 1982).

Criticisms most frequently levelled at IRT are that

* it is not always possible to satisfy the condition of unidimensionality, and multidimensionality may be mistaken for item bias (Crocker and Algina, 1986);

* it is too costly because of sophisticated software and computer time involved (Bleistein, 1986);
* sometimes the process of convergence in estimating the parameters becomes too lengthy to be practical (Ironson, 1982), and
* sufficient sample sizes are problematic and in order to obtain stable, accurate estimates of the parameters, fairly large samples are recommended (Bleistein, 1986).

The present study uses a sample of sufficient size to ensure fairly accurate estimates of parameters. (See section 5.2.) WARMEST, the computer program used, has built-in functions to speed up the process of convergence which brings down the costs involved. The geometry tests that were compiled for this research are believed to be unidimensional in view of the nature of Euclidean geometry. (See Appendices H and I for the test instrument.)

In CTT, differential item functioning of two groups is detected by comparing difficulty values obtained by the respective groups on the same item. If there is a significant difference between the two difficulty values of an item, it is concluded that the two groups responded differently on that item. This argument has shortcomings because, differences in ability between the two groups are hard to eliminate as a confounding factor (Burrill, 1982), and because the distributions of the difficulty values for the two groups are probably not linearly related. Thus, these procedures can lead to erroneous identification of item bias (Barnard, 1992; Warm, 1978). In IRT it is often found that an item favours one group at the lower end of the ability scale and the other group at the higher end of the ability scale (Crocker and Algina, 1986; Lord, 1977b).

Analysis of variance (ANOVA) is one of the most widely used procedures for assessing group bias in the internal structure of a test. According to this method, bias is defined as a group-by-item interaction, indicating that some test items are relatively more difficult for one group than another. Because ANOVA is based on classical item difficulty differences, its use is limited by the problems mentioned in section 4.2.6. It may obscure test bias when

it exists as well as create false impressions of bias (Camilli and Shepard, 1987; Bleistein, 1986; Osterlind, 1983). According to Angoff

> "it is usually the most difficult and the most easy items that are identified as biases for artifactual reasons, due to the choice of methodology and not because such items are inherently discrepant." (Angoff, 1982: 108).

4.6.4 Distractor response analysis

Analysing the responses of different subgroups of the population on distractors is the essential strategy of the distractor response analysis (Osterlind, 1983) and it is a popular way of studying differential item performance. Devising distractors, or wrong answers, among which the correct answer is hidden, is very important (Webb, 1989). It has been mentioned that the primary factor in determining the effectiveness of items as measures of ability, is selecting the distractors (Aiken, 1987). A conventional approach is to give distractors which could reasonably be obtained by various incorrect methods that testees might be expected to employ. Distractors are usually carefully constructed by test developers. When a testee responds incorrectly to a test question, the reaction is seldom pure guessing; incorrect choices usually represent partial knowledge or lack of knowledge (Osterlind, 1983), and often testees' responses to distractors may contain information on misconceptions and errors in reasoning. Analysis of distractor responses is a valuable tool in studying differential item performance of different groups.

> "In an unbiased item a distractor is equally attractive to members of the different groups. (Scheuneman, 1982: 189).

However, Scheuneman warns that distractor attractiveness is a function of the ability of the testee. Groups differing in ability may thus show differing patterns of distractor attractiveness even in unbiased items. She thinks that too great a reliance on distractor data may be misleading.

The distractors of the multiple choice test, used for the present research, are analyzed to obtain additional information on possible differential performance of males and females. (See sections 5.6.3.2 - 5.6.3.5.)

4.6.5 Differential gender item functioning

Gender differences in mathematics achievement are discussed in Chapter 3. Only a few related findings are reported here. Multiple choice testing as such is sometimes to the disadvantage of females (Becker, 1990) and sometimes not (Curtis et al., 1989). Mathematics aptitude tests that produce differential performance in favour of males were adjusted by a model from the Item Response Theory by Young (1991).

Plake et al. (1988) used a modified one-parameter IRT model to investigate the effect of item arrangement (from easy questions to hard questions) in mathematics performance of males and females. They concluded that the differential performance of males and females found in a few items might have been due to chance alone. They did not make use of the three parameter model that makes provision for chance success. Ben-Shakhar and Sinai (1991) found that no correction for guessing may create a small discrimination against females.

A study by Doolittle and Cleary (1987) used a model based on the three parameter logistic model, but adapted for small samples, to analyze their results. Their conclusion that boys did better than girls in geometry is hampered by differential mathematical background in favour of boys and by the small number of items (only eight) used. No deductive reasoning, in an Euclidean sense, was required.

The problems of gender research in mathematics achievement mentioned in this section, such as sample size, number of items and different mathematical backgrounds, are not experienced in the present research. (See section 6.1.4.)

No research on gender differences in the performance on fairly advanced Euclidean geometry (deductive reasoning, proof writing, hypothesis testing), involving analyses on item level, using the three parameter logistic model of IRT, could be traced.

4.7 SUMMARY

Differential item functioning (bias) has been discussed within the framework of Classical Test Theory and Item Response Theory. The advantages and also disadvantages of analysing the differential functioning of testees using the two completely different theories were covered. According to the literature, it seems that the Item Response Theory is the more suitable procedure with which to study differential item functioning. Few studies have used IRT procedures to investigate gender differences at item level in mathematics, and even less in the field of Euclidean geometry.

CHAPTER 5

EMPIRICAL INVESTIGATION

5.1 INTRODUCTION

From the literature study there appears to be a lack of research in general, and empirical research in particular, in the synthetic approach to Euclidean geometry. Most western countries take different approaches to the teaching of geometry on secondary level, as was discussed in section 2.6.3. Several researchers mention the lack of research and the generalized nature of existing research in Euclidean geometry (Clements and Battista, 1992; Hershkowitz et al., 1990; Senk, 1985). There seems to be a need for more specific research into Euclidean geometry with students who are reasonably well prepared in the proof aspects thereof (Senk, 1989). Thus, it was decided to undertake an empirical investigation in South African schools with grade 11 (standard 9) pupils where Euclidean geometry is taught extensively, as compared to other western countries. Pupils at this level have dealt with many theorems and proofs in Euclidean geometry. The main aim of the investigation was to study possible differential gender performance.

Traditionally Euclidean geometry is evaluated in South Africa, especially at more advanced levels, by means of open ended questions. Multiple choice questions were also considered in view of the lack of specific research on a fairly large scale in the field of Euclidean geometry. It has been argued that higher order thinking, which includes deductive reasoning, cannot be evaluated by means of multiple choice questions (items) and that faulty reasoning processes cannot be identified using only multiple choice questions (Prins, 1992). But, on the other hand, multiple choice questions make item analysis possible on a larger scale, and Barnard (1992) claims that the scoring of items is less subjective than that of open ended questions, especially for Euclidean geometry (Barnard, 1993a) where many possible ways of arriving at the correct answer exist. Faulty reasoning processes are unlikely to lead

to correct answers if the distractors are chosen with care (Barnard, 1991). According to Van den Berg (1994), empirical research consistently indicated that items can be used to a large extent to measure the outcomes measured by traditional written examinations. If possible, both ways of testing are advisable (Barnard, 1993a; Aiken, 1987). Two tests in Euclidean geometry were therefore used, one consisting of multiple choice questions and the other of open ended questions. (See section 5.3 for a description of the tests.)

5.2 SAMPLE

The sample consisted of 816 grade 11 (standard 9) pupils from 5 different Transvaal Education Department (TED) high schools. The schools are located on the West and North West side of Johannesburg (part of th luteng province since 1995) and are more or less from the same socio-economic environment. The quality of mathematics teaching in all five school is considered generally good. Three schools are English medium and two schools are Afrikaans medium. The composition of the school population across the schools was fairly constant at the time of the empirical study (1993) and predominantly white. The 816 pupils were taught in 33 different classes by 15 teachers, 14 females and 1 male. More schools were approached for participation, but some could not obtain the necessary permission from the headmaster or from the governing body of the school. After careful consultation with the heads of mathematics departments and teachers involved, it was agreed to use the geometry tests (compiled for this research) as part of the end of the year examinations to ensure full cooperation from the pupils. It was important for the study that the pupils were well prepared for the tests and motivated to do their best.

In TED schools, mathematics is offered on three levels in grade 11, i.e., higher grade, standard grade and lower grade, but in practice very few pupils select lower grade. Standard grade mathematics is of a little lower level than higher grade mathematics, but standard grade pupils are not necessarily the weaker pupils, because it is essentially a matter of choice. In grade 11 most higher grade and standard grade pupils are taught in mixed classes because of similarity of mathematical content in the syllabus and it was therefore not considered necessary to make a distinction between the two levels. Except for a small part of the Euclidean geometry syllabus for grade 11 (standard 9), which could be avoided in the

investigation, all pupils virtually did the same Euclidean geometry. Problems concerning differential gender exposure to mathematics and Euclidean geometry, which is quite prominent in the literature on gender differences (see section 3.4), were thus not a confounding factor in the present research.

The ratio of males to females taking mathematics as a subject differed little within the different schools and for the overall sample there were 410 males and 406 females, approximately 17 to 17,5 years old. This is a favourable situation for educational research in view of the problems mentioned in sections 3.1 and 3.4 relating to the under-representation of females in most samples of gender studies in the field of mathematics and geometry.

Originally it was planned to include a sample of black students as well, but due to problems in traditionally black schools at the time, this could not be realized. A third Afrikaans school was also part of the initial sample, but because the tests were not administered under examination conditions, their result were excluded. It was clear that the pupils in this school did not take the tests seriously, probably because the marks did not count as part of the promotion mark.

5.3 TEST INSTRUMENT

Two tests, based on the **Euclidean** geometry curriculum for grade 11 were compiled by the researcher. The content covered by the tests includes parallel lines, perpendicular lines, equilateral triangles, parallelograms, congruency and the circle theorems, including midpoint, cyclic quadrilaterals and tangents to a circle. The questions range from relatively easy, involving calculations, to quite difficult questions involving multi-step deductive reasoning. All diagrams were supplied.

The first test consisted of 20 four-choice questions, mostly taken from the standardized tests of the Human Sciences Research Council (HSRC). Each question consisted of a stem and four choices. The items were scored dichotomously, i.e., either right or wrong, and therefore the maximum score for the test was 20 marks. The second test, with a maximum

score of 54, consisted of open ended questions. The questions included one theorem, some calculations and several geometry riders. The riders, dependent on 'proof writing' as it is called in the literature, represented approximately 60% of the total score of this test and ranged from those fairly similar to the usual textbook problem to those completely different from them. They were typical of the fourth van Hiele level of formal deduction. The riders were curriculum orientated, but definitely required original, creative thinking. The more difficult ones had probably not been encountered by the pupils before. Both tests covered the same geometrical content. (See Appendices H and I for these tests.)

Before the experiment started, the tests were used in a pilot study which led to refinement of some of the questions and to a better time allocation for both tests. For the multiple choice test 35 minutes were allowed and one hour for the open ended questions. Almost all the students had enough time to complete the tests and therefore the tests can be considered power tests.

5.4 PROCEDURE

All tests were administered during November and early December 1993. Four schools used both tests as part of their end of the year mathematics examinations. One school used the geometry riders as part of the examination and the multiple choice test as a revision test just before the examination. The two tests were written on different days but in the same order, except for the one school that used one of the tests as a revision test. All test material was treated confidentially by the teachers. After scoring the papers for promotion purposes, the answer scripts were given to the researcher.

Marking

The entire sample's answer sheets (n = 816) for the multiple choice test were marked and scored by computer. Item and statistical analyses were done with the aid of computer facilities at the Human Sciences Research Council (HSRC). The SAS computer package was used for the analysis within the framework of the Classical Test Theory and the program WARMEST for the calibration of the parameters of the one dimensional three parameter

logistic model of the Item Response Theory. The program WARMEST uses the weighted likelihood estimation algorithm while keeping the maximum likelihood and the Bayesian procedures as options.

The answer scripts from the five different schools, consisting of the open ended questions on Euclidean geometry were separated for the two genders. About one quarter of each group was randomly selected for remarking by the researcher. Marks were allocated in the same way as for the school leaving examination at the end of grade 12. Not too much emphasis was placed on prescribed format. The emphasis was on correct geometrical reasoning and not on format or reasons for every statement. Odd statements, correct according to the diagram, but not leading on to a conclusion gained no marks. Those resulting in a correct conclusion or leading to a possible correct answer, were given some credit. An 'all or nothing' approach was thus not adopted. A high degree of consistency was guaranteed because all the questions of the selected answer scripts were marked by the researcher. Any correct solution was credited.

Evidence of the following was also noted:

* Logical geometrical reasoning.
* Partial logical geometrical reasoning.
* Rote memorization.
* Assumptions of the result in the proof thereof.
* Being misled by the perceptual features of a diagram.
* True yet redundant statements.
* Simplification using specific values.
* Omission of questions.

The above is not an attempt to classify all possible ways of reasoning or of making the usual mistakes in solving geometry riders, it only describes the most obvious tendencies in reasoning with regard to Euclidean geometry which might possibly show gender differences. (See section 5.7.2 for more details.)

5.5 RESULTS OF EMPIRICAL RESEARCH ON TEST LEVEL

5.5.1 Mean scores

The first step in the analysis was to compare the mean scores that males and females obtained on application of the Euclidean geometry tests. The marks were obtained on the 20 multiple choice questions (Appendix H) by the whole sample of 816 pupils and on a test of open ended questions (Appendix I) by a sub-sample of 203 pupils. The results are given in the following sections.

5.5.1.1 Combined tests

The means of the combined test (a combination of the test consisting of multiple choice questions and the test consisting of open ended questions) showed almost equal performance for the two gender groups. The mean score of the males was 53,5% (n = 410) and that for females 53,3% (n = 406). The respective means were calculated by dividing the obtained mark by the possible mark. For about 25% of the above sample the highest possible score was 74 (multiple choice questions contributing 20 and open ended questions 54). The other 75% of the sample had a maximum possible score of 20 for the multiple choice items.

Subsequently the two components of the measuring instrument were considered separately in order to investigate possible differential gender performance in Euclidean geometry.

5.5.1.2 Multiple choice test

The mean scores of all the males and females in the sample on the multiple choice test are compared in table 5.1. The total score is 20.

TABLE 5.1: Multiple choice test: Mean scores

Geometry test: multiple choice questions			
	Group	Male	Female
n	816	410	406
$\bar{X}$	11.06 (55.30%)	11.09 (55,45%)	11.04 (55,20%)
SD	4,26	4,29	4,24
SE	1.77	1,78	1,75
KR-14	0,83	0,83	0,83
KR-8	0,83	0,83	0,83
KR-20	0,81	0,81	0,81

(n is the number of pupils, $\bar{X}$ is the mean score, SD is the standard deviation, SE is the standard error and KR-8, KR-14 and KR-20 indicate values obtained from the respective Kuder Richardson formulas - See Appendices A, B and C.)

A T-test was done to evaluate the significance of the difference in the means of males and females (T = 0,17). The difference of 0,05 (0,25%) in favour of males is **not** a significant difference on the 1% level. The standard deviations of the two genders, as well as the standard errors of measurement, are also quite close.

Reliability

The reliability of the test was estimated using Kuder-Richardson formulas (see section 4.2.5). The test shows fairly good reliability for the sample as a whole as well as for the males and females separately. A value of 0,83 for reliability can be regarded as acceptable and an indication of good reliability.

Validity

All the teachers involved in the experiment (n = 15) agreed that the tests covered the prescribed syllabus for Euclidean geometry for grade 11 and were of appropriate cognitive level. (See section 5.3 for details of content.) Because the tests were considered to have content validity, they were used as part of the end of the year examination. Although content validity is subjective (see section 4.2.4), it is important in scholastic achievement tests.

5.5.1.3 Open ended test

The open ended questions consisted of geometry riders ranging from calculations to those involving difficult deductive reasoning and hypothesis testing. The riders were scored according to the method described in section 5.4. The total score for the test was 54.

Table 5.2: Open ended test: Mean scores

Geometry: Open ended questions		
	Males	Females
n	101	102
$\overline{X}$	27,52 (50,96%)	27,45 (50,83%)
SD	15,05	14,73

(n is the number of students, $\overline{X}$ the mean score and SD the standard deviation - see Appendix G.)

The difference of 0,13% in mean score on geometry riders is **not** (T = 0,03) statistically significant on the 1% level.

The ratio of the males' SD to females' SD is 1,02 which is indicative of homogeneity of variance. According to Feingold (1994) values between 0,90 and 1,10 are acceptable for this purpose.

5.5.1.4 Instrument correlation

To determine and interpret the interrelationship between the two geometry instruments of different format, correlations were calculated. These correlations are summarized in table 5.3. The end of the year mathematics marks, which also included algebra and trigonometry components, were available for all students and these were also correlated with the geometry test scores.

Table 5.3: Correlations: Multiple choice questions / open ended questions / examination marks

Correlations						
	multiple choice / open ended		multiple choice / examination		open ended / examination	
	n	r	n	r	n	r
Group	203	0,73	816	0,61	203	0,66
Males	101	0,77	410	0,64	101	0,68
Females	102	0,71	406	0,62	102	0,66

(n is the number of students and r is the Pearson's product-moment correlation coefficient.)

The relatively high correlations between the multiple choice test and the open ended test, showed that the two instruments are related to a fairly high degree. The correlation between both instruments and the marks obtained in the examination is also fairly high. Note that the five schools participating in the experiment did not write the same examination papers in mathematics.

The mean for Euclidean geometry only is not a measure of spatial ability as is defined in section 2.2. It is worth noting, however, that the correlation between Euclidean geometry and mathematics in general for this experiment (table 5.3) is between 0,6 and 0,7. It is interesting to note that the correlation for males is a bit higher than that for females.

5.5.2 Discussion

Gender differences reported in mathematics achievement are often observed from early adolescence (Hyde et al., 1990; Swanepoel 1990; Hyde, 1981; Fennema and Carpenter, 1981; Maccoby and Jacklin, 1974). Gender differences in favour of boys have been reported in the following fields:

*	Spatial skills	Johnson and Meade, 1987; Linn and Petersen, 1985; Pattison and Grieve, 1984; McGee, 1979; Maccoby and Jacklin, 1974;
*	Geometry	Frost et al., 1994; Battista, 1990; Hanna et al., 1990; Hyde et al., 1990; Swanepoel, 1990; Doolittle and Cleary, 1987; Fennema and Carpenter, 1981;
*	High cognitive levels	Benbow, 1992; Benbow and Stanley, 1982; Fennema and Carpenter, 1981;
*	Problem solving	Hyde et al., 1990; Armstrong, 1981;
*	Mathematical reasoning	Benbow, 1992; Benbow and Stanley, 1982;
*	Visualization	Battista, 1990.

Given these reported differences, one might expect significant gender differences in solving problems in Euclidean geometry which require some spatial ability and visualization. It involves mathematical reasoning and it qualifies as a higher level cognitive task. It is often only within reach of the high performers, and is considered among the most difficult processes to learn in the secondary school curriculum.

The lack of any significant differential gender performance in mean scores on the geometry achievement tests used in the present research contradicts the above findings. Several explanations will be offered further on.

Problems with existing research (see chapter 3) with regard to gender differences in mathematics achievement include:

* Research is too general to be meaningful.
* Mathematics is multi-dimensional and gender differences vary across the more homogeneous processes within.
* Spatial abilities appear to have differential gender influence on mathematics achievement.
* Samples are not fairly homogeneous (e.g. contain different cultural groups) and may counter-balance gender differences in performance.
* Cross-cultural comparisons often lead to contradictions because of large cultural differences in mathematics achievement.
* Differential gender exposure to previous mathematics courses cannot be controlled sufficiently.
* Research on problem solving often includes a variety of problems, some with a high verbal content.
* Meta-analytical studies include diverse samples and diverse mathematical content.
* Samples that include college entrance tests are self selective.
* Similar mean scores may conceal gender differences on different ability levels.
* Similar mean scores may conceal gender differences in variability.

The research quoted above is characterized by one or more of these problems. It is thus possible that a study, such as the present one, which controls some of these problems, will have contradictory findings. The outcome of the present research is, however, supported by the findings of Senk and Usiskin (1983) in the United States and is also partially related to Tartre's (1990) results. The first-mentioned found equal ability among males and females to write geometry proofs. The difficulty level of the content in their test was a bit lower than that of the test on proof writing used in this research. Their research was complicated by females having less geometry knowledge when entering the high school geometry course. The present study did not use a pre-test because males and females (grade 11) in the sample of the present study have theoretically the same mathematical background due to the school system in South Africa.

Tartre (1990) found no significant gender differences in spatial orientation skills. In section 2.5 it was suggested that some spatial orientation skills relate to the visual aspects of Euclidean geometry.

Battista (1990) suggests that girls and boys perform equally well on logical reasoning and geometric problem solving strategies. The high verbal content of the instrument measuring logical reasoning and the mixture of geometrical content complicates comparison, but Battista's findings may be considered as partially supportive of the present study.

Absence of gender differences in mean scores may sometimes conceal greater variability of one of the groups (Feingold, 1994), which actually represents a gender difference. For the present research the ratios of male variance to female variance are 0,99 and 1,02 for the multiple choice questions and the open ended questions respectively. Values so close to 1 represent similar variability (Feingold, 1994) of the two gender groups with regard to performance in Euclidean geometry.

Comparison of the frequencies of the raw scores of males (n = 410) and females (n = 406) on the multiple choice questions, shows the following results:

*	Performances of 40% and below	31.0% of the males 30,8% of the females

*	Performances above 80%	12,9% of the males
		12,6% of the females

No substantial gender differences in the geometry scores is shown by either the high performers or the ones who obtained a pass mark. This is partly in contrast with the conclusions of Benbow (1992) and Benbow and Stanley (1982), but the male advantage that they found for high performers in mathematical reasoning was shown (Becker, 1990) to be in the field of word problems. Since no word problems were included in this research, comparison seems unnecessary. The absence of significant gender differences in the performance of males and females in geometry in the present research confirms the re-analysis by Becker (1990) of the data that Benbow and Stanley (1982) used for their study.

5.6 RESULTS OF EMPIRICAL RESEARCH ON ITEM LEVEL:

5.6.1 Item Response Theory: Item parameter estimates

The previous comparison of the mean performances of males and females in Euclidean geometry did not disclose any significant gender differences. The mean scores on both tests and the standard deviations show a large degree of resemblance. Classical statistical methods used in conjunction with both multiple choice items and open ended questions in Euclidean geometry have not detected any notable gender differences. The problem with such analysis is that it might exclude more specific gender differences in performance because the test as a whole is taken as point of departure.

Item analysis was thus indicated as a means to find out whether there were any differences between the performance of males and females on item level. But one has to be cautious in using traditional methods of item analysis (Barnard, 1992) as was discussed in chapter 4. The performance of testees is a statistical concept, while the difficulty value of an item is a psychometric concept and these two are not directly comparable. A difficulty value of 0,6 is interpreted as 60% of the testees obtaining the correct answer for the specific item. A performance of 60% on a test as a whole, could indicate that a student mastered 60% of the

content. This can be considered as two identical values, but on different scales and therefore not directly associated.

The **Item Response Theory** (see section 4.3) was developed to overcome such limitations. The performance (ability) of a testee and the difficulty value (**b**) of an item can be expressed on the same scale. The one dimensional three parameter logistic model of Item Response Theory, which was used in the following analysis of the **multiple-choice questions**, provides a very sensitive way of looking at differences on item level.

The estimation of the item parameters was done separately for males, for females and for the group as a whole according to the weighted likelihood estimation method. The computer program WARMEST (see section 5.4) was used to calibrate the items. See Appendices A, B and C for the printouts. An example of such a printout is given in Table 5.4. For each of the three groups and for each of the 20 items the following data are given:

* In the first two columns the item number and the key to the correct answer are given.
* The **a** parameter which indicates the degree to which the item discriminates between high and low performances. The value of **a** is proportional to the slope of the IRF and thus varies according to the level of ability. The value that appears on the printout is the maximum value of **a**, i.e. the **a** that corresponds to the **b** parameter of the item. (For the three parameter logistic model the **a** parameter is found in the region of the **b** parameter, because the value of **c** causes a bit of distortion.)
* The value of the **b** parameter which is an indication of how difficult the item was for the specific group of testees. Since the origin and unit of measurement are arbitrary, it is common practice to set the mean ability to 0 and the unit of measurement to 1. Note that the **b** parameter is found on the horizontal axis which represents **ability** as measured by the specific item.
* The **c** parameter, which is a measure of the degree of guessing that has taken place.
* The **p** value, i.e. the CTT (Classical Test Theory) difficulty index.
* The **d** value, i.e. the CTT discrimination index.
* FITABC, which is a measure of how well the three parameter logistic model of the IRT fits the actual data.

Table 5.4: Item parameter estimates (from Appendix B)

WARMEST: ITEM PA ETER ESTIMATES

TESTID=CRON GRPS..=MAAL N= 410 #ITEMS= 20 MEAN_Z=0.000 SD_Z= 1.000 METHODS: WLE(THETA)

ITEM	KEY	A	B	C	PVAL	DVAL	FITABC	BME(A,B,C)
1	4	0.380	-1.044	0.206	0.724	0.262	0.069	
2	1	0.957	1.199	0.168	0.329	0.419	0.108	
3	2	1.176	-0.798	0.170	0.754	0.495	0.133	
4	3	0.850	0.375	0.175	0.500	0.472	0.283	
5	4	0.817	0.011	0.201	0.595	0.435	0.325	
6	3	1.261	0.791	0.130	0.346	0.512	0.185	
7	2	1.387	-0.237	0.169	0.639	0.565	0.208	
8	2	0.674	-2.023	0.197	0.895	0.267	0.108	
9	2	0.960	0.557	0.156	0.446	0.481	0.045	
10	3	1.050	-0.842	0.199	0.785	0.447	0.084	
11	1	1.191	0.800	0.188	0.398	0.453	0.192	
12	3	1.031	0.852	0.170	0.383	0.468	0.045	
13	3	1.380	-0.217	0.127	0.612	0.575	0.092	
14	2	1.009	0.692	0.167	0.412	0.470	0.035	
15	1	1.474	0.777	0.112	0.324	0.531	0.260	
16	4	1.159	-0.487	0.139	0.685	0.506	0.367	
17	1	1.093	-0.928	0.172	0.788	0.433	0.046	
18	4	1.082	0.656	0.093	0.359	0.537	0.258	
19	3	1.468	-0.599	0.110	0.712	0.563	0.591	
20	1	0.653	1.297	0.220	0.400	0.326	0.108	

The values of the parameters have been discussed in section 4.3.5. Note how well the model that was chosen, fits the data. Values less than 1 (Barnard, 1992; 1991) are considered acceptable. Approximately 90% and 85% of the FITABC values are less than 0,3 for females and males respectively. The data for the group as a whole has 80% of values less than 0,4. No values of FITABC exceeds 0,6. According to Hambleton (1989) a good fit is essential before conclusions can be drawn.

It is important to note that, although the three ability means (for males, females and the whole group) are each equal to 0, they are independent of each other since they are arbitrary. (See Appendices A, B and C.) During the calibration of items the parameter estimates are placed on a scale that is defined by the mean and standard deviation of the specific sample of testees. Before the parameters can be compared, they must be transformed to a common scale. The relationship between the scales derived from two different calibrations is linear because they differ only in origin and unit of measurement (Barnard, 1993b; 1992; Hambleton et al., 1991). It was decided to transform the parameters for males onto the scale of females. The transformation was done with the computer program MINEQUAT (see Appendix D). These values could then be directly compared to the parameters for females (Appendix C). The values could have been mapped the other way round with exactly the same outcome (Barnard 1993b; Hambleton et al., 1991).

5.6.2 Item Response Theory: Numerical values of parameters

If the numerical values of the **b** parameters of the males (Appendix D) and females (Appendix C) are compared, it is clear that for 13 of the 20 items, females have higher indices than males. Note that a greater **b** parameter, translates into a higher position on the ability scale. That means that 13 items were relatively more difficult for females than for males, though some of the differences are small. Eleven of the **a** parameters are higher for females which means that 11 items discriminated better (between high and low performers) for females than for males. The means of the **a**, **b** and **c** parameters, calculated for 20 items, are given in table 5.5. Although differences are small, the mean values for **b** show that the test was in general of above average difficulty for females and just below average difficulty for males. The test discriminates better for females than for males. The means of the **c**

parameter are rather low which means that little guessing took place. Because each item consisted of a stem and four choices, the statistical random chance for guessing was 0,25. For males not one of the items (Appendix B) had an index for guessing greater than 0,25 and for females (Appendix C) only one value of **c** exceeded that. This implies that guessing on the multiple choice questions was not a confounding factor in this study. Omission of items did not play a differential role either. The average percentage for questions omitted was 0,7% and 0,8% for males and females respectively.

The difference in the mean values of the **b** parameter is 0,11 on a scale from 0 to 1. It is important to note that even small differences in the values of the **b** parameter can translate into actual differences, as is evident from the item analysis in section 5.6.3.2. Becker (1990) refers to an average difference of 0,13 as considerable.

Table 5.5: Item Response Theory: Means of parameters

Means: Item parameters		
Parameters	Male	Female
a	1,14	1,33
b	-0,01	+0,10
c	0,16	0,17

5.6.3 Item response theory: Differential item performance

5.6.3.1 Introduction

In section 3.4 it has been argued that the absence of mean differences in the performance of different groups is not sufficient evidence for a conclusion that such differences do not exist (Feingold, 1994; Becker, 1990; Doolittle and Cleary, 1987). Becker and Forsyth (1994) indicated an interaction between gender differences in mathematical problem solving and ability levels of the students. This argument is supported by the investigation of Tartre

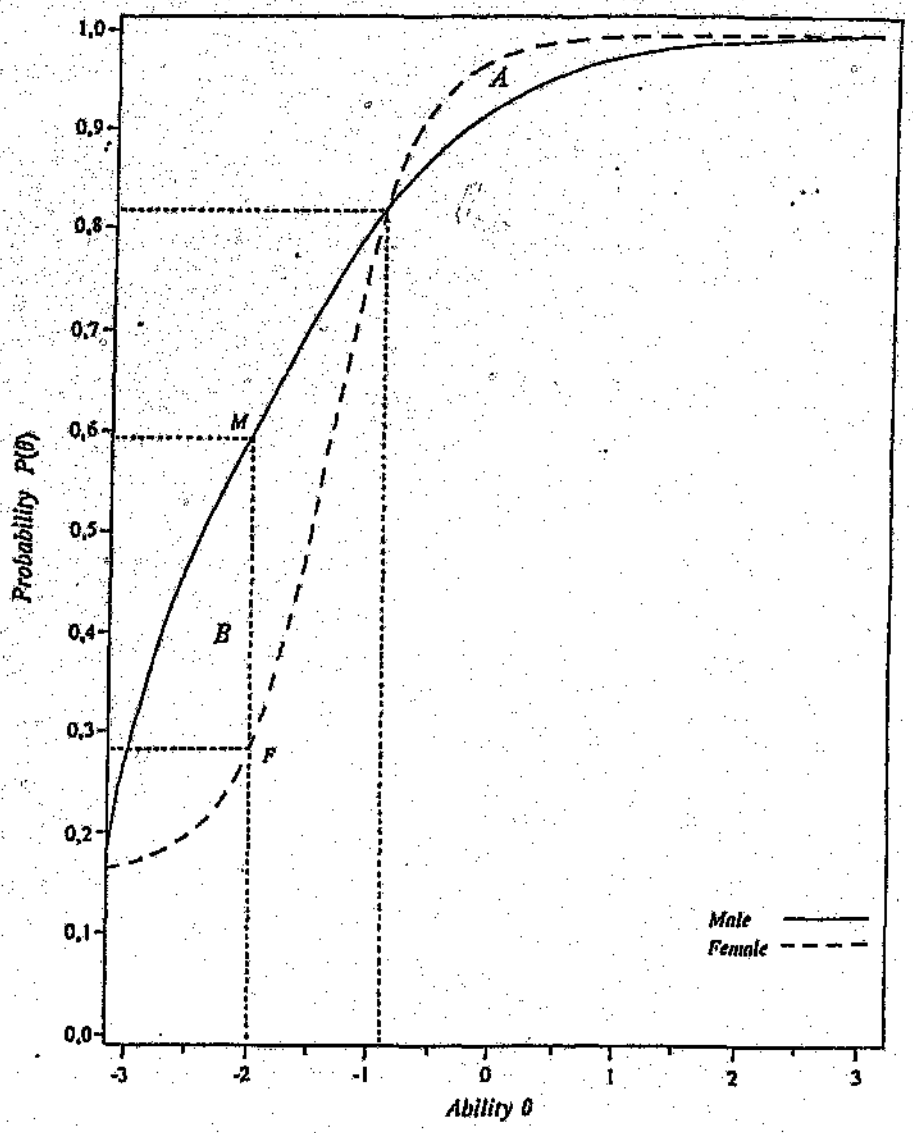
(1990). It appears that examining male and female performance at different ability levels reveals more information on the relationship between the gender of students and academic performance than is achieved by merely considering group mean scores, either on a complete test or on a single item. Brusselmans-Dehairs and Henry (1994) advise the use of models from the Item Response Theory to overcome these problems.

It was also suggested in section 4.6.2, that methods based on the Item Response Theory are very sensitive in identifying differential item performance. But differential item functioning (DIF) is a complex concept within the framework of the Item Response Theory. Barnard (1991) and Hambleton et al. (1991) say that an item shows DIF if individuals, having the same ability, but from different groups, do not have the same probability of obtaining the correct answer for the item. Another way of considering differential item performance is by comparison of the area between the Item Response Functions - IRFs (see section 4.3.5) of the two groups (Barnard, 1991). It can be best illustrated by means of graphical representation.

Figure 5.1 represents the IRFs of item 8 which was identified as producing differential item functioning in the present research. Female F is functioning at ability level -2,0 and her probability of succeeding on item 8 is about 0,28. Male M is functioning at a similar ability level but his chances of getting this item right is almost 0,6, which is much higher than that for female F. The same is clearly not true for females functioning at average ability level, namely, 0. Their probability of obtaining the correct answer is higher than that of males of average ability. At an ability level of -0,8 males and females have the same probability of being successful on this item. This kind of differential item functioning is called non-uniform (Hambleton et al., 1991) because it is not in favour of one group over all the ability levels. If the area between the curves for males and females is considered, it is clear that area B is larger than area A, which means that the item functions in favour of males more often than in favour of females. Formulas to calculate such areas have been developed, but they are not reliable for the three parameter logistic model. When the **c** parameters are not the same, the area B, for example, will be infinite if it calculated over the entire range of ability, i.e. from $-\infty$ to $+\infty$ (Hambleton et al., 1991). Sometimes the area is calculated within reasonable boundaries, for example, $-3 < b < 3$.

Figure 5.1: Item Response Function -

Differential item performance of males and females



The area between the IRFs is related to the values of the parameters. A less complicated way (Barnard, 1991) of detecting differential item functioning, based on the values of the parameters only, has been developed by Warm (1978). The latter method was used in the present research to identify items that function differently in favour of one of the two gender groups.

The absolute values of the differences between the **a**, **b** and **c** parameters for the two groups respectively are calculated and denoted by d(a), d(b) and d(c). An item is biased, or results in differential item performance, if it meets any **one or more** of the following four conditions:

$$d(a) > 0{,}8$$
$$d(b) > 0{,}5$$
$$d(c) > 0{,}15$$
$$d(a) + d(b) > 1{,}00$$

The degree to which an item favours one group or the other is determined by the conditions that it satisfies, in the following order:

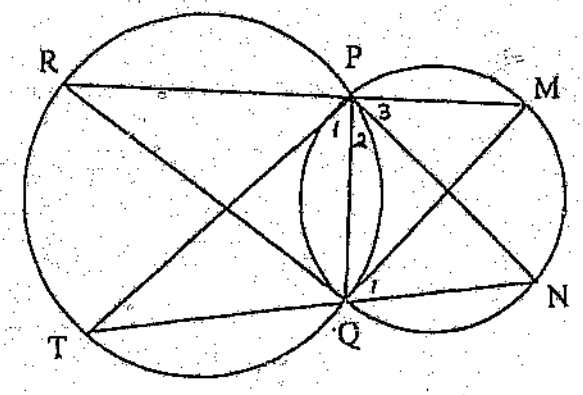
d(a), d(b) and d(a) + d(b)
d(b) and d(a) + d(b)
d(a) and d(a) + d(b)
d(b)
d(a)

Note that the values of the **c** parameter are ignored because guessing was not really a significant factor according to the data obtained in the present empirical research (see appendices A, B, C, and D). This is contradictory to Ben-Shakhar and Sinai's (1991) conclusion that no correction for guessing may create a small discrimination against females.

Four of the 20 items of the multiple choice test (Appendix H) can be identified as resulting in differential item performance with respect to gender. They are items 8, 4, 11 and 12 in order of the most to the least biased. These four items will now be discussed in more detail. The distractors to these four items are also considered to supply more information on differential item performance. (See appendices E and F for distractors.)

5.6.3.2 Item 8

The circles do not have equal diameters.
Which statement is wrong ?



A. $\hat{R} = \hat{T}$
B. $\hat{P}_1 = \hat{P}_2$
C. $\hat{M} = \hat{N}$
D. $\hat{Q}_1 = \hat{P}_3$

Table 5.6: Values of item parameters for item 8

Parameter	a	b	c
Male	0,729	-1,924	0,197
Female	1,668	-1,355	0,163
d	0,939 *	0,569 *	0,034
d(a) + d(b)	1,508 *		

(**d** is the absolute value of the differences.)

This item satisfies (marked *) three conditions (set out in section 5.6.3.1) and is thus the most liable to produce differential item performance. The item was experienced as much more difficult by females than by males . It differentiated much better for females with respect to low and high performers. The difference in the degree of guessing that took place for males and for females respectively is too little to be of importance. Generally, guessing did not take place to a great extent for either of the gender groups.

The question appears to be relatively easy as the negative b values indicate. The fact that the statement of the question is in the negative, may have had an influence, although the only other question in negative form is question 13, which shows small differences, namely

d(a) = 0,148 and d(b) = 0,038. The girls found question 13 only slightly more difficult than the boys.

Item 8: Distractors (Table 5.7)

The high frequencies of both males and females on the correct response (*) is in agreement with the negative values for the **b** parameter. The item was relative easy for both groups. The distractors show a similar pattern for both groups and no signs of gender differences could be identified.

Note that the total frequency of females in appendix F is 407. The answer sheet of one female was damaged and was ignored for the rest of the analysis.

Item 8: Item Response Function (Figure 5.2)

The Item Response Functions of males and females for item 8 provide important information that cannot be deduced from the item parameters alone (Table 5.6) or from the key and the distractors in (Table 5.7). The item features rather low on the ability scale as is seen from the negative values of the **b** parameter and also from the position of both curves, rather to the left of the average, viz. 0,00 on the horizontal axis. The lowest performing females found item 8 relatively much more difficult than low performing males; that is true up to ability level -1,00 from where females outperform males slightly, except for the top performers of both gender groups that performed equally well on this item.

If only the average performance on this item is considered, it seems that both groups performed almost equally well (Appendices E and F). Nor is there any indication in the performance on the different distractors (Table 5.7) that suggests any difference in performance of the two gender groups.

However, when studying the Item Response Functions of both groups, clear differential gender performance shows. A large proportion of males who generally perform low, answered this item correctly while the same is not true for females. Item 8 was the easiest item for both gender groups, but the low performing females found the item more difficult than low performing males.

Figure 5.2: Item Response Functions for males and females - item 8

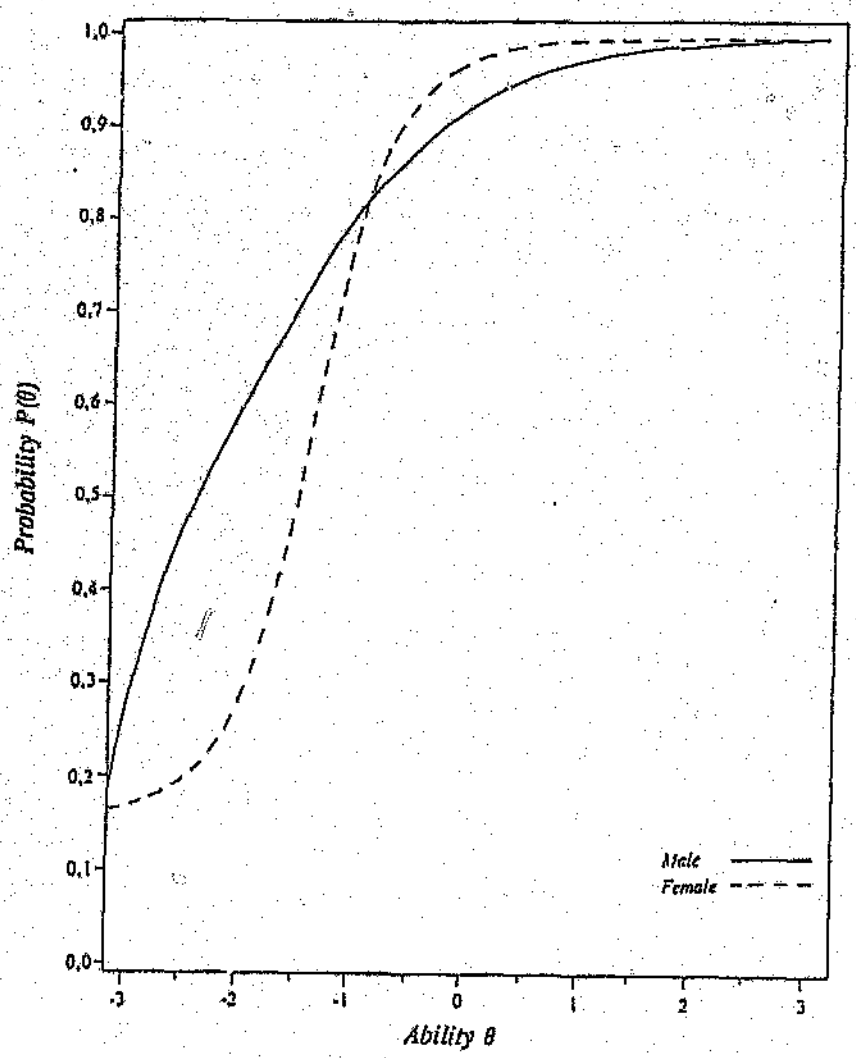


Table 5.7: Frequencies of key and distractors for item 8

Key and Distractors	Males	Females
$\hat{R} = \hat{T}$	5	5
$\hat{P}_1 = \hat{P}_2$ *	367	365
$\hat{M} = \hat{N}$	9	13
$\hat{Q}_1 = \hat{P}_3$	26	24

(* Indicates the correct response (key) - Appendices E and F)

The rather convex shape of the Item Response Function for males indicates a low discrimination value of the item for males. In other words, the item was too easy for males to have a good discrimination value. It seems that a fairly easy Euclidean geometry item, that has a high success rate (90% for males and 89% for females), nevertheless poses problems to low performing females. It discriminates well between low and average performing females while the same is not true for males. It appears that low performing females thus do not have the ability to identify 'angles in the same circle segment' from a simple diagram.

The solution of this item does not require deductive geometrical thinking but depends on visual interpretation of the diagram. Visual interpretations of simple diagrams, usually quadrilaterals, are related to the first two van Hiele levels (section 2.6.5) of thinking but research on these levels does not report on gender differences (Teppo, 1991; Usiskin and Senk, 1990; Smith and de Villiers, 1990; Human et al., 1989; Senk, 1989). However, neither exposure to formal geometry (De Villiers, 1987; Burger and Shaughnessy, 1986; Mayberry, 1983) nor age or grade (Burger and Shaughnessy, 1985) guarantees progress beyond the first two van Hiele levels. The greater inability of low performing females (as compared to low performing males) to visually interpret the diagram of item 8 may be rooted in insufficient informal geometrical experiences usually associated with the first two van Hiele levels. In section 2.6.4 several problems regarding a too formal approach to the teaching of Euclidean geometry in secondary schools in South Africa are discussed.

5.6.3.3 Item 4

O is the centre of the circle.

PT = PR

Express x in terms of y.

A. x = ¼y

B. x = 2y

C. x = 4y

D. x = ½y

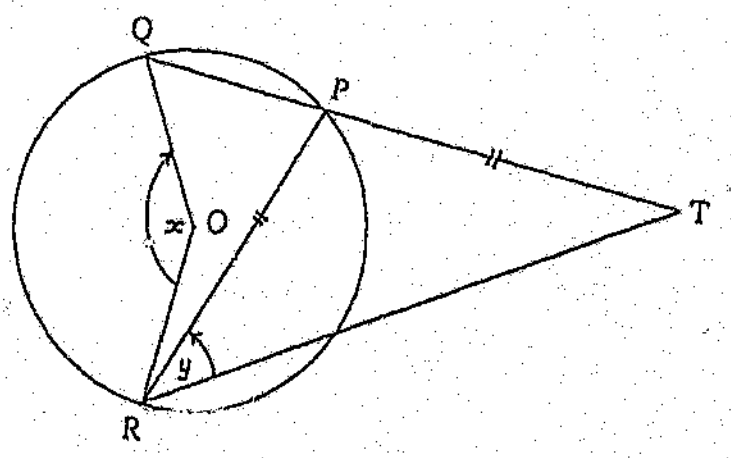


Table 5.8: Values of item parameters for item 4

Parameter	a	b	c
Male	0,919	0,294	0,175
Female	1,744	0,773	0,164
d	0,825 *	0,479	0,011
d(a) + d(b)	1,304 *		

(**d** is the absolute value of the differences.)

This item is characterized by differential item functioning (*) on grounds of two of the criteria. The lower **b** parameter of the males implies that the item was easier for males than for females. To find an explanation is not that simple. The question has an algebraic component to it as well as the geometrical one. It was not lack of knowledge of the theorem that handicapped the girls, because on the items that tested knowledge of the theorem (items 3, 5 and 6 - Appendix H) no gender differences were displayed and the overall performance was quite good.

Item 4: Distractors (Table 5.9)

It is interesting to note that the inverse algebraic process (displayed in the first distractor), was not as attractive as the other distractors. Females have a higher frequency on the second distractor than on the correct answer and males also found this distractor attractive. See the next paragraph for the discussion.

Item 4: Item Response Function (Figure 5.3)

The Item Response Functions of males and females for item 4 show differential gender performance in favour of most of the males. Both graphs are more to the right on the ability axis than those for item 8, which means that both groups experienced this item more difficult than item 8.

Figure 5.3: Item Response Functions for males and females - Item 4

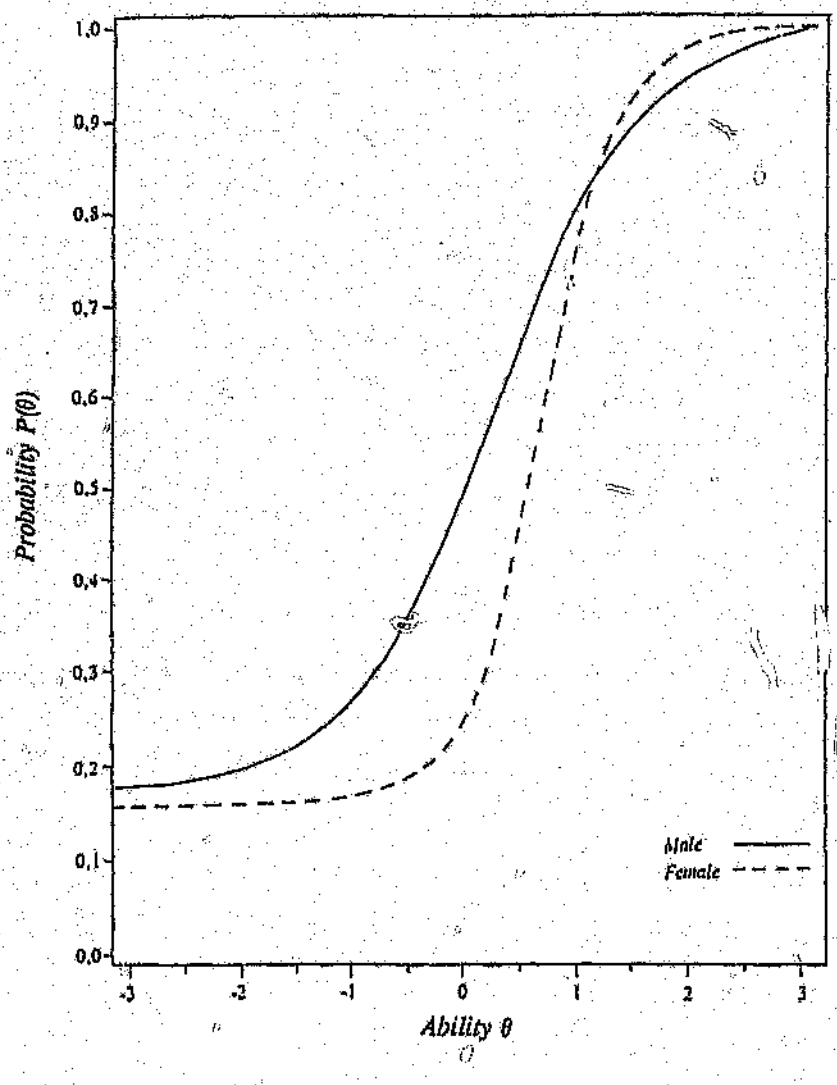


Table 5.9: Frequencies of key and distractors for item 4

Key and Distractors	Males	Females
x = ¼y	21	15
x = 2y	145	165
x = 4y *	205	156
x = ½y	37	70

(* indicates the correct response - Appendices E and F)

The weakest males, as well as females, were unsuccessful, but the females almost stayed on guessing level up to average ability, whereas the males were fairly successful at the average level of ability for this item. As in the previous item, the high performing females outperformed the males slightly, but in this case they only caught up with males at ability level 1,0. All other females experienced this item as more difficult than males did.

A study of both the Item Response Function of females and their responses on the distractors (Table 5.9), seems to suggest that the low performing females did not guess randomly because of the low number of responses on the first distractor. If algebraic mistakes in calculating the size of the angle were the females' main problem, a higher frequency on the first distractor, which is the inverse of the correct response, is expected. Item 6 (Appendix H), involving the same midpoint theorem and which also contains an algebraic component, shows no differential gender performance. One also expects females of average and above average ability levels to be able to perform these algebraic operations which are computational in nature and sometimes show gender differences in favour of females (Frost et al., 1994; Hyde et al., 1990; Kimball, 1989; Doolittle and Cleary, 1987).

Further possibilities that seem likely because of a higher frequency on the second distractor than on the correct answer, are the following. Females might have been constrained by geometrical rigidity more than is the case for males. Hoz (1981) refers to the inability of pupils to perceive a diagram in a different way as the perceptual component of geometrical rigidity. (See section 2.6.4.) Pupils know that x must be twice something, but they are not flexible enough in their geometrical thinking to pursue the idea. Add the frequencies on the second and fourth distractors (also something equal to twice something else) and these account for 57% of the females' answers. Hoz (1981) did not report on genders separately. A related deficiency is that of being deceived by the visual appearance of a diagram. Presmeg (1993) points out how students are inclined to go by the visual appearances of a diagram but Presmeg (1993) reported no gender differences in the tendency of students to be misled by the visual appearances of diagrams. In the diagram of this item, x **appears** to be twice the size of y. If females reasoned in this way, it contradicts the report of Presmeg (1993). Males also chose this distractor more than the other distractors, but not to the same degree as females did.

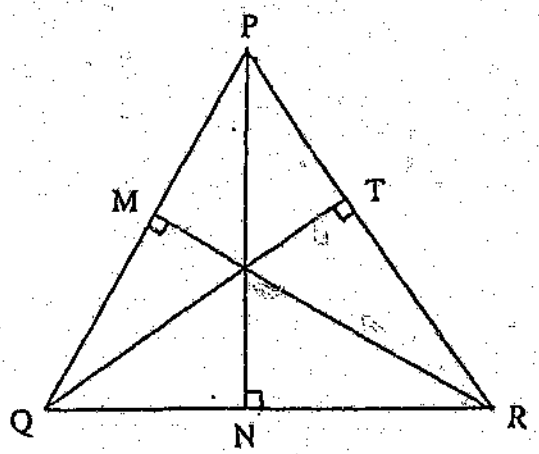
The integration of spatial and analytical skills seems to be important for mathematical problem solving (Eisenberg and Dreyfus, 1991; Tall, 1991), for geometry (Clements and Battista, 1992; Battista, 1990) and probably for the solution of item 4. This item showed differential gender performance in favour of most males but high ability females, however, slightly outperformed males of similar ability.

5.6.3.4 Item 12

Items 11 and 12 are similar in geometrical content. Both items involve cyclic quadrilaterals that are identified in the same way, viz., equal angles subtended by the same chord. This specific way of identifying a cyclic quadrilateral, tends to be more difficult for certain females than for males. The two items will be discussed separately and then together.

Item 12

$\triangle$ PQR with perpendiculars as in diagram.
Which one of the following is
a cyclic quadrilateral ?



A. MNRT
B. PMNT
C. PQNT
D. MQNT

Table 5.10: Values of parameters for item 12

Parameter	a	b	c
Male	1,115	0,736	0,170
Female	1,696	1,167	0,177
d	0,581	0,431	0,007
d(a) + d(b)	1,012 *		

(**d** is the absolute value of the differences.)

This item was more difficult for females than for males and also discriminated better among them. The sum of the differences (*) show that this item produced differential item performance. The direction of the benefit is clear from the diagram of the Item Response Function (Fig. 5.4). A large group of males, ranging in ability level from -1,75 to 1,75 found this item easier than females.

Item 12: Distractors (Table 5.11)

The point in the centre of the triangle was intentionally not marked, so that the most simple prototype cyclic quadrilateral that plays a role in the results of item 11 (see section 5.6.3.5), is not present in this case. The males do not show a tendency to choose one distractor more than the other. For females a specific trend is observed. The first and the last distractor represent 53% of the total female frequency. If the distractors are studied carefully, it is clear that the three distractors are symmetrical in the sense that they represent mirror images of each other. Why then the tendency of females to choose two of the three so often? One explanation that seems acceptable is the position of the quadrilaterals represented in the first and last distractors. Both are in the standard textbook position, although they are not cyclic quadrilaterals. Similar tendencies are present in item 11, which is discussed in the next section. Presmeg (1986) stated that the image of a standard figure may induce inflexible thinking which prevents the recognition of a concept in a non-standard diagram. Females seem to experience this problem to a greater extent than males do. Kabanova-Meller (1950) refers to an inability to distinguish between the essential and unessential concepts in a diagram. The equal angles were recognized, but the essential feature, that of the same chord subtending them, was missed. The above-mentioned studies were not aimed at gender differences.

Item 12: Item Response Function (Figure 5.4)

The Item Response Functions of males and females for item 12 show similar tendencies as that of item 11 and are discussed in the next section together with that of item 11.

Figure 5.4: Item Response Functions for males and females - item 12

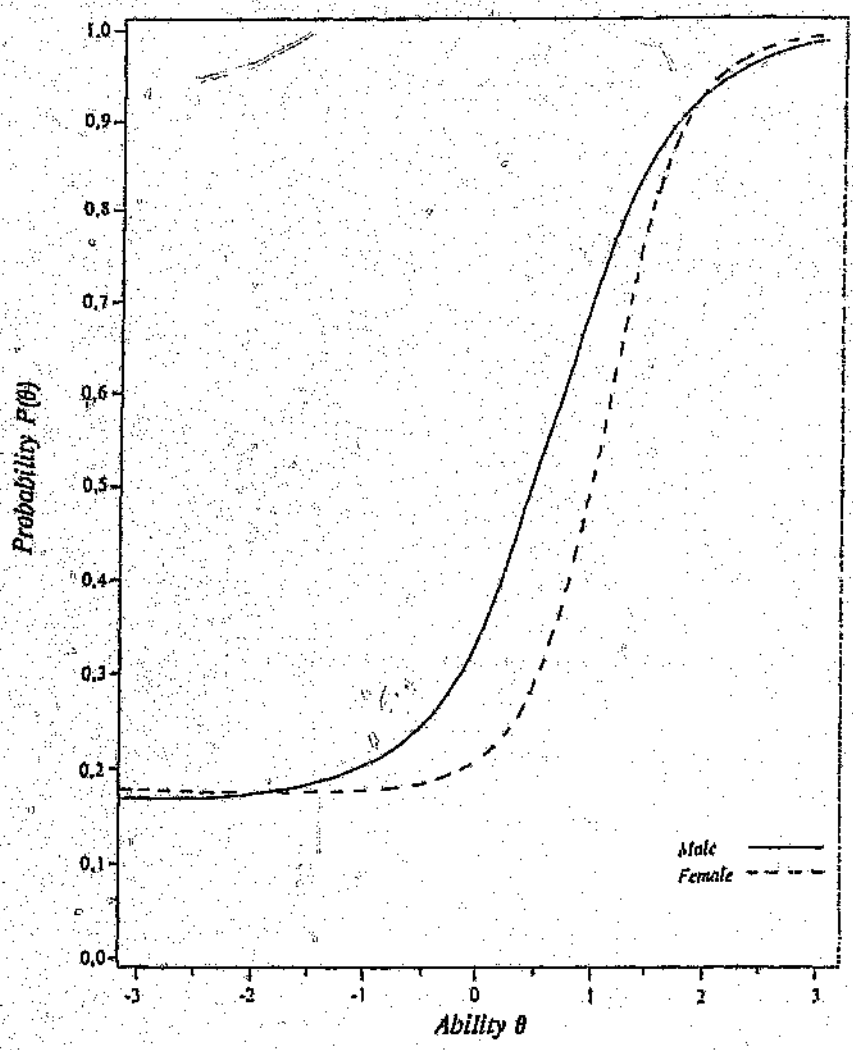


Table 5.11: Frequencies of key and distractors for item 12

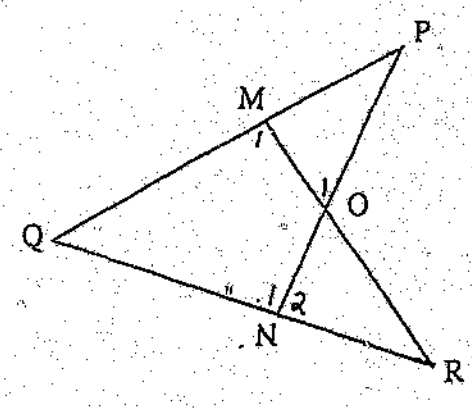
Key and Distractors	Male	Female
MNRT	73	112
PMNT	82	67
PQNT *	157	124
MQNT	91	102

(* indicates the correct response - Appendices E and F)

5.6.3.5 Item 11

$\hat{M}_1 = \hat{N}_1$

Which statement is necessarily true ?



A. PMNR is a cyclic quadrilateral.

B. $\hat{O}_1 = \hat{Q}$

C. MQNO is a cyclic quadrilateral.

D. $\hat{N}_2 = \hat{M}_1$

Table 5.12: Values of item parameters for item 11

Parameter	a	b	c
Male	1,288	0,687	0,188
Female	2,014	0,965	0,134
d	0,726	0,278	0,054
d(a) + d(b)	1,004 *		

(d is the absolute value of the differences.)

This item is a high discriminator for females and also more difficult for females. The item produced differential item functioning and the Item Response Function (Fig. 5.5) provides evidence that it is in favour of males except for the top performers.

Item 11: Distractors (Table 5.13)

The high frequencies of both groups on the third distractor may represent the influence of the prototype diagram of a cyclic quadrilateral. This is evidence of the pupils' limited and standardized accumulation of geometric images (Zykova, 1954). Similar results have been

reported by Hershkowitz et al. (1990), Hershkowitz (1989), Presmeg (1993; 1986) and Kabanova-Meller (1950).

It appears that for item 11 females were more influenced than males by the visual appearance of the diagram. Their prototype image of a cyclic quadrilateral is probably that of a quadrilateral with four sides and four angles. The responses of females on the third distractor is 39% in comparison with 30% for the correct answer. The large number of female responses on the last distractor is also an indication that they considered this specific quadrilateral to be a cyclic quadrilateral and N_2 to be the exterior angle. The tendency shown by females on items 11 and 12 is in contrast with the comment by Hershkowitz et al. (1990) that in spite of individual differences in visual ability, visual distractors act in very much the same way on different individuals and populations.

The results of items 11 and 12 are similar, probably due to the similarity of geometrical content of the items. These research results are of interest because they are also repeated in the open ended questions. (See section 5.7.1.)

Item 11: Item Response Function (Figure 5.5)

The Item Response Functions of item 11 (Fig. 5.5) and item 12 (Fig. 5.4) showed great similarity except for the fact that males tended to guess more on item 11. Both low performing males and females experienced some difficulty in identifying the quadrilaterals, but low performing females were even less successful than low performing males. High performing females did not have the same problem and performed as well as males and sometimes even overtook them. The females, however, only caught up with males at an ability level of approximately 1,5 for these two items. Both items seem to require some integration of spatial and analytical modes of thought (section 3.3.4). As for item 4 (section 5.6.3.3) the high ability (section 4.3.5) females were as successful or slightly more successful than males of similar ability levels.

Figure 5.5: Item Response Functions of males and females - item 11

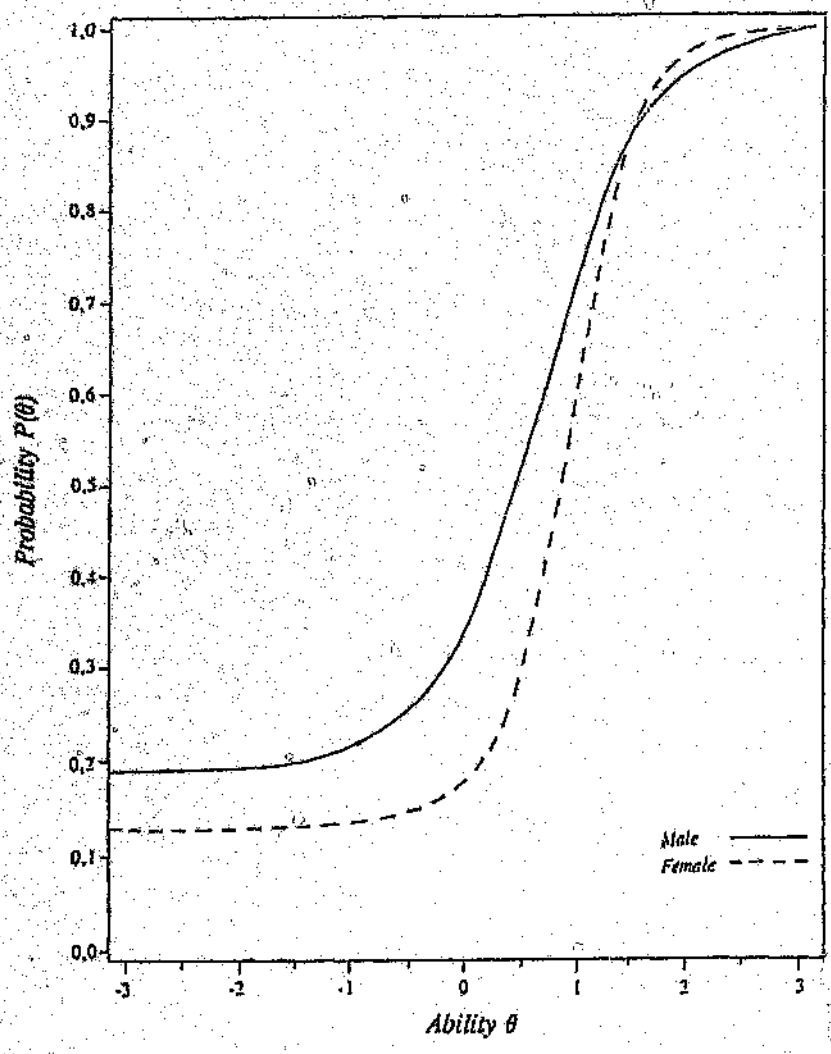


Table 5.13: Frequencies of key and distractors for item 11

Key and Distractors	Males	Females
PMNR is a cyclic quadrilateral *	163	123
$\hat{O}_1 = \hat{Q}$	48	45
MQNO is a cyclic quadrilateral	151	160
$\hat{N}_2 = \hat{M}_1$	44	77

(* indicates the correct response - Appendices E and F)

5.7 DIAGNOSTIC ANALYSIS OF OPEN ENDED TEST INSTRUMENT

The geometry test consisting of open ended questions is given in Appendix I. The mean scores of males (50,96%) and females (50,83%) showed no statistically significant difference as was discussed in section 5.5.1.3. (See Appendix G.)

5.7.1 Analysis of subdivisions of questions

Analysis of questions and subdivisions of questions provides more information on the performance of males and females on Euclidean geometry. The mean scores of males and females on the subdivisions of the open ended geometry questions are given in table 5.14. If mean scores of males and females are compared, a large degree of similarity is observed. The highest differences were found for questions 2.1.1, 2.2.1 and 3.1 but **none** of these differences is statistically significant.

The difference in mean performance on question 2.1.1 is 0,34 which comes to 9%. Question 2.1.1 is similar to item 12 of the multiple choice test (see section 5.6.3.4). Although the difference in performance on the open ended question is not statistically significant, the same tendency is observed. Females seem to find it more difficult than males to identify a cyclic quadrilateral of this specific nature, where two equal angles are subtended by the same chord and where the circle is not shown in the diagram. The females who can identify the quadrilaterals, perform as well as males in answering the rest of the question 2.1. In question 2.2 the cyclic quadrilateral is given. In the first part of the question (2.2.1), the cyclic quadrilateral must be used and now the advantage (0,2), is in favour of females. This advantage comes to 10% if expressed as a percentage.

Question 3.1 is a theorem (bookwork) and it appears that females know their theory better than males. The difference of 0,49 is equal to 8%. Memorization plays a role in this type of question. (See also section 5.7.2.)

Table 5.14: Mean scores on open ended questions

Analysis of open ended questions			
Question	Total score	Males (n = 101)	Females (n = 102)
		mean	mean
1.1	1	0,62	0,60
1.2	7	4,65	4,51
1.3	7	2,07	2,12
2.1.1	4	3,22	2,88
2.1.2	3	1,38	1,02
2.2.1	2	1,40	1,60
2.2.2	5	2,39	2,30
3.1	6	3,10	3,59
3.2.1	3	2,35	2,22
3.2.2	3	2,19	2,12
3.2.3	3	1,30	1,36
3.3.1	5	2,06	2,00
3.3.2	5	1,21	1,13
Total	54	27,52 (50,96%)	27,45 (50,83%)

The other mean gender differences in performance on this test are negligible, in spite of a variety of question types. The last question, 3.3.2, which was the most difficult one for both genders, with mean scores of 1,21 (24,2%) and 1,13 (22,6%) for males and females respectively, showed very little gender difference in performance. This question was aimed at the high performers and again it displayed a great degree of resemblance with the results

of the high performers on the multiple choice test. In section 5.5.2 it was mentioned that 12,9% males (n = 410) and 12,5% females (n = 406) obtained more than 80% on the multiple choice questions. If question 3.3 as a whole is considered, the mean scores are 3,27 or 32,7% and 3,13 or 31,3% for males and females respectively. This question required hypothesis testing and deductive reasoning (usually associated with the fourth van Hiele level - section 2.6.5) as well as a fair amount of visualization. It appears that these aspects did not result in differential gender performance.

Questions 1.2 and 1.3 were included to obtain more information about the performance of students on geometry questions with specific values for angles (question 1.2), and similar questions (question 1.3) where the angles are expressed in terms of symbols. It is interesting to note how similar the tendencies were for males and females. Males' mean performance dropped from 4.65 to 2,07 (66% to 30%) and females' mean performance from 4.51 to 2,12 (64% to 30%). For both gender groups the mean score on question 1.3, almost identical to question 1.2, dropped to less than half that of the mean score on question 1.2. The additional algebraic component to question 1.3 was experienced as much more difficult by males and females.

5.7.2 Diagnostic evaluation of open ended questions

The open ended questions were also evaluated according to the different aspects mentioned in section 5.4. Few gender differences were observed, and where found, differences were very small.

Logical geometrical reasoning and partial logical geometrical reasoning

Both the above-mentioned ways of reasoning refer to correct answers - the first refers to students who could produce a proof in the minimum number of steps for the line of thinking that was used and the second refers to students who arrived at the correct answer but in a lengthy way. The two ways of reasoning do not represent all the correct answers and must rather be seen as the two extreme ways of arriving at the correct answer. Females showed more evidence than males did for both ways of reasoning. The successful students of both

gender groups arrived at the correct answer in a logical way more often than in a roundabout, lengthy way. It must be stressed that numerous correct solution strategies were encountered.

The following are two examples of what is considered to be logical geometrical reasoning. Both are correct solutions to question 2.2 that were given by males and by females. Many other ways of reasoning also represented logical geometrical reasoning.

2.2.1 $\hat{A} = \hat{E}_1$

2.2.2 $\hat{B} = \hat{F}_2$ (exterior angle, cyclic quad)

$\therefore$ $\hat{E}_1 = \hat{C}$ (angles of $\triangle$)

$\therefore$ DMCN a cyclic quad. (exterior angle = opposite angle)

2.2.1 $\hat{A} = \hat{E}_1$

2.2.2 $\hat{N}_1 = \hat{K}_1$ (exterior angle = opposite angle)

$= \hat{M}_p$ (")

$\therefore$ DMCN a cyclic quad. (exterior angle = opposite angle)

The following is an example of what was considered as a lengthy, roundabout way of reasoning.

2.2.1 $\hat{A} = \hat{E}_1$

2.2.2 $\hat{N}_1 + \hat{N}_2 = 180°$ (opposite angles, cyclic quad.)

and $\hat{K}_1 + \hat{K}_2 = 180°$ (")

but $\hat{N}_2 = \hat{K}_2$ (exterior angle = opposite angle)

$\therefore$ $\hat{N}_1 = \hat{K}_1$

and from there on in a similar lengthy manner to deduce that two other angles were also equal. This is only one of many variations of roundabout reasoning.

Rote Memorization

Rote memorization is characterized by an obvious lack of geometrical reasoning probably due to the memorization of theorems, for example, the correct statements, but in the wrong order. Only one theorem (bookwork) was included in the open ended test (question 3.1). Females tended to show more evidence of memorization than males.

True, yet redundant statements

Redundant information was considered to be present if a student made sensible statements leading to an answer, but also wrote down redundant statements which were true according to the diagram, but not part of the line of thinking. This type of error was not very prominent but was made with almost equal frequency by males and females.

Being misled by the perceptual features of a diagram

Because of the prominence of the role that diagrams play in the literature on geometry and visualization, some different ways in which students interpreted diagrams, were considered. Special note was taken when students were misled by the perceptual features of the diagram, e.g., assuming lines perpendicular (especially in question 2.2.2) or lines parallel (especially in question 3.3), because they appeared that way in the diagram. Females were not misled by the diagrams more often than males were, as was hypothesized in the analysis of the multiple choice instrument (see section 5.6.3.3).

Assumptions

Females tend to assume more often what was required to prove, than males did, but as before, the difference was small.

Simplification of the question by assuming specific values for angles

Only one male student made this error. Teachers probably warn against this type of error continuously.

Omission of questions

Females tended to omit more subdivisions of questions than males did.

5.8 SUMMARY

Performance on Euclidean geometry was evaluated for possible gender differences. A secondary school sample representative of both males and females, who were well prepared generally, was used. The gender groups had similar previous exposure to mathematics and Euclidean geometry.

The measuring instrument consisted of two geometry tests - one with open ended questions and the other with multiple choice questions. The mean scores on both these tests showed **no** statistically significant differential gender performance. The gender groups also showed **no** difference in variability (outside the limits generally accepted) in their performance on Euclidean geometry.

Further analysis on item level was done by means of the three parameter logistic model of the Item Response Theory. The computer program WARMEST, which is based on the weighted likelihood estimation algorithm, was used to estimate the item parameters. Of a total of 20 items, **four items** which produced differential item performance were identified by comparing the IRFs and the parameters of the items. Of these four items, one was the easiest item for both gender groups, and the other three were among the more difficult items, as experienced by both groups. All four items have two characteristics in common, namely, that these four items were experienced as more difficult by females (higher **b** parameters - section 4.3.5), and the low ability females caused the lower mean performance of females on these four items in general. On all four items high ability females were as successful or

even more so, than males of the same ability level. Some Euclidean geometry tended to be especially difficult for low ability females. It is thus possible that high ability females make up for their less successful counterparts resulting in comparable mean performance to that of males.

The open ended questions when analyzed diagnostically, (granted subjectively), confirmed some of the findings on the multiple choice questions. Females tend to find it more difficult than males to identify cyclic quadrilaterals where a circle has to be imagined. Two of the four items that produced differential item performance in favour of most males have the same characteristic of a circle that has to be imagined.

Females were inclined to show evidence of logic geometrical reasoning more often than males did, but they also appeared to rely more than males on memorization of theorems (bookwork) in Euclidean geometry. Females omitted a few more subdivisions of questions than males. Both gender groups supplied redundant information and were misled by diagrams to more or less the same extent.

Overall, the complete analysis displayed more resemblances than differences in the performance of males and females on secondary level Euclidean geometry.

CHAPTER 6

MAJOR OUTCOMES, CONCLUSIONS, LIMITATIONS AND RECOMMENDATIONS

6.1 INTRODUCTION

Gender differences in academic performance have become a vital aspect in developing countries because of shortages of personnel in critical fields. Females capable of learning the mathematics required for science or mathematics careers are limited in their options and are a relative loss to the community if they do not develop their mathematical potential to the full extent. It is therefore a matter of concern to learn about the frequent occurence of differential gender performance, to the disfavour of females, in mathematics in general and in geometry in particular.

6.1.1 Major aim of Study

The major aim of this study is to investigate possible differential gender performance in Euclidean geometry. Euclidean geometry is taught synthetically in South African secondary schools and, as a large portion of the geometry curriculum is completed at the end of grade 11 (standard 9), it was decided to involve pupils in this phase in the study. Conditions for research were favourable due to a national core curriculum for mathematics and a uniform school system in a large number of secondary schools in South Africa.

The Classical Test Theory and the one dimensional three parameter logistic model of the Item Response Theory were used in the analysis of the data in order to investigate possible gender differences in performance on test level as well as on item level.

6.1.2 Euclidean geometry

The traditional synthetic approach to the teaching of Euclidean geometry came into disfavour, largely as a result of poor teaching practices. Problems originating from too rigorous teaching methods resulted in the replacement of pure Euclidean geometry in many western countries by other approaches to geometry, for example, transformation geometry, analytical geometry and vectors. Indications are that some of these countries regret their decision to change to other approaches of teaching geometry. Asian countries, like Japan, still teach a substantial amount of Euclidean geometry and are two school years ahead of most western countries in this respect. Although it cannot be ascribed to the greater amount of Euclidean geometry, it is well known that Japanese pupils outperform pupils from western countries by far in some large-scale international mathematics competitions (Robitaille and Garden, 1989). Western countries that still teach a fair amount of traditional deductive Euclidean geometry are, inter alia, South Africa and the United States.

In chapter 2, it has been argued that Euclidean geometry deserves a place in the curriculum for several reasons, especially if the teaching approach is initially informal. Euclidean geometry certainly

* develops spatial and visualization skills - vital elements in cognitive development;
* lends itself to a study and understanding of the real physical world;
* is an inexhaustible source of non-routine problems suitable for creative problem solving in teaching;
* develops logical reasoning invaluable in other spheres of life;
* is a vehicle for representation and model building, and
* remains the most suitable example of a mathematical system to teach at secondary level.

The cultural value of teaching such a historic topic as Euclidean geometry, should not be underestimated. For more than 2 000 years, Euclidean geometry played a dominant role in the development of civilization, and without becoming involved in complicated detail,

secondary pupils can discover the fundamental role that Euclidean geometry played in the development of mathematics and science in general.

6.1.3 Gender differences

The negative consequences of differential gender performance in mathematics and geometry have been outlined. Existing research in this field is hampered by confounding factors some of which are the following:

* Females are under-represented in most samples of secondary pupils. No technique can extrapolate data as to the ability of females where the data does not exist.
* The retention rate in secondary schools of males and females until the age of approximately 18, differs considerably across samples and across countries.
* Gender differences are reported, while only mean performances are compared. It has been shown that gender differences in academic performance can be present in spite of equal mean performance.
* Mathematics is often considered unidimensional, while algebraic and geometrical solution processes differ in nature and sometimes result in different outcomes for the same sample. Also spatial ability, usually related to geometry, is not unidimensional as often assumed. One component of spatial ability, viz. spatial orientation skills, seems to be more closely related to performance in Euclidean geometry.
* Differential gender exposure to previous courses in mathematics, especially to geometry, is a research problem in most countries.
* It has been recently realised that gender differences in performance are not linearly related over all ability levels. The statistical analysis of most studies does not take this important factor into consideration.

6.1.4 Favourable conditions in South Africa for gender research

Favourable conditions for research with regard to gender differences in performance in geometry exist in a large number of South African secondary schools. The mathematics taught in all secondary schools in South Africa was determined by the same core curriculum,

at the time (1993) of the empirical component of this research. At the start of grade 11, all mathematics pupils theoretically had the same mathematical background, and in particular, the same geometrical background, because Euclidean geometry is a compulsory component of mathematics up to grade 12. Another important advantage for the present research is the fact that females were not under-represented in the sample, which consisted of the grade 11 mathematics pupils from five predominantly white schools in the Johannesburg area. Nor were they under-represented in the grade 12 (school leaving) examination of the Transvaal Education Department (TED). The latter examination was written by about 35 000 pupils. These facts eliminate some of the problems mentioned above concerning differential gender exposure to mathematics, different retention rates for males and females till school leaving age, and the under-representation of females in samples for gender research in the field of Euclidean geometry.

Most previous gender research in the field of mathematics in South Africa did **not** consider Euclidean geometry, was only concerned with basic concepts of Euclidean geometry or was hampered by problems in the historically black schools. Attempts to include a sample of non-white pupils in the present research failed due to unrest in schools at the time.

6.2 MAJOR OUTCOMES AND CONCLUSIONS

The sample of 816 grade 11 pupils of five secondary schools consisted of 406 females and 410 males. The gender groups were comparable as regards the aspects mentioned in section 6.1.4.

The test instrument consisted of multiple choice questions in the one test and open ended questions, which included proof writing, deductive reasoning and hypothesis testing, in the second test. The latter skills are all considered to be on the fourth van Hiele level of geometrical thinking.

The statistical analysis of the data comprised analysis within the framework of the Classical Test Theory by means of SAS computer programs. Within the framework of the Item

Response Theory, the parameters for the one dimensional three parameter logistic model, were estimated by means of the computer program WARMEST.

6.2.1 Comparison at mean level

Mean scores on both the multiple choice test and the open ended test showed **no** statistically significant gender differences in performance in Euclidean geometry. Also the standard deviations of males and females on both instruments indicated very little difference (see sections 5.5.1.1 to 5.5.1.3).

The result of the present research thus contradicts some findings in South Africa (see section 3.3.1.5) and earlier findings (see section 3.3.1.1)

* of male superiority in spatial abilities from adolescence onwards (Maccoby and Jacklin, 1974) and also the confirmation of Hyde (1981) in a meta-analysis (It is not claimed that Euclidean geometry refers to all the components of spatial ability.);

* that males scored higher than females on geometry over all cognitive levels (Fennema and Carpenter, 1981), and

* that although cognitive gender differences are disappearing, the male advantage in space relations remained up to 1980 (Feingold, 1988).

The present results seem to be partially supportive of some of the more recent meta-analytical studies (Friedman, 1994) stating that gender differences in mathematics performance are declining. For geometry specifically, a moderate (Friedman, 1994) to only slight (Frost et al., 1994; Hyde et al., 1990) male advantage was found. (See sections 3.3.1.3 to 3.3.1.5.) Friedman (1994) and Hyde et al. (1990) found that increasing selectivity of samples increased the male advantage. (See also sections 3.3.1.6 and 3.4.) The samples involved in the international study (section 3.3.1.6) that found a male advantage in geometry performance for 10 out of the 15 participating countries, were also, with few exceptions, highly selective. The results of this study suggests that samples which are representative of both gender groups

are more likely to result in an absence of differential gender performance in geometry. The findings that are supportive or partially supportive of the present study are discussed in section 3.3.1.5. No gender related differences were found in

* students' ability in geometrical proof writing (Senk and Usiskin, 1983);
* in spatial orientation skills (Tartre, 1990), and
* in a re-analysis of the geometrical content of the well-reported Benbow and Stanley (1982) data by Becker (1990), and
* the problem solving strategies used in solving geometry problems (Battista, 1990).

6.2.2 Comparison at item level

The absence of differential gender performance in mean scores on the test instruments does not necessarily imply the absence of more specific differences (Becker and Forsyth, 1994; Feingold, 1994; Tartre, 1990) in academic performance. Therefore, the Euclidean geometry content of the multiple choice test was considered at item level (see section 5.6).

The estimated parameters of the three parameter logistic model of the Item Response Theory were compared for the two gender groups according to certain guidelines (see section 5.6.3.1) for the determination of differential item performance. This statistical model is considered to be very sensitive in identifying differences between groups. The interaction of the ability levels of students and gender differences in the academic achievement of students, identified by Becker and Forsyth (1994), Feingold (1994) and Tartre (1990), could be addressed by the present research by implementing the Item Response Theory in its capacity to compare data at different ability levels. (See also sections 4.3 and 4.6.)

Four of the 20 items were identified as producing differential item performance (section 5.6.3). The differences on all four items were non-uniform, i.e., not in favour of the same gender for all ability levels. However, the differences were in favour of males more often than for females and also over larger ability ranges than for females. This is best illustrated by the Item Response Functions of both gender groups for these four items (see sections 5.6.3.2 - 5.6.3.5). (See also section 4.3.5.)

The following three common trends were observed in the performance on these four Euclidean geometry items:

* All four items were experienced as being more difficult (higher **b** parameters - see sections 4.3.5 and 4.3.6) by females than by males (Tables 5.6, 5.8, 5.10 and 5.12). Doolittle and Cleary (1987) reported similar results for geometry items, although the geometry was of a lower cognitive level.

* The females who perform at the lower ability levels (section 4.3.5) for these four items, were consistently outperformed by males of similar ability. High performing females performed as well on these four items, and sometimes even a little better, than high performing males. On all four items, females caught up with males at some point along the higher ranges of the ability scale. (See the Item Response Functions of these items - Figures 5.2 - 5.5.) A much clearer picture of differential item performance is obtained in this way. The IRT contributed to a better understanding of gender differences in performance in Euclidean geometry by comparing the entire score distributions of males and females.

* The responses on three items appear to indicate that females functioning at the higher ability levels (section 4.3.5) integrate spatial and analytical skills (as represented in Euclidean geometry) as well as males do or sometimes even slightly better. (See sections 3.3.1.5. and 3.3.4 and Figures 5.3, 5.4 and 5.5.)

The four items represented a variety of geometrical content (Appendix H - items 4, 8, 11 and 12). One item was easy and the other three among the more difficult items as judged from the **b** parameters (Appendix A) of these items.

Item 11 and item 12

Items 11 and 12 (sections 5.6.3.4 and 5.6.3.5) involved the aspect of identifying a specific type of cyclic quadrilateral, viz. where two equal angles were subtended by the same line segment and the circle was not shown in the diagram. Females of lower ability levels could

not identify such cyclic quadrilaterals as well as males could. A similar tendency was observed on the open ended questions (see section 5.7.1). It is suggested that females may be more influenced by the prototype diagram (see section 2.6.4) portraying a cyclic quadrilateral as a quadrilateral with four sides and four angles or by the other prototype cyclic quadrilateral with two equal angles, subtended by the same chord, in the usual upright position. High frequencies on certain distractors seem to suggest that the image of the standard textbook figure induces inflexible geometrical thinking (see section 2.6.4) in females functioning on the lower ability levels. This inflexibility inhibits females in working with a concept if the related diagram is non-standard, to a larger degree than is the case for males functioning at similar ability levels. Brown and Presmeg (1993) argue that weaker students tend to rely on 'concrete images' (non-moving pictures) in mathematics, but they did not analyze this aspect in relation to gender. Reports on these phenomena were generally not aimed at gender differences (Brown and Presmeg, 1993; Presmeg, 1993; Hershkowitz et al., 1990; Hoz, 1981; Zykova, 1954; Krivnova-Meller, 1950).

Item 4

The algebraic component to item 4 (section 5.6.3.3) did not seem to be linked to the differential gender performance. Low ability females seem to be more handicapped by geometrical rigidity (Hoz, 1981), which again can be linked to the visual appearance of the diagram (sections 5.6.3.3 and 2.6.4).

Item 8

Item 8 (section 5.6.3.2) was the easiest item for both gender groups (see Appendices B and C). Low ability females were less successful than low ability males on this item which only required the identification of angles on the circle, subtended by the same chord. The item discriminated well between low and high ability females, but not so for males. Again it appears that the low ability females could not apply this basic knowledge of the circle theorems of Euclidean geometry.

6.2.3 Analysis of open ended questions

The most striking feature of this analysis was the large degree of resemblance in the performance of males and females in Euclidean geometry (see Table 5.14). Not one of the subdivisions of the open ended questions showed any statistically significant differential gender performance, in spite of a large variety of questions (Appendix I). Where gender differences in reasoning processes and errors made were observed, they were very small. (See section 5.7.) Few students from both gender groups were misled by the diagrams and few made use of redundant information. This can probably be ascribed to the fact that the students were well prepared, functioning mostly under examination conditions, and would rather not attempt a question if they were in doubt. Females tended to show evidence of logical geometrical reasoning on the one hand, and of memorization on the other, a few times more often than males did. They also omitted more subdivisions of questions than males did. The latter may explain the fact that on the open ended questions, females were in general not misled more than males were by the obvious perceptual features of the diagrams such as perpendicular lines and parallel lines. However, the responses of the females on question 2.1.1 showed the same tendency that was observed on the multiple choice test. (See sections 5.6.3.4 and 5.6.3.5.) Females could not identify the cyclic quadrilaterals as well as males did, where the circle was not shown in the diagram.

6.2.4 Conclusion

The comparison of mean performance only is not a sufficient measure for detecting differential gender performance. Differences in favour of one gender group at lower ability levels, for example, can be cancelled out at other levels of ability, resulting in equal mean performance. Overall, 20% of the multiple choice questions showed differential item performance in favour of most males, On these items, the females' disadvantage was caused by the subgroup performing at the lower levels of ability, while females functioning at high levels of ability equalled or even outperformed males functioning at similar ability levels. This interaction of ability levels of students and gender differences in performance in Euclidean geometry was clearly illustrated by the Item Response Functions. The one

dimensional three parameter logistic model of the Item Response Theory proved to be effective in identifying differential item performance.

In general, the performance of males and females in secondary Euclidean geometry, displayed **more resemblances than differences.**

6.3 LIMITATIONS OF THE STUDY

* The mean level of difficulty of both Euclidean geometry tests was sufficient. However, one or two items, functioning at the ability level of the top 5% students, would have provided more information on that specific subgroup. The highest **b** parameter was 1,3 (Appendix A). Items with **b** parameters of at least 2 would have provided more information on the real top performers.

* The results of this study can not be generalized to include students who are not exposed to reasonably good teaching of Euclidean geometry. Because extremely poor teaching methods are often blamed for Euclidean geometry falling into disfavour, students subjected to such practices were purposely excluded.

* The research results cannot necessarily be generalized to include other cultural groups in South Africa or to other countries. In terms of this research this is not considered a limitation, because this study was specifically aimed at a more homogeneous group, as was discussed in section 3.4.

* The number of pupils used to calibrate the multiple choice questions could have been larger. Attempts to involve more pupils were unsuccessful.

6.4 RECOMMENDATIONS

* Research studies that compare groups with the purpose of identifying differences in performance should consider models from the Item Response Theory, where sample

sizes allow for it. The one dimensional three parameter logistic model is especially suitable for such comparisons.

* Euclidean geometry is not within reach of only a select few students. Good teaching results in fairly good average performance of students. The specific gender differences in performance on Euclidean geometry that were noticed should be brought to the attention of teachers in general so that these problems can be addressed.

* It seems that Euclidean geometry classified as the fourth van Hiele level of thinking can be evaluated successfully by means of multiple choice questions. It is recommended that more use of this type of questioning is made in school leaving examinations, especially in view of the increasing numbers of students who will write these examinations in the near future in South Africa.

* Research on the van Hiele levels of geometrical thinking should include some gender research.

* Gender research in mathematics performance should include samples comparably representative of both gender groups. The under-representation of one of the groups can lead to distorted results.

* Similar studies should be conducted with pupils of other population groups and also with multicultural classes to see if the same results are obtained.

* A study including more items, similar to those which showed differential gender performance (see sections 5.6.3.2 - 5.6.3.5), should be undertaken to confirm the results obtained on gender differences in performance in Euclidean geometry. A qualitative study should also be done to explore the causes of these observed gender differences in performance.

BIBLIOGRAPHY

AIKEN, L.R. (1987) Testing with multiple-choice items. Journal of Research and Development in Education. 20,4: 44-58.

AIKEN, L.R. (1970) Attitudes towards mathematics. Review of Educational Research. 40, 4: 551-596.

ALLEN, M.J. & YEN, W.M. (1979) Introduction to measurement theory. Monterey: Brooks/Cole Publishing Company.

ANASTASI, A. (1958) Differential Psychology. (3rd ed.) New York: Macmillan.

ANGOFF, W.A. (1982) Use of difficulty and discrimination indices for detecting item bias. In R.A. Berk (ed) 1982. Handbook of methods for detecting test bias. Baltimore: Johns Hopkins University Press.

ARMSTRONG, J.M. (1981) Achievement and participation of women in mathematics: Results of two national surveys. Journal for Research in Mathematics Education. 12, 5: 356-372.

ARY, D., JACOBS, L.C. & RAZAVIEH, A. (1985) Introduction to research in education. Third edition. New York: Holt, Rinehart and Winston.

BAART, L. (1992) Measuring the earth. Pythagoras. 28: 39-42.

BAENNINGER, M. & NEWCOMBE, N. (1989) The role of experience in spatial test performance: A meta-analysis. Sex-Roles. 20, 5/6: 327-344.

BAILEY, S. (1993) The current status of gender equity research in American schools. Educational Psychologist. 28, 4: 321-339.

BAKER, F.B. (1985) The basics of item response theory. Portsmouth: Heinemann.

BARNARD, J.J. (1993a) Die gebruik van veelkeusevrae in eksamens. Pythagoras. 31, 22-27.

BARNARD, J.J. (1993b) Equating examination standards. South African Journal of Education. 13, 4: 145-149.

BARNARD, J.J. (1992) The feasibility of using the minimum chi-square method in equating scores calibrated by means of a weighted likelihood estimation algorithm. Ed.D. thesis. Newport University, California.

BARNARD, J.J. (1991) Die gebruik van die logistiese drie-parameter itemresponsmodel in onderwys. Ongepubliseerde Ph.D. verhandeling. Pretoria: Universiteit van Pretoria.

BARNARD, J.J. & OOSTHUIZEN, W.L. (1992) Nuwe era in opvoedkundige meting. Suid-Afrikaanse Tydskrif vir Opvoedkunde. 12, 1: 8-11.

BATTISTA, M.T. (1990) Spatial visualization and gender differences in high school geometry. Journal for Research in Mathematics Education. 21, 1: 47-60.

BAYA'A, N.F. (1990) Mathematics anxiety, mathematics achievement, gender and socio-economic status among Arab secondary students in Israel. International Journal of Mathematical Education in Science and Technology. 21: 319-324.

BECKER, B.J. (1990) Item characteristics and gender differences on the SAT-M for mathematically able youths. American Educational Research Journal. 27, 1: 65-87.

BECKER, D.F. & FORSYTH, R.A. (1994) Gender differences in mathematics problem solving and science: A longitudinal analysis. International Journal of Educational Research. 21, 4: 407-416.

BECKER, J.P. (1992) Results of U.S-Japan cross-national research on students' problem solving behaviours. In I. Wirszup & R. Streit (eds) 1991. Developments in school mathematics education around the world. Volume 3. Proceedings of the Third UCSMP International Conference on Mathematics Education. Reston, Virginia: National Council of Teachers of Mathematics.

BECKER, J.R. (1981) Differential treatment of females and males in mathematics classes. Journal for Research in Mathematics Education. 12, 1: 40-53.

BELL, A.W. (1976) A study of pupils' proof-explanations in mathematical situations. Educational Studies in Mathematics. 7: 23-40.

BELL, F.H. (1978) Teaching and learning mathematics. Iowa: Wm. C. Brown Company Publishers.

BENBOW, C.P. (1992) Academic achievement in mathematics and science of students between ages 13 and 23: Are there differences among students in the top one percent of mathematical ability ? Journal of Educational Psychology. 84, 1: 51-61.

BENBOW, C.P. & STANLEY, J.C. (1983) Sex-differences in mathematical reasoning ability: More facts. Science. 222: 1029-1031.

BENBOW, C.P. & STANLEY, J.C. (1982) Consequences in high school and college of sex differences in mathematical reasoning ability: A longitudinal perspective. American Educational Research Journal. 19, 4: 598-622.

BEN-CHAIM, D., LAPPAN, G. & HOUANG, R.T. (1988) The effect of instruction on spatial visualization skills of middle school boys and girls. American Educational Research Journal. 25, 1: 51-71.

BEN-SHAKHAR, G. & SINAI, Y. (1991) Gender differences in multiple-choice tests: The role of differential guessing tendencies. Journal of Educational Measurement. 28, 1: 23-35.

BERK, R.A. (ed) (1982) Handbook of methods for detecting test bias. Baltimore: Johns Hopkins University Press.

BISHOP, A.J. (1989) Review of research on visualization in mathematics education. Focus on Learning Problems in Mathematics. 11, 1: 7-15.

BISHOP, A.J. (1983) Space and geometry. In R. Lesh & M. Landau (eds) 1983. Acquisition of mathematics concepts and processes. New York: Academic Press.

BLEISTEIN, C.A. (1986) Application of Item Response Theory to the study of differential item characteristics: A review of the literature. Research Memorandum. Educational Testing Service, Princeton, New Jersey. ED 268 160.

BLITHE, T., FORBES, S., CLARK, M., ROBINSON, E. & WHITWELL, K. (1994) Gender differences in New Zealand mathematics performance at the secondary - tertiary interface. International Journal of Educational Research. 21, 4: 427-438.

BLOOM, B.S. (ed), ENGELHART, M.D., FURST, E.J., HILL, W.H. & KRATHWOHL, D.R. (1956) Taxonomy of educational objectives. Handbook I: Cognitive domain. New York: David McKay Company, inc.

BLOOM, W. (1988) Spaces and geometry. In A. Hirst and K. Hirst (eds) 1988. Proceedings of the Sixth International Congress on Mathematical Education. Hungary: Malév.

BLOOM, W. (1984) The teaching of geometry. In M. Carrs (ed) 1986. Proceedings of the Fifth International Congress on Mathematical Education. Boston: Birkhäuser.

BOHLIN, C.F. (1994) Learning style factors and mathematics performance: Sex-related differences. International Journal of Educational Research. 21, 4: 387-398.

BOYER, C.B. (1968) A history of mathematics. New York: John Wiley & sons.

BRANDON, P.R., NEWTON, B.J. & HAMMOND, O.W. (1987) Children's mathematics achievement in Hawaii: Sex differences favouring girls. American Educational Research Journal. 24, 3: 437-461.

BREEN, C.J. (1992a) Curriculum development: Current trends in South Africa. In M. Moodley, R.A. Njisane & N.C. Presmeg (eds) 1992. Mathematics education for in-service and pre-service teachers. Pietermaritzburg: Shuter & Shooter.

BREEN, C.J. (1992b) Informal geometry. In M. Moodley, R.A. Njisane & N.C. Presmeg (eds) 1992. Mathematics education for in-service and pre-service teachers. Pietermaritzburg: Shuter & Shooter.

BROWN, C.A., CARPENTER, T.P., KOUBA, V.L., LINDQUIST, M.M., SILVER, E.A. & SWAFFORD, J.O. (1988) Secondary school results for the fourth NAEP: Mathematics assessment. Mathematics Teacher. 81, 5: 337-347.

BROWN, D.L. & PRESMEG, N.C. (1993) Types of imagery used by elementary and secondary school students in mathematical reasoning. Proceedings of the Seventeenth International Conference - Psychology of Mathematics Education. Volume II. Tsukuba, Japan: University of Tsukuba.

BRUSSELMANS-DEHAIRS, C. & HENRY, G.F. (1994) Introduction. International Journal of Educational Research. 21, 4:353-360.

BURGER, W.F. & SHAUGHNESSY, J.M. (1986) Characterizing the van Hiele levels of development in geometry. Journal for Research in Mathematics Education. 17, 1: 31-48.

BURRILL, L.E. (1982) Comparative studies of item bias methods. in R.A. Berk (ed) 1982. Handbook of methods for detecting test bias. Baltimore: Johns Hopkins University Press.

BURSTEIN, L. (1992) (ed) The IEA study of mathematics III: Student growth and classroom processes. Oxford: Pergamon press.

BURTON, D.M. (1985) The history of mathematics. Massachusetts: Allyn and Bacon.

BURTON, G. & PETERSON, K. (1985) Sex differences in geometry students perceptions of ideal and actual behaviour. Journal of Research and Development in Education. 18, 4: 55-59.

BURTON, L. (ed) (1990) Gender and mathematics: An international perspective. New York: Cassell Educational Ltd.

CAMILLI, G. & SHEPARD, L.A. (1987) The inadequacy of ANOVA for detecting test bias. Journal of Educational Statistics. 12, 1: 87-99.

CHAZAN, D. (1993) High school geometry students' justification for their views of empirical evidence and mathematical proof. Educational Studies in Mathematics. 24: 359-387.

CHIPMAN, S.F., MARSHALL, S.P. & SCOTT, P.A. (1991) Content effects on word problem performance: A possible source of test bias. American Educational Research Journal. 28, 4: 897-915.

CLEMENTS, D.H. and BATTISTA, M.T. (1992) Geometry and spatial reasoning. In Grouws, D.A. (ed) 1992. Handbook of research on mathematics teaching and learning. New York: Macmillan Publishing Company.

CLEMENTS, M.A. (1979) Sex differences in mathematical performance: An historical perspective. Educational Studies in Mathematics. 10: 305-322.

CONNOR, J.M. & SERBIN, L.A. (1985) Visual-spatial skill: Is it important for mathematics ? Can it be taught ? In S.F. Chipman, L.R. Brush & D.M. Wilson. (eds) 1985. Women and Mathematics: Balancing the equation. London: Lawrence Erlbaum ass. Publishers.

CROCKER, L. & ALGINA, J. (1986) Introduction to classical and modern test theory. New York: Holt, Rinehart and Winston.

CRONJÉ, L.S. (1991) Die invloed van 'n metode van geleide ontdekking, waarby die geskiedenis van Wiskunde integreer word, op die houding van St. 9-leerlinge teenoor Meetkunde. Ongepubliseerde M.Sc. navorsingsverslag. Johannesburg: Universiteit van die Witwatersrand.

CROWLEY, M.L. (1990) Criterion-referenced reliability indices associated with the van Hiele geometry test. Journal for Research in Mathematics Education. 21, 238-241.

CROWLEY, M.L. (1987) The van Hiele model of the development of geometric thought. In M.M. Lindquist & A.P. Schulte (eds) 1987. Learning and teaching geometry, K-12. 1987 Yearbook. Virginia: National Council of Teachers of Mathematics.

CURIO, F.R. (1987) Comprehension of mathematical relationship expressed in graphs. Journal for Research in Mathematics Education. 18, 382-393.

CURTIS, P.J.D., de VILLIERS, J.U. & POLONSKY, M.J. (1989) Inter-group differences in examination scores resulting from multiple-choice and essay questions. South African Journal of Education. 9, 2: 260-262.

DAVIS, P.J. & HERSH, R. (1981) The mathematical experience. Boston: Birkhäuser.

De la ROSA, B. (1988) Historical perspectives - an essential ingredient of the equipment of the mathematics teacher. In du Toit, G. (ed) 1988. Proceedings: Ninth National Congress on Mathematics Education. The Mathematical Association of South Africa.

DEPARTMENT OF EDUCATION AND CULTURE - ADMINISTRATION: HOUSE OF ASSEMBLY (TRANSVAAL EDUCATION DEPARTMENT) (1992) Mathematics: Scheme of work std. 6.

De VILLIERS, M.D. (1991) Pupils' need for conviction and explanation within the context of geometry. Pythagoras. 26: 18-27.

De VILLIERS, M.D. (1990) The role and function of proof in mathematics. Pythagoras. 24: 17-23.

De VILLIERS, M.D. (1988) Enkele konsepsuele raamwerke van leerlinge in meetkunde. In G. du Toit (red.) 1988. Verrigtinge: Negende Nasionale Kongres oor Wiskunde-onderwys. Wiskundegenootskap van Suid-Afrika.

De VILLIERS, M.D. (1987) Algemene beheersingsvlakke van sekere wiskundige werkwyses deur voornemende wiskunde-onderwysers. Suid-Afrikaanse Tydskrif vir Opvoedkunde. 7, 1: 34-41.

De VILLIERS, M.D. (1986) Navorsingsontwikkelinge en onderrig-implikasies van die van Hiele teorie van meetkundige denke. In G. Oberholzer (red) 1986. Verrigtinge: Agtste Nasionale Kongres oor Wiskunde-onderwys. Die Wiskundegenootskap van Suid-Afrika.

De VILLIERS, M.D. & OBERHOLZER, G. (1986) Geometry by Algebra or why use a cartesian plane ? In G. Oberholzer (ed) 1986. Proceedings: Eighth National Congress on Mathematical Education. The Mathematical Association of South Africa.

DIEUDONNé, J.A. (1973) Should we teach 'modern' mathematics ? American Scientist. 61, 1: 16-19.

DOOLITTLE, A.E. (1985) Understanding differential item performance as a consequence of gender differences in academic background. Paper presented at the Annual Meeting of the American Educational Research Association. Chicago. ED 263 218.

DOOLITTLE, A,E. & CLEARY, T.A. (1987) Gender-based differential item performance in mathematics achievement items. Journal of Educational Measurement. 24, 2: 157-166.

DREYFUS, T. & HADAS, N. (1987) Euclid may stay - and even be taught. In M.M. Lindquist & A.P. Shulte. (eds) Learning and teaching geometry, K-12. 1987 yearbook. Virginia: National Council of Teachers of Mathematics.

DRISCOLL, M. (1982) Research within reach: Secondary school mathematics. Virginia: National Council of Teachers of Mathematics.

DRIVER, G. (1980) How West Indians do better at school (especially girls). New Society. 1980: 111-114.

Du PLESSIS, S.J.P., van WYK, C.K. & VOLSCHENK, P.G. (1993) Beter akademiese prestasie van damestudente as manstudente: nuwe perspektief. Suid-Afrikaanse Tydskrif vir Hoër Onderwys. 7, 1: 21-30.

Du TOIT, P.J. (1921) Meetkunde. Stellenbosch: Ecclesia Drukkery.

DWECK, C.S., DAVIDSON, W., NELSON, S. & ENNA, B. (1978) Sex differences in learned helplessness: II. The contingencies of evaluative feedback in the classroom and III. An experimental analysis. Developmental Psychology. 14, 3: 268-276.

DWECK, C.S. & GILLIARD, D. (1975) Expectancy statements as determinants of reactions to failure: Sex differences in persistence and expectancy change. Journal of Personality and Social Psychology. 32, 6: 1077-1084.

EARLE, H.F. (1972) Attitudes toward geometry. Ph.D. dissertation. University of Maryland, USA.

EBEL, R.L. (1986) Essentials of educational measurement. New-Jersey: Prentice-Hall.

EISENBERG, T. & DREYFUS, T. (1991) On the reluctance to visualise in mathematics. In W. Zimmerman & S. Cunningham (eds) 1991. Visualization in teaching and learning mathematics. USA: Mathematical Association of America.

EGSGARD, J. & FLETCHER, T. (1984) Senior Secondary School. Geometry. In M. Carrs (ed) 1986. Proceedings of the Fifth International Congress on Mathematical Education. Boston: Birkhäuser.

ERNEST, P. (1991) The Philosophy of mathematics education. London: The Falmer Press.

ETHINGTON, C.A. (1992) Gender differences in a psychological model of mathematics achievement. Journal for Research in Mathematics Education. 23, 2: 166-181.

ETHINGTON, C.A. (1990) Gender differences in mathematics: An international perspective. Journal for Research in Mathematics Education. 21: 74-80.

ETHINGTON, C.A. & WOLFE, L.M. (1984) Sex-differences in a causal model of mathematics achievement. Journal for Research in Mathematics Education. 15: 361-377.

FARRELL, M.A. (1987) Geometry for secondary school teachers. In M.M. Lindquist & A.P. Shulte. (eds) Learning and teaching geometry, K-12. 1987 Yearbook. Virginia: National Council of Teachers of Mathematics.

FEINGOLD, A. (1994) Gender differences in variability in intellectual abilities: A cross-cultural perspective. Sex Roles. 30, 1-2: 81-92.

FEINGOLD, A.F. (1993) Cognitive gender differences: A developmental perspective. Sex-Roles. 29, 1/2: 91-113.

FEINGOLD, A. (1992) Sex-differences in variability in intellectual abilities: A new look at an old controversy. Review of Educational Research. 62, 1: 61-84.

FEINGOLD, A. (1988) Cognitive gender differences are disappearing. American Psychologist. 43, 2: 95-103.

FENNEMA, E. (1990) Justice, equity and mathematics education. In E. Fennema and G. Leder (eds) 1990. Mathematics and Gender. New York: Teachers College Press.

FENNEMA, E. (1981) Women and mathematics: Does research matter ? Journal for Research in Mathematics Education. 12, 5: 380-385.

FENNEMA, E. and CARPENTER, T.P. (1981) Sex-related differences in mathematics: Results from national assessment. Mathematics Teacher. 74: 554-559.

FENNEMA, E. & HART, L.E. (1994) Gender and the JRME. Journal for Research in Mathematics Education. 25, 6: 648-659.

FENNEMA, E. and LEDER, G.C. (eds) (1990) Mathematics and gender. New York: Teachers College Press.

FENNEMA, E. & SHERMAN, J. (1978) Sex-related differences in mathematics achievement and related factors: A further study. Journal for Research in Mathematics Education. 9, 3: 189-203.

FENNEMA, E. & SHERMAN, J. (1977) Sex-related differences in mathematics achievement, spatial visualization, and affective factors. American Educational Research Journal. 14, 1: 51-71.

FENNEMA, E. & TARTRE, L.A. (1985) The use of spatial visualization in mathematics by girls and boys. Journal for Research in Mathematics Education. 16, 3: 184-206.

FERRINI-MUNDY, J. (1987) Spatial training for calculus students: sex differences in achievement and in visualization ability. Journal for Research in Mathematics Education. 18, 2: 126-140.

FLAUGHER, R.L. (1978) The many definitions of test bias. American Psychologist. 33, 7: 671-679.

FREUDENTHAL, H. (1973) Mathematics as an educational task. Dordrecht: Reidel.

FRIEDMAN, L. (1994) Meta-analytical contributions to the study of gender differences in mathematics - The relationship of mathematical and spatial skills. International Journal of Educational Research. 21, 4: 361-371.

FRIEDMAN, L. (1989) Mathematics and the gender gap: A meta-analysis of recent studies on sex differences in mathematics tasks. Review of Educational Research. 59: 185-213.

FROST, L.A., HYDE, J.S. & FENNEMA, E. (1994) Gender, mathematics performance, and mathematics-related attitudes and affect: A meta-analytic synthesis. International Journal of Educational Research. 21, 4: 373-386.

GARAFOLO, J. (1989) Beliefs, responses and mathematics education: Observations from the back of the classroom. School Science and Mathematics. 7: 451-455.

GARDEN, R.A. (1989) Students achievement: Population B. In D.F. Robitaille & R.A. GARDEN (eds) 1989. The IEA Study of mathematics II: Contexts and outcomes of school mathematics. Oxford: Pergamon Press.

GILES, F. (1982) Geometry for all. Mathematics Teaching. 100: 30-37.

GOLDIN, G.A. (1990) Epistemology, constructivism, and discovery learning in mathematics. In R.B. Davis, C.A.Maher & N. Noddings (eds) 1990. Constructivist views on the teaching and learning of mathematics. Reston: National Council of Teachers of Mathematics.

GOWER, E. (1983) Geometry in the secondary school: Some considerations. In M. Zweng et al., (eds) 1983. Proceedings of the Fourth International Congress on Mathematics Education. Boston: Birkhäuser.

GRUNBAUM, B. (1983) Shouldn't we teach geometry ? In M. Zweng (ed) 1983. Proceedings of the Fourth International Congress on Mathematical Education. Boston: Birkhäuser.

GUILFORD, J.P. (1965) Fundamental statistics in psychology and education. New York: Mc.Graw-Hill Book company.

GUTIÉRREZ, A., JAIME, A. & FORTUNY, J.M. (1991) An alternative paradigm to evaluate the acquisition of the van Hiele levels. Journal for Research in Mathematics Education. 22, 3: 237-251.

HALPERN, D.F. (1989) The disappearance of cognitive gender differences: What you see depends on where you look. American Psychologist. 44: 1156-1157.

HALPERN, D.F. (1986) A different answer to the question, "Do sex-related differences in spatial abilities exist?" American Psychologist. 41, 2: 1014-1015.

HAMBLETON, R.K. (1989) Principles and selected applications of Item Response Theory. In R.L. Linn (1989) Educational measurement. (3rd Edition) London: MacMillan

HAMBLETON, R.K. & SWAMINATHAN, H. (1985) Item Response Theory: Principles and applications. Boston: Kluwer Nijhoff Publishing.

HAMBLETON, R.K., SWAMINATHAN, H. & ROGERS, H. (1991) (eds) Fundamentals of Item Response Theory. Volume 2. Newbury Park: Sage publications.

HANNA, G. (1994) Cross-cultural gender differences in mathematics education. International Journal of Educational Research. 21, 4: 417-426.

HANNA, G. (1986) Sex differences in mathematics achievement of eighth graders in Ontario. Journal for Research in Mathematics Education. 17, 231-237.

HANNA, G. (1983) Rigorous proof in mathematics education. Ontario: Ontario Institute for Studies in Education.

HANNA, G., KUNDIGER, E. & LAROUCHE, C. (1990) Mathematical achievement of grade 12 girls in fifteen countries. In Burton, L. (ed) 1990. Gender and Mathematics: An International Perspective. New York: Birkhäuser.

HANSEN, M. (1993) What is Geometry ? Mathematics in school. 22, 4: 9-11.

HARADA, K.; NOHDA, N. & GALLOU-DUMIEL, E. (1993) The role of conjectures in geometrical proof-problem solving: France - Japan collaborative research. In

Proceedings of the Seventeenth International Conference - Psychology of Mathematics Education, volume III. Tsukuba, Japan: University of Tsukuba.

HEMBREE, R. (1990) The nature, affects and relief of mathematics anxiety. Journal for Research in Mathematics education. 21, 1: 33-46.

HENDRICKS, M.A. (1991) The nature and value of mathematics. Spectrum. 29, 2: 32-35.

HENSEL, R.A.M. (1989) Mathematical achievement: Equating the sexes. School Science and Mathematics. 89, 8: 646-653.

HERSH, R. (1993) Proving is convincing and explaining. Educational Studies in Mathematics. 24,: 389-399.

HERSHKOWITZ, R. with the collaboration of D. BEN-CHAIM, C. HOYLES, G. LAPPAN, M. MITCHELMORE & S. VINNER (1990) Psychological aspects of Learning Geometry. In P. Nesher & J. Kilpatrick (eds) 1990. Mathematics and Cognition. Cambridge: Cambridge University Press.

HERSHKOWITZ, R. (1989) Visualization in geometry - two sides of the coin. Focus on learning problems in mathematics. 11, 1: 61-76.

HERTZOG, C. & CARTER, L. (1982) Sex differences in the structure of intelligence: A confirmatory factor analysis. Intelligence. 6, 287-303.

HOFFER, A. (1981) Geometry is more than proof. Mathematics Teacher. 74: 11-18.

HOFFMAN, D.J. (1985) Toetsopstelling met behulp van item- en toets- inligtingskrommes. Suid-Afrikaanse Tydskrif vir Sielkunde. 15,3: 90-96.

HOWSON, A.G. & KAHANE, J.P. (eds) (1986) School mathematics in the 1990s. Cambridge: Cambridge University Press.

HOZ, R. (1981) The effects of rigidity on school geometry learning. Educational Studies in Mathematics. 12: 171-190.

HULL, L.W.H. (1969) The superstition of educated men. Mathematics Teaching. 43: 26-33.

HUMAN, P.G. & NEL, J.H. in samewerking met M.D. de VILLIERS, T.P. DREYER & F.G. WESSELS. (1989) Alternatiewe aanbiedingsstrategieë vir Meetkunde Onderwys: 'n Teoretiese en empiriese studie. Enwous-verslag 2, 1989. Universiteit van Stellenbosch.

HYDE, J.S. (1981) How large are cognitive gender differences? American Psychologist. 36, 892-901.

HYDE, J.S., FENNEMA, E. & LAMON, S.J. (1990) Gender differences in mathematics performance: A meta-analysis. Psychological Bulletin. 107, 2: 139-155.

IRONSON, G.H. (1982) Use of Chi-square and latent trait approaches for detecting item bias. In R.A. Berk (ed) Handbook of methods for detecting test bias. Baltimore: The Johns Hopkins University Press.

IRONSON, G.H. & SUBKOVIAK, M.J. (1979) A comparison of several methods of assessing item bias. Journal of Educational Measurement. 16, 4: 209-225.

JOHNSON, E.S. (1984) Sex differences in problem solving. Journal of Educational Psychology. 76, 6: 1359-1371.

JOHNSON, E.S. & MEADE, A.C. (1987) Developmental patterns of spatial ability: An early sex difference. Child Development. 58: 725-740.

JONES, J.S. (1969) The history of mathematics as a teaching tool. Thirty-first yearbook. Washington: National Council of Teachers of Mathematics.

KABANOVA-MELLER, E.N. (1950) The role of the diagram in the application of geometric theorems. In J. Kilpatrick & I. Wirszup (eds) 1970. Soviet studies in the psychology of learning and teaching mathematics. Volume IV. Chicago: University of Chicago.

KAPADIA, R. (1985) Bring back geometry. The Mathematical Intelligencer. 7, 2: 53-65.

KIFER, E. & ROBITAILLE, D.F. (1989) Attitudes, preferences and Opinions. In D.F. Robitaille & R.A. Garden (eds) 1989. The IEA study of mathematics II: Contexts and Outcomes of school mathematics. New York: Pergamon Press.

KIMBALL, M.M. (1989) A new perspective on Women's math achievement. Psychological Bulletin. 105, 2: 198-214.

KLINE, M. (1980) Mathematics: The loss of certainty. New York: Oxford University Press.

KRIEL, D.J. (1975) Die verdere onderrig van meetkunde. Sintetiese meetkunde in die senior sekondêre fase. Spectrum. 13, 3: 32-34.

KRUTETSKII, V.A. (1976) The psychology of mathematical abilities in schoolchildren. Chicago: University of Chicago.

LAKATOS, I (1967) A renaissance of empiricism in the recent philosophy of mathematics. In T. Tymoczko (ed) 1985. New directions in the philosophy of mathematics. Boston: Birkhäuser.

LAKATOS, I. (1976) Proofs and Refutations. Cambridge: Cambridge University Press.

LAKATOS, I. (1979) What does a mathematical proof prove ? In Tymoczko, T. (ed) 1985. New directions in the philosophy of mathematics. Boston: Birkhäuser.

LANG, B. & RUANE, P. (1981) Geometry in English secondary schools. Educational Studies in Mathematics. 12: 123-132.

LEAN, G. & CLEMENTS, M.A. (1981) Spatial ability, visual imagery and mathematical performance. Educational Studies in Mathematics. 12: 267-299.

LEDER, G.C. (1993) Reconciling affective and cognitive aspects of mathematics learning: reality or a pious hope ? In Proceedings of the Seventeenth International Conference - Psychology of Mathematics Education. Volume I. Tsukuba, Japan: University of Tsukuba.

LEDER, G.C. (1992) Mathematics and gender: Changing perspectives. In Grouws, D.A. (ed) 1992. Handbook of Research on Mathematics Teaching and Learning. New York: Macmillan publishing company.

LEDER, G.C. (1990) Gender differences in mathematics: An overview. In E. Fennema and G.C. Leder (eds) 1990. Mathematics and gender. New York: Columbia University, Teachers College Press.

LEDER, G.C. (1986) Mathematics learning and socialization processes. In L. Burton (ed) 1986. Girls into maths can go. London: Holt, Rinehart and Winston.

LEDER, G.C. & FENNEMA, E. (1990) Gender differences in mathematics: A synthesis. In E. Fennema and G. Leder (eds) 1990. Mathematics and gender. New York: Teachers College Press, Columbia University.

LE ROUX, P.G. (1986) Enkele gegewens aangaande die medaljewenners van die Suid-Afrikaanse Wiskunde-olimpiade van 1966 - 1983. Verrigtinge: Agtste Nasionale

Kongres van die Wiskunde Genootskap van Suid-Afrika. 1986: 222-235. Stellenbosch: Universiteit van Stellenbosch.

LINN, M.C. & PETERSEN, A.C. (1985) Emergence and characterization of sex differences in spatial ability: A meta-analysis. Child Development. 56, 1479 - 1498.

LORD, F.M. (1977a) A study of item bias, using item characteristic curve theory. In Ype H. Poortinga (ed) 1977. Basic problems in cross-cultural psychology. Amsterdam: Swets and Zeitlinger B.V.

LORD, F.M. (1977b) Practical applications of item characteristic curve theory. Journal of Educational Measurement. 14, 2: 117-138.

LORD, F.M. & NOVICK, M.R. (1968) Statistical theories of mental test scores. Massachusetts: Addison-Wesley.

LOVE, E. (1989) Raising some questions on geometry. Mathematics Teaching. 129: 3-6.

LUMMIS, M. & STEVENSON, H.W. (1990) Gender differences in beliefs and achievement: A cross-cultural study. Developmental Psychology. 26,2: 254-263.

LYNN, R. (1992) Sex differences on the differential aptitude test in British and American adolescents. Educational Psychology. 12, 2: 101-106.

MACCOBY, E.E. & JACKLIN, C.N. (1974) The psychology of sex-differences. Stanford, CA: Stanford University Press.

MALAN, A.P. (1971) Gefundeerde onderrig in meetkunde. Spectrum. 9, 4: 251-252.

MALAN, A.P. (1967) Vernuwing in matrikulasie-eksamen. Spectrum. 5, 1: 38-39.

MALAN, F.P.R. (1986) Onderrigstrategieë vir die oorgang van partisiedenke na hiërargiese denke in die klassifikasie van vierhoeke: enkele gevallestudies. Interne ENWOUS-verslag no. 3. Stellenbosch: Universiteit van Stellenbosch.

MARSHALL, S.P. (1983) Sex differences in mathematical errors: An analysis of distractor choices. Journal for Research in Mathematics Education. 14, 4: 325-336.

MASINGILA, J.O. (1993) Secondary geometry: A lack of evolution. School Science and Mathematics. 1: 38-44.

MASTERS, G.N. (1994) Towards effective assessment of educational attainment in the classroom and schools. Paper presented at the Second International Conference on Educational Evaluation and Assessment, July 1994. Pretoria: University of South Africa.

MAYBERRY, J. (1983) The van Hiele levels of geometric thought in undergraduate preservice teachers. Journal for Research in Mathematics Education. 14, 1: 58-69.

McCARTHY, S.A. (1992) Mathematical achievement and gender. Curriculum and Teaching. 7, 2: 47-53.

McGEE, M.G. (1979) Human spatial abilities: Psychometric studies and environmental, genetic hormonal, and neurological influences. Psychological Bulletin. 86, 5: 889-918.

McLARTY, J.R., NOBLE, A.C. & HUNTLEY, R.M. (1989) Effects of item wording on sex bias. Journal of Educational Measurement. 26, 3: 285-293.

McLEOD, D.B. (1992) Research on affect in mathematics education: A reconceptualization. In D.A. Grouws (ed) 1992. Handbook of research on mathematics teaching and learning. New York: MacMillan Publishing company.

McLEOD, D.B. (1988) Affective issues in mathematical problem solving: Some theoretical considerations. Journal for Research in Mathematics Education. 19, 2: 134-141.

MEECE, J.L., PARSONS, J.E., KACZALA, C.M., GOFF, S.B. & FUTTERMAN, R. (1982) Sex differences in math achievement: Toward a model of academic choice. Psychological Bulletin. 91, 2: 324-348.

MIWA, T. (1992) School mathematics in Japan and the U.S.: Focusing on recent trends in elementary and lower secondary school. In I. Wirszup & R. Streit (eds) 1992. Developments in school mathematics education around the world. Reston: National Council of Teachers of Mathematics.

MORITZ, R.E. (1958) On mathematics and mathematicians. New York: Dover Publications Inc.

NICHOLSON, M.J. (1985) Comments on the discussions. Subgroup: Geometry for the secondary school. In A. Olivier (ed) 1985. The Australian Experience: Impressions of ICME 5. The Mathematical Association of South Africa.

NOHDA, N. (1992) Geometry teaching in Japanese school mathematics. Pythagoras. 28: 18-25.

NOSS, R. (1987) Children's learning of geometrical concepts through Logo. Journal for Research in Mathematics Education. 18, 343-362.

NYANGENI, N.P. (1993) Gender differences in achievement in, and attitude towards mathematics of standard ten pupils in the Umtata district of Transkei. Unpublished M. Ed. research report, University of Transkei, Umtata.

O'DAFFER, P.G. (1990) Inductive and deductive reasoning. Mathematics Teacher. 84, 5: 378-384.

OOSTHUIZEN, W.L., SWART, J. & GILDENHUYS, D.G. (1988) Vakdidaktiek: Wiskunde vir die sekondêre skool. Pretoria: De Jager-Haum.

OSSERMAN, R. (1983) Structure vs. substance: The fall and rise of geometry. In M. Zweng (ed) 1983. Proceedings of the Fourth International Congress on Mathematical Education. Boston: Birkhäuser.

OSTERLIND, S.J. (1983) Test item bias. In J.L. Sullivan & R.G. Niemi (eds) 1983. Quantitative applications in the social sciences. Beverly Hills: Sage Publications.

OWEN, K. (1989) Test and item bias: The suitability of the Junior Aptitude Tests as a common test battery for white, Indian and black pupils in standard 7. Report P-96. Pretoria: Human Sciences Research Council.

PASK, G. (1976) Styles and strategies of learning. British Journal of Educational Psychology. 46: 128-148.

PATTISON, P. & GRIEVE, N. (1984) Do spatial skills contribute to sex differences in different types of mathematical problems? Journal of Educational Psychology. 76, 4: 678-689.

PEREIRA-MENDOZA, L. & ROBBINS, M. (1977) A study of the objectives of high school geometry as perceived by teachers and university mathematics educators. School Science and Mathematics. 77, 3: 189-195.

PETERSON, K., BURTON, G. & BAKER, D. (1983) Geometry students' role-specific self-concept: Success, teacher and sex differences. Journal of Educational Research. 77, 2: 122-126.

PIAGET, J. (1970) Science of education and the psychology of the child. New York: Orion Press.

PLAKE, B.S., PATIENCE, B.M. & WHITNEY, D.R. (1988) Differential item performance in mathematics achievement test items: Effect on item arrangement. Educational and Psychological Measurement. 48, 885-894.

POLYA, G. (1957) How to solve it. 2nd Edition. New Jersey: Princeton University Press.

POSAMENTIER, A.S. (1989) Geometry: A remedy for the malaise of middle school mathematics. Mathematics Teacher. 82, 9: 678-680.

PRAWAT, R.S. & FLODEN, R.E. (1994) Philosophical perspectives on constructivist views of learning. Educational Psychology. 29, 1: 37-48.

PRESMEG, N.C. (1993) Mathematics - A bunch of formulas ? Interplay of beliefs and problem solving styles. Proceedings of the Seventeenth International Conference - Psychology of Mathematics Education. Volume III. Tsukuba, Japan: University of Tsukuba.

PRESMEG, N.C. (1992) Curriculum developme UK, USA and other countries. In M. Moodley, R.A. Njisane & N.C. Presmeg (eds.) 1992. Mathematics education for in-service and pre-service teachers. Pietermaritzburg: Shuter & Shooter.

PRESMEG, N.C. (1986) Visualization in school mathematics. For the Learning of Mathematics. 6, 3: 42-46.

PRETORIUS, G.J.P. (1989) Die verband tussen affektiewe stabiliteit en effektiewe wiskunde onderrig en leer in die sekondêre skool. Ongepubliseerde Ph.D. verhandeling. Pretoria: Universiteit van Pretoria.

PRINS, E. (1992) Veelkeusevrae in senior sekondêre wiskunde? Pythagoras. 30: 33-35.

RAND, A. (1986) Measurement. In H. Binswanger (ed) 1986. The Ayn Rand lexicon. New York: New American Library.

REDDY, D. (1987) An investigation of the extent to which the learning of informal geometry, specifically Euclidean transformation geometry, contributes to the meaningful acquisition of geometric facts. A qualitative study confined to the transformations that generate quadrilaterals. Unpublished M.Ed. dissertation. Cape Town: University of Cape town.

REYES, L.H. (1984) Affective variables and mathematics education. The Elementary School Journal. 84: 558-581.

REYNOLDS, R.C. (1982) Methods for detecting construct and predictive bias. In R.A. Berk (ed) 1982. Handbook of methods for detecting test bias. London: Johns Hopkins University Press.

ROBITAILLE, D.F. (1989) Students' achievements: Population A. In D.F. Robitaille & R.A. Garden (eds) 1989. The IEA study of mathematics II: Contexts and outcomes of school mathematics. New York: Pergamon Press.

ROBITAILLE, D.F. & GARDEN, R.A. (eds) (1989) The IEA study of mathematics II: Contexts and outcomes of school mathematics. New York: Pergamon Press.

ROBITAILLE, D.F. & TAYLOR, A.R. (1989) Changes in patterns of achievement between the first an second mathematics studies. In D.F. ROBITAILLE & R.A. GARDEN (eds) 1989. The IEA study of mathematics II: Contexts and outcomes of school mathematics. Oxford: Pergamon press.

ROBITAILLE, D.F. & TRAVERS, K.J. (1992) International studies of achievement in mathematics. In D.A. Grouws (ed) 1992. Handbook of research on mathematics teaching and learning. New York: Macmillan publishing company.

RUDISILL, E.M. & MORRISON, L.J. (1989) Sex-differences in mathematics achievement: An emerging case for physiological factors. School Science and Mathematics. 89, 7: 571-577.

SCHEUNEMAN, J.D. (1982) A Posteriori analyses of biased items. In R. A. Berk (ed) 1982. Handbook of methods for detecting test bias. Baltimore: Johns Hopkins University Press.

SCHIDDELL, J.L. & ETHINGTON, C.A. (1994) Teaching of geometry in the eighth grade mathematics curriculum: Findings from the second international mathematics study. Focus on Learning Problems in Mathematics. 16, 2: 51-61.

SCHOENFELD, A.H. (1985) Mathematical problem solving. New York: Academic Press Inc.

SCHUMANN, H. & de VILLIERS, M.D. (1993) Continuous variation of geometric figures: interactive theorem finding and problems in proving. Pythagoras. 31: 9-20.

SCHUSTER, S. (1975) On the teaching of geometry: A potpouri. Euclides. 50, 4/5: 167-176.

SCHUTTE, H.J. (1983) Greek mathematical thought and modern mathematics. Spectrum. 21, 2: 44-46.

SCOTT-HODGETTS, R. (1986) Girls and mathematics: The negative implications of success. In L. Burton (ed) 1986. Girls into maths can go. London: Holt, Rinehart and Winston.

SENK, S. (1989) Van Hiele levels and achievement in writing geometry proofs. Journal for Research in Mathematics Education. 20, 3: 309-321.

SENK, S. (1985) How well do students write geometry proofs ? Mathematics Teacher. 78: 448-456.

SENK, S. & USISKIN, Z. (1983) Geometry proof writing: A new view on sex differences in mathematics ability. American Journal of Education. 91: 187-201.

SHAUGHNESSY, J.M. & BURGER, W.F. (1985) Spadework prior to deduction in geometry. Mathematics Teacher. 78, 6: 419-427.

SHERARD, W.H. (1981) Why is geometry a basic skill ? Mathematics teacher. 74: 19-21.

SHERMAN, J. (1980) Mathematics, spatial visualization, and related factors: changes in girls and boys, grades 8 - 11. Journal of Educational Psychology. 72, 4: 476 - 482.

SHERMAN, J. and FENNEMA, E. (1977) The study of mathematics by high school girls and boys: Related variables. American Educational Research Journal. 14, 2: 159-168.

SHIBATA, T. (1973) The role of axioms in contemporary mathematics and in mathematical education. In A.G. Howson (ed) 1973. Developments in mathematical education. Proceedings of the Second International Congress on Mathematical Education. Cambridge: Cambridge University Press.

SMITH, E.C. & DE VILLIERS, M.D. (1990) Kritiese vergelyking van twee van Hiele-toetsinstumente. Suid-Afrikaanse Tydskrif vir Opvoedkunde. 10, 1: 68-74.

SMITH, S.E. & WALKER, W.J. (1988) Sex differences on New York State Regents examinations: Support for the differential course-taking hypothesis. Journal for Research in Mathematics Education. 19, 81-85.

SNYMAN, J.J. (1984) 'n Strategie vir probleemoplossing in Meetkunde-onderrig. Spectrum. 22, 2:33-37.

SNYMAN, J.J. (1982) Kreatiewe denke en begripsvorming as komponente van probleemoplossing by Meetkunde-onderrig in die senior-sekondêre fase. Ongepubliseerde M. ed. verhandeling. Johannesburg: Randse Afrikaanse Universiteit.

STOKER, J. (1991) Developing active mathematical thinkers through language interaction: A curriculum development project in primary mathematics. In A. Olivier (ed) 1991. Fourteenth National Convention on Mathematics and natural Science Education - Proceedings. Cape Town: CTP Book Printers.

SUBKOVIAK, M.J., MACK, J.S., IRONSON, G.H. & CRAIG, R.D. (1984) Empirical comparison of selected item bias detection procedures with bias manipulation. Journal of Educational Measurement. 21,1: 49-58.

SUYDAM, M.N. (1985) The shape of instruction geometry: Some highlights from research. Mathematics Teacher. 76, 6: 481-486.

SWANEPOEL, C.H. (1990) Bemeestering van basiese meetkundekonsepte: geslagsverskille. Suid-Afrikaanse Tydskrif vir Opvoedkunde. 10: 5-6.

TALL, D. (1991) Intuition and rigour: The role of visualization in the calculus. In W. Zimmerman & S. Cunningham (eds) 1991. Visualization in teaching and learning mathematics. USA: Mathematical Association of America.

TARTRE, L.A. (1990) Spatial skills, gender and mathematics. In E. Fennema and G.C. Leder (eds) 1990. Mathematics and gender. New York: Teachers College Press.

TEPPO, A. (1991) Van Hiele levels of geometric thought revisited. Mathematics Teacher. 84, 3: 210-221.

THOM, R. (1973) Modern mathematics: does it exist ? In A.G.Howson (ed) Developments in mathematical education. Proceedings of the Second International Congress on Mathematical Education. Cambridge: Cambridge University Press.

THOM, R. (1971) Modern mathematics: An educational and philosophic error? American Scientist. 59: 695-699.

TOUMASIS, C. (1991) Geometry is motion: A dynamic approach to the teaching of school Euclidean geometry. International Journal for Mathematics Education in Science and Technology. 22, 2: 229-242.

TYMOCZKO, T. (ed) (1985) New directions in the philosophy of mathematics. Boston: Birkhäuser.

USISKIN, Z. (1987) Resolving the continuing dilemmas in school geometry. In M.M. Lindquist and A.P. Shulte (eds) Learning and teaching geometry, K-12, 1987 yearbook. Reston, Virginia: National Council of Teachers of Mathematics.

USISKIN, Z. (1982) Van Hiele levels and achievement in secondary school geometry. (Final report of the Cognitive Development and Achievement in Secondary School Geometry Project) Chicago: University of Chicago. ED 220 288.

USISKIN, Z. & SENK, S. (1990) Evaluating a test of van Hiele levels: A response to Crowley and Wilson. Journal for Research in Mathematics Education. 21, 243-245.

VAN DEN BERG, A.R., (1994) Multiple choice questions - what they can and cannot measure. Paper presented at the Second International Conference on Educational Evaluation and Assessment, July 1994. University of South Africa, Pretoria.

VAN HIELE-GELDOF, D. (1957) De Didaktiek van de Meetkunde in de Eerste Klas van het V.H.M.O. Amsterdam: J.M. Meulenhoff et al.

VAN HIELE, P.M. (1973) Begript en Inzicht. Purmerend, Nederland: Muuses.

VAN ZYL, A.J. (1942) Mathematics at the cross-roads. Cape Town: Maskew Miller Ltd.

VISSER, D. (1989) Mathematics - the critical occupational filter for women. South African Journal of Science. 85: 212-214.

VISSER, D. (1988a) Mathematics anxiety and continued participation in mathematics. Spectrum. 26, 2: 38-40.

VISSER, D. (1988b) The influence of socialization and emotional factors on mathematics achievement and participation. In A. Borbás (ed) Proceedings of the Twelfth Annual Conference of the International Group for the Psychology of Mathematics Education. Volume II. Veszprèm, Hungary: OOK Printing House.

VISSER, D. (1987) Sex differences in adolescent mathematics behaviour. South African Journal for Psychology. 17, 4: 137-145.

VISSER, D. (1985) Vroue en Wiskunde: Fokus op geslagsverskille. Verslag P-55. Raad vir Geesteswetenskaplike Navorsing. Pretoria.

VOLMINK, J.D. (1988) Acquisition of concepts and construction of meaning in geometry. Unpublished Ph. D. dissertation. Cornell University.

VON GLASERSFELD, E. (1990) An exposition of constructivism: Why some like it ra In R.B. Davis, C.A. Maher & N. Noddings (eds) 1990. Constructivist views on the teaching and learning of mathematics. Reston: National Council of Teachers of Mathematics.

VORSTER, J.F. (1988) 'n Evaluering van die bestaande metodes en modelle vir die moderering en gelykstelling van eksamenstandaarde met spesiale verwysing na die gebruik van itembanke. Ongepubliseerde D.Ed. verhandeling. Potchefstroom: Potchefstroom Universiteit vir Christelike Hoër Onderwys.

WAGEMAKER, H. & KNIGHT, G. (1989) The students of mathematics. In D.F. Robitaille & R.A. Garden (eds) 1989. The IEA study of mathematics II: Contexts and outcomes of school mathematics. New York: Pergamon Press.

WAINER, H. & STEINBERG, L.S. (1992) Sex differences in performance on the mathematics section of the scholastic aptitude test: A bidirectional validity study. Harvard Educational Review. 62, 3: 323-336.

WALKERDINE, V. (1989) Counting girls out. London: Virago Press.

WARM, T.A. (1978) A primer of item response theory. Oklahoma: U.S. Coastal Guard Institute.

WARM, T.A. (1985) Weighted likelihood estimation of ability in item response theory with tests of finite length. Technical report. Oklahoma: U.S. Coast Guard Institute.

WEBB, J.H. (1989) Multiple-choice questions in mathematics. South African Journal of Education. 9, 1: 216-218.

WHITMAN, N.C., HASHIMOTO, Y., ISODA, M., IIJIMA, Y. & NOHDA, N. (1993) The attained geometry curriculum in Japan and Hawaii relative to the van Hiele level theory. Proceedings of the Seventeenth International Conference - Psychology of Mathematics Education. Tsukuba, Japan: University of Tsukuba.

WILSON, M. (1990) Measuring a van Hiele geometry sequence: A reanalysis. Journal for Research in Mathematics Education. 21: 231-237.

WOLLEAT, P.L., PEDRO, D.J., BECKER, A.D. & FENNEMA, E. (1980) Sex differences in high school students' causal attributions of performance in mathematics. Journal for Research in Mathematics Education. 11: 356-366.

WOLFAARDT, J.B. (1990) Achievement test item analysis: A comparison of traditional and modern methods. South African Journal of Psychology. 20, 4: 282-285.

YAKIMANSKAYA, I.S. (1991) The development of spatial thinking in school children. In P. S. Wilson & E.J. Davis (eds) 1991. Soviet Studies in Mathematics Education. Volume 3. Reston, Virginia: National Council of Teachers of Mathematics.

YOUNG, J.W. (1991) Gender bias in predicting college academic performance: A new approach using Item Response Theory. Journal of Educational Measurement. 28, 1: 37-47.

ZAMBO, R. & FOLLMAN, J. (1994) Gender-related differences in problem solving at the 6th and the 8th grade levels. Focus on Learning Problems in Mathematics. 16, 2: 20-38.

ZIMMERMAN, W. & CUNNINGHAM, S. (eds) (1991) Visualization in teaching and learning mathematics. USA: The Mathematical Association of America.

ZYKOVA, V.I. (1954) The psychology of Sixth-grade pupils' mastery of geometric concepts. In J. Kilpatrick & I. Wirszup (eds) 1969. Soviet studies in the psychology of learning and teaching mathematics. Volume 1. Chicago: University of Chicago.

APPENDIX A_1

Raw scores (sample)

WARMEST 5
RAW SCORE FREQUENCY DISTRIBUTION

TESTID=CRON GRPSEL=3388 N= 816 MEAN=11.06 STD DEV= 4.26 KR-20= 0.81

RAW	FREQ	RAW	FREQ	RAW	FREQ	RAW	FREQ
0	2	20	18				
1	3						
2	3						
3	9						
4	23						
5	29						
6	57						
7	60						
8	66						
9	69						
10	58						
11	75						
12	73						
13	52						
14	49						
15	47						
16	39						
17	32						
18	35						
19	19						

KR14- = 0.827
STANDARD ERROR = 1.770
KR9- = 0.829

APPENDIX A_2

Item parameter estimates (sample)

WARMEST 5; ITEM PARAMETER ESTIMATES
TESTID=CRCN GRPSEL=&&&& N= 816 #ITEMS= 20 MEAN_Z=0.000 SD_Z= 1.000 METHODS: WLE(THETA), BME(A,B,C)

ITEM	KEY	A	B	C	PVAL	QVAL	FITABC
1	4	0.310	-1.191	0.206	0.718	0.251	0.063
2	1	1.362	1.271	0.168	0.313	0.388	0.051
3	2	1.171	-0.734	0.123	0.747	0.502	0.341
4	3	1.107	0.621	0.178	0.441	0.486	0.220
5	4	0.709	0.063	0.194	0.586	0.415	0.315
6	3	1.425	0.723	0.127	0.362	0.534	0.189
7	2	1.255	-0.124	0.188	0.625	0.528	0.229
8	2	0.960	-1.659	0.173	0.896	0.331	0.271
9	2	1.038	0.391	0.123	0.460	0.521	0.117
10	3	0.979	-0.984	0.170	0.797	0.431	0.200
11	1	1.425	0.893	0.160	0.351	0.475	0.311
12	3	1.208	1.959	0.176	0.344	0.436	0.099
13	3	1.311	-0.313	0.103	0.637	0.561	0.271
14	2	0.972	0.804	0.221	0.441	0.417	0.085
15	1	1.425	0.796	0.088	0.314	0.543	0.443
16	4	1.116	-0.517	0.137	0.696	0.505	0.289
17	1	1.257	-1.039	0.145	0.818	0.436	0.269
18	4	1.045	0.607	0.095	0.385	0.529	0.192
19	3	1.355	-0.613	0.103	0.716	0.542	0.545
20	1	0.804	0.966	0.213	0.419	0.385	0.022

APPENDIX B_1

Raw scores (males)

WARMEST 5
RAW SCORE FREQUENCY DISTRIBUTION

TESTID=CRON GRPSEL=MALE N= 410 MEAN=11.09 STD DEV= 4.29 KR-20= 0.81

RAW	FREQ	RAW	FREQ	RAW	FREQ	RAW	FREQ
0	2	20	12				
1	1						
2	2						
3	4						
4	12						
5	14						
6	30						
7	27						
8	35						
9	30						
10	27						
11	40						
12	36						
13	31						
14	27						
15	26						
16	13						
17	16						
18	16						
19	9						

KR14- = 0.828
STANDARD ERROR = 1.781
KR8- = 0.830

APPENDIX B_2

Item parameter estimates (males)

WARMEST 5: ITEM PARAMETER ESTIMATES

TESTID=CRON GRPSEL=MALE N= 410 #ITEMS= 20 MEAN_Z=0.000 SD_Z= 1.000 METHODS: WLE(THETA), BME(A,B,C)

ITEM	KEY	A	B	C	PVAL	DVAL	FITABC
1	4	0.380	-1.064	0.206	0.724	0.262	0.069
2	1	0.957	1.199	0.168	0.329	0.419	0.108
3	2	1.176	-0.708	0.170	0.754	0.495	0.133
4	3	0.850	0.375	0.175	0.500	0.472	0.283
5	4	0.817	0.011	0.201	0.595	0.435	0.325
6	3	1.261	0.791	0.130	0.346	0.512	0.185
7	2	1.387	-0.237	0.169	0.639	0.565	0.208
8	2	0.674	-2.023	0.197	0.895	0.267	0.108
9	2	0.960	0.557	0.156	0.446	0.481	0.045
10	3	1.050	-0.842	0.199	0.785	0.447	0.084
11	1	1.191	0.800	0.188	0.398	0.453	0.192
12	3	1.031	0.852	0.170	0.383	0.448	0.045
13	3	1.380	-0.217	0.127	0.612	0.575	0.092
14	2	1.009	0.692	0.167	0.412	0.470	0.035
15	1	1.474	0.777	0.112	0.324	0.531	0.260
16	4	1.159	-0.487	0.139	0.685	0.506	0.367
17	1	1.093	-0.528	0.172	0.788	0.433	0.046
18	4	1.082	0.656	0.093	0.359	0.537	0.258
19	3	1.468	-0.599	0.110	0.712	0.563	0.591
20	1	0.653	1.297	0.220	0.400	0.326	0.108

APPENDIX C_1

Raw scores (females)

WARMEST 5
RAW SCORE FREQUENCY DISTRIBUTION

TESTID=CRON GRPSEL=VROU N= 406 MEAN=11.04 STD DEV= 4.24 KR-20= 0.81

RAW	FREQ	RAW	FREQ	RAW	FREQ	RAW	FREQ
0	0	20	6				
1	2						
2	1						
3	5						
4	11						
5	15						
6	27						
7	33						
8	31						
9	39						
10	30						
11	34						
12	37						
13	21						
14	22						
15	21						
16	26						
17	16						
18	19						
19	10						

KR14- = 0.829
STANDARD ERROR = 1.747
KR8- = 0.829

APPENDIX C_2

Item parameter estimates (males)

WARMEST 5: ITEM PARAMETER ESTIMATES

ITEM PARAMETER ESTIMATES

TESTID=CRON GRPSEL=VRCU N= 406 #ITEMS= 20 MEAN_Z=0.000 SD_Z= 1.000 METHODS: WLE(THETA), BME(A,B,C)

ITEM	KEY	A	B	C	PVAL	QVAL	FITABC
1	4	0.290	-1.215	0.202	0.714	0.241	0.054
2	1	1.433	1.237	0.169	0.296	0.359	0.181
3	2	1.183	-0.660	0.161	0.741	0.511	0.126
4	3	1.744	0.773	0.164	0.392	0.509	0.246
5	4	0.670	0.202	0.212	0.576	0.396	0.103
6	3	1.873	0.716	0.149	0.379	0.557	0.252
7	2	1.149	0.013	0.209	0.611	0.491	0.037
8	2	1.668	-1.355	0.163	0.897	0.397	0.526
9	2	1.241	0.352	0.128	0.473	0.562	0.193
10	3	1.034	-1.021	0.175	0.808	0.415	0.055
11	1	2.014	0.965	0.134	0.303	0.504	0.192
12	3	1.696	1.167	0.177	0.305	0.405	0.095
13	3	1.344	-0.291	0.165	0.660	0.549	0.195
14	2	1.202	0.911	0.278	0.468	0.368	0.241
15	1	1.518	0.842	0.086	0.305	0.555	0.109
16	4	1.153	-0.421	0.203	0.704	0.506	0.157
17	1	1.724	-1.070	0.171	0.847	0.444	0.150
18	4	1.206	0.674	0.163	0.414	0.524	0.323
19	3	1.201	-0.546	0.173	0.719	0.522	0.190
20	1	1.204	0.735	0.215	0.441	0.446	0.234

APPENDIX D

Item parameters (males on scale of females)

MINEQUAT: MALES ON SCALE OF FEMALES
ANIT=20 A=0.92488 B=-0.0524685
SUMOFM=0.05039003 INDEXA=0.3477413 INDEXB=0.2199761

ITEMPARAMETER OUTPUT OF PROGRAM MINEQUAT

OBS	TESTID	SELFACT	I	A	B	C
1	CRON	MANL	1	0.411	-1.037	0.206
2	CRON	MANL	2	1.035	1.056	0.168
3	CRON	MANL	3	1.272	-0.707	0.170
4	CRON	MANL	4	0.919	0.294	0.175
5	CRON	MANL	5	0.883	-0.042	0.201
6	CRON	MANL	6	1.363	0.679	0.130
7	CRON	MANL	7	1.500	-0.272	0.169
8	CRON	MANL	8	0.729	-1.924	0.197
9	CRON	MANL	9	1.038	0.463	0.166
10	CRON	MANL	10	1.135	-0.831	0.199
11	CRON	MANL	11	1.288	0.687	0.108
12	CRON	MANL	12	1.115	0.736	0.170
13	CRON	MANL	13	1.492	-0.253	0.127
14	CRON	MANL	14	1.091	0.588	0.167
15	CRON	MANL	15	1.594	0.666	0.112
16	CRON	MANL	16	1.253	-0.503	0.139
17	CRON	MANL	17	1.182	-0.911	0.172
18	CRON	MANL	18	1.170	0.554	0.093
19	CRON	MANL	19	1.587	-0.606	0.110
20	CRON	MANL	20	0.706	1.147	0.220

APPENDIX E_1

CTT analysis (males)

SAS

ITEM NR ITEM NR	KEY KORREKTE AFLEIER	DISTRACTER NR AFLEIER NR	FREQ AANTAL	PERS.VAN AANTAL GEDOEN	PERS.VAN TOTALE AANTAL	GEMIDDELD VIR TOETS	VERSKIL (GEMID - TOT.GEMID)	KRITERIUM NR	ITEM-KRITERIUM KOR.	ITEM STANDAARD-AFWYKING	PRODUK
1		NG	0	0.0	0.0	0.0	0.0				
		1	94	23.0	22.9	9.9	-1.2				
		2	7	1.7	1.7	5.7	-5.4				
		3	10	2.5	2.4	7.4	-3.7				
	*.....	4	297	72.8	72.4	11.8	0.7				
		5	0	0.0	0.0	0.0	0.0				
		NB	2	0.5	0.5	0.0	-11.1				
		VERK	113	27.7	27.6	9.3	-1.8				
								1	0.262(0.161)	0.447	0.117
2		NG	1	0.2	0.2	11.0	-0.1				
	*.....	1	135	33.2	32.9	13.7	2.6				
		2	41	10.1	10.0	8.9	-2.2				
		3	15	3.7	3.7	9.4	-1.7				
		4	216	53.1	52.7	10.1	-1.0				
		5	0	0.0	0.0	0.0	0.0				
		NB	2	0.5	0.5	0.0	-11.1				
		VERK	275	67.6	67.1	9.8	-1.3				
								1	0.419(0.322)	0.470	0.197
3		NG	1	0.2	0.2	11.0	-0.1				
		1	6	1.5	1.5	6.5	-4.6				
	*.....	2	309	75.9	75.4	12.3	1.2				
		3	5	1.2	1.2	10.0	-1.1				
		4	87	21.4	21.2	7.4	-3.7				
		5	0	0.0	0.0	0.0	0.0				
		NB	2	0.5	0.5	0.0	-11.1				
		VERK	101	24.8	24.6	7.4	-3.7				
								1	0.495(0.413)	0.431	0.213
4		NG	0	0.0	0.0	0.0	0.0				
		1	21	5.1	5.1	9.4	-1.7				
		2	145	35.5	35.4	9.3	-1.8				
	*.....	3	205	50.2	50.0	13.1	2.0				
		4	37	9.1	9.0	8.4	-2.7				
		5	0	0.0	0.0	0.0	0.0				
		NB	2	0.5	0.5	0.0	-11.1				
		VERK	205	50.2	50.0	9.1	-2.0				
								1	0.472(0.374)	0.500	0.236

SAS

ITEM NR / ITEM NR	KEY / KORREKTE AFLEIER	DISTRACTER NR / AFLEIER NR	FREQ / AANTAL	PERS.VAN AANTAL GEDOEN	PERS.VAN TOTALE AANTAL	GEMIDDELD VIR TOETS	VERSKIL (GEMID - TOT.GEMID)	KRITERIUM NR	ITEM-KRITERIUM KOR.	ITEM STANDAARD-AFWYKING	PRODUK
5		NG	1	0.2	0.2	1.0	-10.1				
		1	55	13.5	13.4	8.5	-2.6				
		2	33	8.1	8.0	9.8	-1.3				
		3	75	18.4	18.3	9.0	-2.1				
	*.....	4	244	60.0	59.5	12.6	1.5				
		5	0	0.0	0.0	0.0	0.0				
		NB	2	0.5	0.5	0.0	-11.1				
		VERK	166	40.8	40.5	8.8	-2.3				
								1	0.435(0.335)	0.491	0.214
6		NG	4	1.0	1.0	11.0	-0.1				
		1	143	35.4	34.9	9.0	-2.0				
		2	60	14.9	14.6	11.3	0.2				
	*.....	3	142	35.1	34.6	14.1	3.0				
		4	59	14.6	14.4	9.0	-2.1				
		5	0	0.0	0.0	0.0	0.0				
		NB	2	0.5	0.5	0.0	-11.1				
		VERK	268	66.3	65.4	9.5	-1.6				
								1	0.512(0.423)	0.476	0.244
7		NG	4	1.0	1.0	12.5	1.4				
		1	39	9.7	9.5	7.8	-3.3				
	*.....	2	262	64.9	63.9	12.9	1.8				
		3	76	18.8	18.5	7.7	-3.4				
		4	27	6.7	6.6	8.3	-2.8				
		5	0	0.0	0.0	0.0	0.0				
		NB	2	0.5	0.5	0.0	-11.1				
		VERK	148	36.6	36.1	7.9	-3.2				
								1	0.565(0.481)	0.480	0.271
8		NG	1	0.2	0.2	13.0	1.9				
		1	5	1.2	1.2	9.0	-2.1				
	*.....	2	367	90.2	89.5	11.5	0.4				
		3	9	2.2	2.2	7.0	-4.1				
		4	26	6.4	6.3	8.2	-2.9				
		5	0	0.0	0.0	0.0	0.0				
		NB	2	0.5	0.5	0.0	-11.1				
		VERK	43	10.6	10.5	7.7	-3.3				
								1	0.267(0.199)	0.306	0.082

SAS ITEM NR ITEM NR	KEY KORREKTE AFLEIER	DISTRACTER NR AFLEIER NR	FREQ AANTAL	PERS.VAN AANTAL GEDOEN	PERS.VAN TOTALE AANTAL	GEMIDDELD VIR TOETS	VERSKIL (GEMID - TOT.GEMID)	KRITERIUM NR	ITEM-KRITERIUM KOR.	ITEM STANDAARD-AFWYKING	PRODUK
9		NG	3	0.7	0.7	8.3	-2.8				
		1	76	18.8	18.5	8.8	-2.3				
	*	2	183	45.2	44.6	13.4	2.3				
		3	82	20.2	20.0	10.0	-1.1				
		4	64	15.8	15.6	9.1	-2.0				
		5	0	0.0	0.0	0.0	0.0				
		NB	2	0.5	0.5	0.0	-11.1				
		VERK	227	56.0	55.4	9.2	-1.9				
								1	0.481(0.385)	0.497	0.239
10		NG	1	0.2	0.2	4.0	-7.1				
		1	41	10.1	10.0	7.1	-4.0				
		2	11	2.7	2.7	7.8	-3.3				
	*	3	322	79.1	78.5	12.1	1.0				
		4	33	8.1	8.0	8.2	-2.8				
		5	0	0.0	0.0	0.0	0.0				
		NB	2	0.5	0.5	0.0	-11.1				
		VERK	88	21.6	21.5	7.4	-3.7				
								1	0.447(0.365)	0.411	0.183
11		NG	2	0.5	0.5	10.5	-0.6				
	*	1	163	40.1	39.8	13.5	2.4				
		2	48	11.8	11.7	10.5	-0.6				
		3	151	37.2	36.9	9.3	-1.8				
		4	44	10.8	10.7	9.7	-1.4				
		5	0	0.0	0.0	0.0	0.0				
		NB	2	0.5	0.5	0.0	-11.1				
		VERK	247	60.8	60.2	9.5	-1.6				
								1	0.453(0.355)	0.489	0.222
12		NG	5	1.2	1.2	11.8	0.7				
		1	73	18.1	17.8	9.8	-1.3				
		2	82	20.3	20.0	10.3	-0.8				
	*	3	157	39.0	38.3	13.6	2.5				
		4	91	22.6	22.2	8.6	-2.5				
		5	0	0.0	0.0	0.0	0.0				
		NB	2	0.5	0.5	0.0	-11.1				
		VERK	253	62.8	61.7	9.5	-1.6				
								1	0.468(0.373)	0.486	0.228

SAS ITEM NR ITEM NR	KEY KORREKTE AFLEIER	DISTRACTER NR AFLEIER NR	FREQ AANTAL	PERS.VAN AANTAL GEDOEN	PERS.VAN TOTALE AANTAL	GEMIDDELD VIR TOETS	VERSKIL (GEMID - TOT.GEMID)	KRITERIUM NR	ITEM-KRITERIUM KOR.	ITEM STANDAARD-AFWYKING	PRODUK
13		NG	4	1.0	1.0	9.3	-1.8				
		1	54	13.4	13.2	7.9	-3.2				
		2	54	13.4	13.2	8.5	-2.6				
	*.....	3	251	62.1	61.2	13.1	2.0				
		4	45	11.1	11.0	7.8	-3.3				
		5	0	0.0	0.0	0.0	0.0				
		NB	2	0.5	0.5	0.0	-11.1				
		VERK	159	39.4	38.3	8.0	-3.1				
								1	0.575(0.491)	0.487	0.280
14		NG	9	2.3	2.2	11.9	0.8				
		1	142	35.6	34.6	10.1	-1.0				
	*.....	2	169	42.4	41.2	13.5	2.4				
		3	75	18.8	18.3	8.1	-3.0				
		4	13	3.3	3.2	8.8	-2.3				
		5	0	0.0	0.0	0.0	0.0				
		NB	2	0.5	0.5	0.0	-11.1				
		VERK	241	60.4	58.8	9.4	-1.7				
								1	0.470(0.374)	0.492	0.231
15		NG	4	1.0	1.0	9.8	-1.3				
	*.....	1	133	32.9	32.4	14.4	3.3				
		2	48	11.9	11.7	9.6	-1.5				
		3	79	19.6	19.3	9.1	-2.0				
		4	144	35.6	35.1	9.8	-1.3				
		5	0	0.0	0.0	0.0	0.0				
		NB	2	0.5	0.5	0.0	-11.1				
		VERK	277	68.6	67.6	9.5	-1.6				
								1	0.531(0.446)	0.468	0.249
16		NG	4	1.0	1.0	9.3	-1.8				
		1	20	5.0	4.9	8.9	-2.1				
		2	34	8.4	8.3	7.9	-3.2				
		3	69	17.1	16.8	7.7	-3.3				
	*.....	4	281	69.6	68.5	12.6	1.5				
		5	0	0.0	0.0	0.0	0.0				
		NB	2	0.5	0.5	0.0	-11.1				
		VERK	129	31.9	31.5	7.9	-3.2				
								1	0.506(0.419)	0.464	0.235

SAS ITEM NR ITEM NR	KEY KORREKTE AFLEIER	DISTRACTER NR AFLEIER NR	FREQ AANTAL	PERS.VAN AANTAL GEDOEN	PERS.VAN TOTALE AANTAL	GEMIDDELD VIR TOETS	VERSKIL (GEMID - TOT.GEMID)	KRITERIUM NR	ITEM-KRITERIUM KOR.	ITEM STANDAARD-AFWYKING	PRODUK
17		NG	5	1.2	1.2	9.8	-2.3				
	*.....	1	323	80.7	78.3	12.1	1.0				
		2	28	7.0	6.8	7.8	-3.3				
		3	21	5.2	5.1	7.3	-3.8				
		4	28	7.0	6.8	7.4	-3.7				
		5	0	0.0	0.0	0.0	0.0				
		NB	5	1.2	1.2	5.6	-5.5				
		VERK	87	21.7	21.2	7.5	-3.6				
								1	0.433(0.351)	0.409	0.177
18		NG	7	1.8	1.7	12.0	0.9				
		1	24	6.0	5.9	9.0	-2.1				
		2	118	29.1	28.3	9.3	-1.8				
		3	111	27.9	27.1	9.5	-1.6				
	*.....	4	147	36.9	35.9	14.2	3.1				
		5	0	0.0	0.0	0.0	0.0				
		NB	5	1.3	1.2	5.6	-5.5				
		VERK	263	66.1	64.1	9.4	-1.7				
								1	0.537(0.450)	0.480	0.258
19		NG	0	0.0	0.0	0.0	0.0				
		1	11	2.7	2.7	6.8	-4.3				
		2	85	21.0	20.7	7.1	-3.9				
	*.....	3	292	72.1	71.2	12.6	1.5				
		4	17	4.2	4.1	8.8	-2.3				
		5	0	0.0	0.0	0.0	0.0				
		NB	5	1.2	1.2	5.6	-5.5				
		VERK	118	29.1	23.8	7.3	-3.8				
								1	0.563(0.484)	0.453	0.255
20		NG	0	0.0	0.0	0.0	0.0				
	*.....	1	164	40.9	40.0	12.8	1.7				
		2	57	14.2	13.9	8.2	-2.8				
		3	115	28.7	28.0	11.3	0.2				
		4	65	16.2	15.9	9.4	-1.7				
		5	0	0.0	0.0	0.0	0.0				
		NB	9	2.2	2.2	7.9	-3.2				
		VERK	246	61.3	60.0	9.9	-1.1				
								1	0.326(0.218)	0.490	0.160

APPENDIX F_1

CTT analysis (females)

SAS ITEM NR ITEM NR	KEY KORREKTE AFLEIER	DISTRACTER NR AFLEIER NR	FREQ AANTAL	PERS.VAN AANTAL GEDOEN	PERS.VAN TOTALE AANTAL	GEMIDDELD VIR TOETS	VERSKIL (GEMID - TOT.GEMID)	KRITERIUM NR	ITEM-KRITERIUM KOR.	ITEM STANDAARD-AFWYKING	PRODUK
1		NG	0	0.0	0.0	0.0	0.0				
		1	101	24.8	24.8	9.9	-1.2				
		2	5	1.2	1.2	7.2	-3.8				
		3	11	2.7	2.7	6.4	-4.7				
	*.....	4	290	71.3	71.3	11.7	0.6				
		5	0	0.0	0.0	0.0	0.0				
		NB	0	0.0	0.0	0.0	0.0				
		VERK	117	28.7	28.7	9.4	-1.6				
								1	0.241(0.137)	0.453	0.109
2		NG	1	0.2	0.2	16.0	5.0				
	*.....	1	121	29.8	29.7	13.4	2.3				
		2	57	14.0	14.0	9.2	-1.8				
		3	20	4.9	4.9	8.7	-2.7				
		4	208	51.2	51.1	10.4	-0.7				
		5	0	0.0	0.0	0.0	0.0				
		NB	0	0.0	0.0	0.0	0.0				
		VERK	286	70.4	70.3	10.1	-1.0				
								1	0.357(0.257)	0.457	0.163
3		NG	3	0.7	0.7	12.3	1.3				
		1	6	1.5	1.5	6.2	-4.9				
	*.....	2	301	74.5	74.0	12.3	1.3				
		3	8	2.0	2.0	7.9	-3.2				
		4	89	22.0	21.9	7.3	-3.8				
		5	0	0.0	0.0	0.0	0.0				
		NB	0	0.0	0.0	0.0	0.0				
		VERK	106	26.2	26.0	7.4	-3.6				
								1	0.510(0.427)	0.439	0.224
4		NG	1	0.2	0.2	14.0	3.0				
		1	15	3.7	3.7	9.1	-1.9				
		2	165	40.6	40.5	9.5	-1.5				
	*.....	3	156	38.4	38.3	13.8	2.7				
		4	70	17.2	17.2	9.0	-2.1				
		5	0	0.0	0.0	0.0	0.0				
		NB	0	0.0	0.0	0.0	0.0				
		VERK	251	61.8	61.7	9.4	-1.7				
								1	0.507(0.414)	0.486	0.246

SAS ITEM NR ITEM NR	KEY KORREKTE AFLEIER	DISTRACTER NR AFLEIER NR	FREQ AANTAL	PERS.VAN AANTAL GEDOEN	PERS.VAN TOTALE AANTAL	GEMIDDELD VIR TOETS	VERSKIL (GEMID - TOT.GEMID)	KRITERIUM NR	ITEM-KRITERIUM KOR.	ITEM STANDAARD-AFWYKING	PRODUK
5		NG	1	0.2	0.2	16.0	5.0				
		1	43	10.6	10.6	7.6	-3.5				
		2	23	5.7	5.7	7.8	-3.2				
		3	105	25.9	25.8	9.9	-1.1				
	*.....	4	235	57.9	57.7	12.5	1.4				
		5	0	0.0	0.0	0.0	0.0				
		NB	0	0.0	0.0	0.0	0.0				
		VERK	172	42.4	42.3	9.1	-2.0				
								1	0.395(0.290)	0.494	0.195
6		NG	8	2.0	2.0	11.0	-0.0				
		1	129	32.3	31.7	9.6	-1.4				
		2	43	10.8	10.6	10.1	-0.9				
	*.....	3	154	38.6	37.8	14.1	3.0				
		4	73	18.3	17.9	7.7	-3.3				
		5	0	0.0	0.0	0.0	0.0				
		NB	0	0.0	0.0	0.0	0.0				
		VERK	253	63.4	62.2	9.2	-1.8				
								1	0.557(0.471)	0.485	0.270
7		NG	2	0.5	0.5	9.0	-2.0				
		1	45	11.1	11.1	8.8	-2.3				
	*.....	2	248	61.2	60.9	12.7	1.7				
		3	93	23.0	22.9	8.2	-2.9				
		4	19	4.7	4.7	9.1	-2.0				
		5	0	0.0	0.0	0.0	0.0				
		NB	0	0.0	0.0	0.0	0.0				
		VERK	159	39.3	39.1	8.4	-2.6				
								1	0.491(0.396)	0.488	0.240
8		NG	0	0.0	0.0	0.0	0.0				
		1	5	1.2	1.2	5.2	-5.8				
	*.....	2	365	89.7	89.7	11.6	0.6				
		3	13	3.2	3.2	7.7	-3.3				
		4	24	5.9	5.9	5.4	-5.6				
		5	0	0.0	0.0	0.0	0.0				
		NB	0	0.0	0.0	0.0	0.0				
		VERK	42	10.3	10.3	6.1	-4.9				
								1	0.397(0.333)	0.304	0.121

SAS ITEM NR ITEM NR	KEY KORREKTE AFLEIER	DISTRACTER NR AFLEIER NR	FREQ AANTAL	PERS.VAN AANTAL GEDOEN	PERS.VAN TOTALE AANTAL	GEMIDDELD VIR TOETS	VERSKIL (GEMID - TOT.GEMID)	KRITERIUM NR	ITEM-KRITERIUM KOR.	ITEM STANDAARD-AFWYKING	PRODUK
9		NG	3	0.7	0.7	7.7	-3.4				
		1	74	18.3	18.2	8.3	-2.7				
	*.....	2	192	47.5	47.2	13.6	2.5				
		3	59	14.6	14.5	9.4	-1.6				
		4	79	19.6	19.4	8.8	-2.2				
		5	0	0.0	0.0	0.0	0.0				
		NB	0	0.0	0.0	0.0	0.0				
		VERK	215	53.2	52.8	8.8	-2.2				
								1	0.562(0.473)	0.499	0.280
10		NG	5	1.2	1.2	10.8	-0.2				
		1	50	12.4	12.3	7.3	-3.8				
		2	11	2.7	2.7	7.1	-4.0				
	*.....	3	329	81.8	80.5	11.9	0.9				
		4	12	3.0	2.9	7.1	-4.0				
		5	0	0.0	0.0	0.0	0.0				
		NB	0	0.0	0.0	0.0	0.0				
		VERK	78	19.4	19.2	7.4	-3.6				
								1	0.415(0.334)	0.394	0.163
11		NG	2	0.5	0.5	3.5	-7.5				
	*.....	1	123	30.4	30.2	14.3	3.2				
		2	45	11.1	11.1	10.4	-0.7				
		3	160	39.5	39.3	9.3	-1.8				
		4	77	19.0	18.9	10.1	-0.9				
		5	0	0.0	0.0	0.0	0.0				
		NB	0	0.0	0.0	0.0	0.0				
		VERK	284	70.1	69.8	9.6	-1.4				
								1	0.504(0.417)	0.459	0.232
12		NG	2	0.5	0.5	13.5	2.5				
		1	112	27.7	27.5	9.8	-1.2				
		2	67	16.5	16.5	9.8	-1.3				
	*.....	3	124	30.6	30.5	13.6	2.6				
		4	102	25.2	25.1	10.0	-1.0				
		5	0	0.0	0.0	0.0	0.0				
		NB	0	0.0	0.0	0.0	0.0				
		VERK	283	69.9	69.5	9.9	-1.1				
								1	0.405(0.308)	0.460	0.186

SAS ITEM NR ITEM NR	KEY KORREKTE AFLEIER	DISTRACTER NR AFLEIER NR	FREQ AANTAL	PERS.VAN AANTAL GEDOEN	PERS.VAN TOTALE AANTAL	GEMIDDELD VIR TOETS	VERSKIL (GEMID - TOT.GEMID)	KRITERIUM NR	ITEM-KRITERIUM KOR.	ITEM STANDAARD-AFWYKING	PRODUK
13		NG	6	1.5	1.5	11.3	0.3				
		1	43	10.7	10.6	7.4	-3.6				
		2	47	11.7	11.5	7.4	-3.6				
	*.....	3	265	67.1	66.1	12.7	1.7				
		4	42	10.5	10.3	8.1	-2.9				
		5	0	0.0	0.0	0.0	0.0				
		NB	0	0.0	0.0	0.0	0.0				
		VERK	138	34.4	33.9	7.8	-3.2	1	0.548(0.462)	0.473	0.259
14		NG	6	1.5	1.5	10.3	-0.7				
		1	104	26.0	25.6	10.8	-0.3				
	*.....	2	191	47.7	46.9	12.7	1.6				
		3	73	18.2	17.9	8.5	-2.5				
		4	32	8.0	7.9	8.0	-3.1				
		5	0	0.0	0.0	0.0	0.0				
		NB	1	0.2	0.2	11.0	-0.0				
		VERK	216	54.0	53.1	9.6	-1.5	1	0.367(0.258)	0.499	0.183
15		NG	6	1.5	1.5	11.0	-0.0				
	*.....	1	124	31.0	30.5	14.6	3.5				
		2	38	9.5	9.3	9.7	-1.3				
		3	58	14.5	14.3	8.8	-2.2				
		4	180	45.0	44.2	9.6	-1.5				
		5	0	0.0	0.0	0.0	0.0				
		NB	1	0.2	0.2	11.0	-0.0				
		VERK	283	70.7	69.5	9.5	-1.6	1	0.555(0.473)	0.460	0.255
16		NG	2	0.5	0.5	9.0	-2.0				
		1	35	8.7	8.6	8.5	-2.6				
		2	30	7.4	7.4	7.1	-4.0				
		3	52	12.9	12.8	7.5	-3.5				
	*.....	4	287	71.0	70.5	12.4	1.4				
		5	0	0.0	0.0	0.0	0.0				
		NB	1	0.2	0.2	11.0	-0.0				
		VERK	120	29.7	29.5	7.7	-3.3	1	0.506(0.419)	0.456	0.231

APPENDIX F_4

SAS ITEM NR ITEM NR	KEY KORREKTE AFLEIER	DISTRACTER NR AFLEIER NR	FREQ AANTAL	PERS.VAN AANTAL GEDOEN	PERS.VAN TOTALE AANTAL	GEMIDDELD VIR TOETS	VERSKIL (GEMID - TOT.GEMID)	KRITERIUM NR	ITEM-KRITERIUM KOR.	ITEM STANDAARD-AFWYKING	PRODUK
17		NG	3	0.7	0.7	6.7	-4.4				
	*.....	1	345	85.6	84.3	11.8	0.8				
		2	20	5.0	4.9	7.1	-3.9				
		3	16	4.0	3.9	6.6	-4.5				
		4	22	5.5	5.4	6.0	-5.1				
		5	0	0.0	0.0	0.0	0.0				
		NB	1	0.2	0.2	11.0	-0.0				
		VERK	62	15.4	15.2	6.6	-4.4				
								1	0.444(0.372)	0.359	0.159
18		NG	10	2.5	2.5	9.7	-1.3				
		1	27	6.8	6.6	8.3	-2.7				
		2	105	26.5	25.8	8.8	-2.2				
		3	96	24.2	23.6	9.8	-1.3				
	*.....	4	168	42.4	41.3	13.7	2.6				
		5	0	0.0	0.0	0.0	0.0				
		NB	1	0.3	0.2	11.0	-0.0				
		VERK	239	60.4	58.7	9.2	-1.9				
								1	0.524(0.431)	0.492	0.258
19		NG	3	0.7	0.7	7.7	-3.4				
		1	19	4.7	4.7	7.7	-3.3				
		2	82	20.3	20.1	7.4	-3.7				
	*.....	3	293	72.7	72.0	12.4	1.4				
		4	9	2.2	2.2	7.9	-3.2				
		5	0	0.0	0.0	0.0	0.0				
		NB	1	0.2	0.2	11.0	-0.0				
		VERK	114	28.3	28.0	7.5	-3.5				
								1	0.521(0.437)	0.449	0.234
20		NG	0	0.0	0.0	0.0	0.0				
	*.....	1	179	44.9	44.0	13.2	2.1				
		2	67	16.8	16.5	9.3	-1.7				
		3	88	22.1	21.6	10.0	-1.0				
		4	65	16.3	16.0	8.4	-2.7				
		5	0	0.0	0.0	0.0	0.0				
		NB	8	2.0	2.0	10.8	-0.3				
		VERK	228	57.1	56.0	9.4	-1.7				
								1	0.446(0.345)	0.496	0.222

APPENDIX G

Open ended questions - mean and standard deviation

OPEN ENDED QUESTIONS - MALES

Analysis Variable : X

N Obs	Minimum	Maximum	Range	Mean	Std Dev
101	1.0000000	54.0000000	53.0000000	27.5247525	15.0514744

OPEN ENDED QUESTIONS - FEMALES

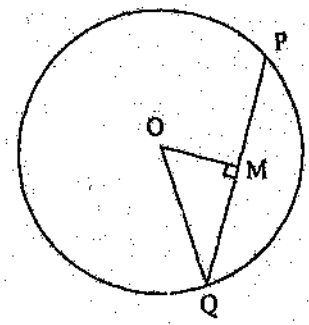
Analysis Variable : X

N Obs	Minimum	Maximum	Range	Mean	Std Dev
102	0.5000000	54.0000000	53.5000000	27.4509804	14.7281487

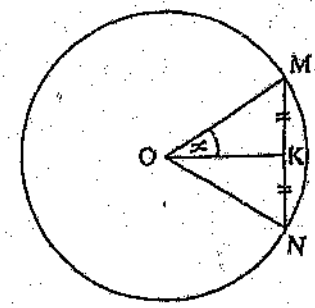
APPENDIX H_1

EUCLIDEAN GEOMETRY - TEST 1

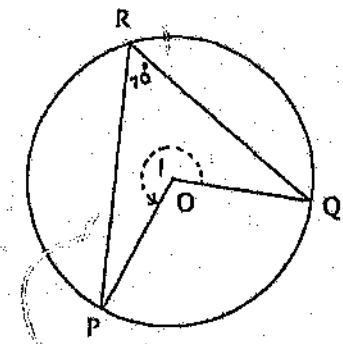
1. O is the centre of the circle.
   OM ⊥ PQ
   OM = 3 meters and OQ = 5 meters.
   What is the length of PQ in meters ?
   A. 4
   B. 5
   C. 6
   D. 8



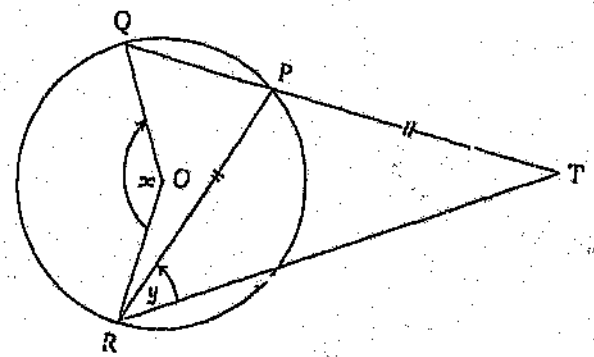
2. In circle O
   OM = MN and MK = KN.
   $K\hat{O}M = x$
   Calculate x.
   A. 30°
   B. 45°
   C. 60°
   D. Not enough information to calculate x.



3. O is the centre of the circle.
   P, Q and R are points on the circle.
   $\hat{R} = 70°$.
   Calculate the magnitude of $\hat{O}_1$.
   A. 280°
   B. 220°
   C. 210°
   D. 140°



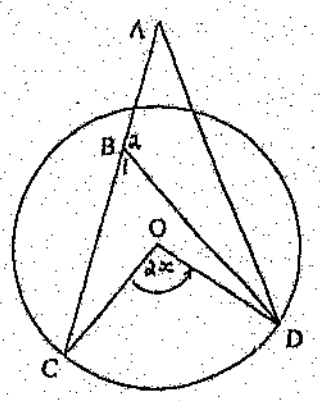
4. O is the centre of the circle.
   PT = PR
   Express x in terms of y.
   A. x = ¼y
   B. x = 2y
   C. x = 4y
   D. x = ½y



APPENDIX H_2

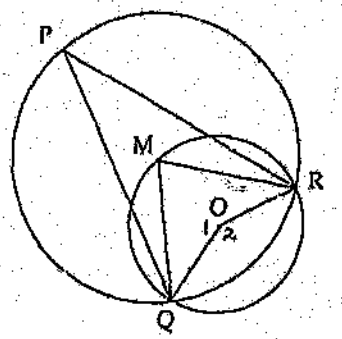
5. O is the centre of the circle.
$\hat{O}_1 = 2x$ (see fig).
B lies on AC, inside the circle.
Which statement is always true ?

A. $\hat{B}_1 = x$
B. $\hat{B}_2 = 2x$
C. $\hat{A} = x$
D. $\hat{A} < x$



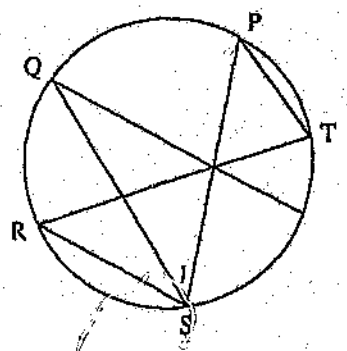
6. M and O are the centres of the circles and $\hat{O}_1 = 2x$.
Calculate $\hat{P}$ in terms of x.

A. ½x
B. 90° - x
C. 90° - ½x
D. x



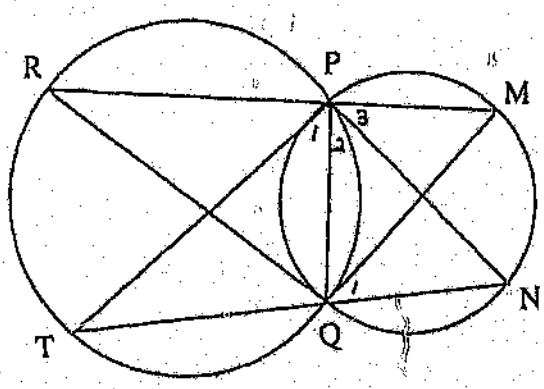
7. P, Q, R, S and T are points on the circle.
Which statement is always true ?

A. $\hat{P} = \hat{Q}$
B. $\hat{P} = \hat{R}$
C. $\hat{R} = \hat{Q}$
D. $\hat{T} = \hat{S}_1$



8. The circles do not have equal diameters.
Which statement is wrong ?

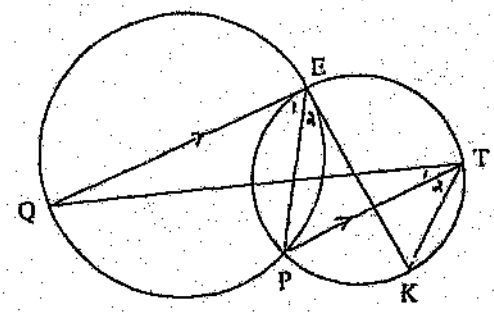
A. $\hat{R} = \hat{T}$
B. $\hat{P}_1 = \hat{P}_2$
C. $\hat{M} = \hat{N}$
D. $\hat{Q}_1 = \hat{P}_2$



APPENDIX H_3

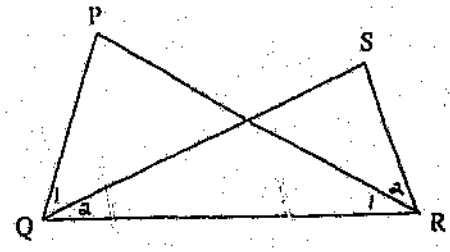
9. Two circles cut each other in E and P and QE // PT.
Which statement is always true ?

A. $\hat{E}_2 = \hat{T}_1$
B. $\hat{E}_1 = \hat{K}$
C. $\hat{E}_2 = \hat{Q}$
D. $\hat{E}_1 = \hat{E}_2$



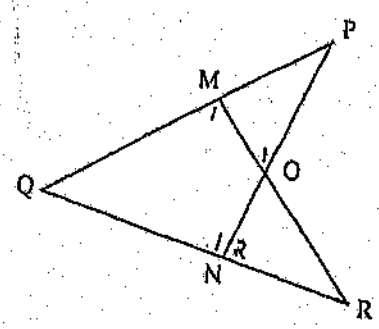
10. PQRS will be a cyclic quadrilateral if ...

A. $\hat{Q}_1 = \hat{S}$
B. $\hat{R}_1 = \hat{S}$
C. $\hat{Q}_1 = \hat{R}_2$
D. $\hat{Q}_2 = R_2$



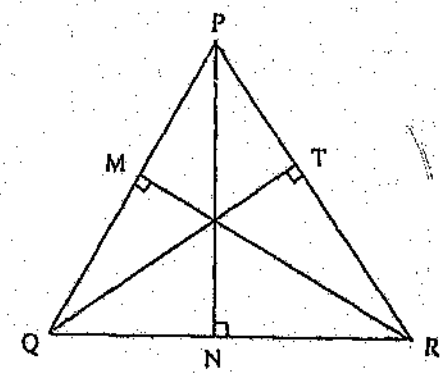
11. $\hat{M}_1 = \hat{N}_1$
Which statement is necessarily true ?

A. PMNR is a cyclic quadrilateral.
B. $\hat{O}_1 = \hat{Q}$
C. MQNO is a cyclic quadrilateral.
D. $\hat{N}_2 = \hat{M}_1$



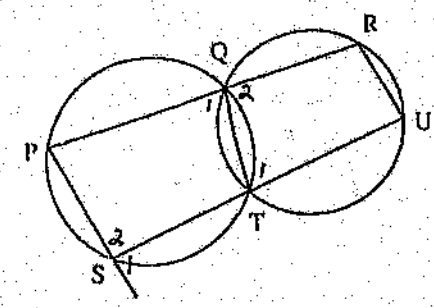
12. ΔPQR with perpendiculars as in diagram.
Which one of the following is a cyclic quadrilateral ?

A. MNRT
B. PMNT
C. PQNT
D. MQNT



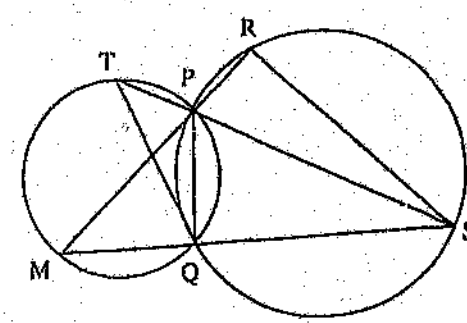
13. PR is not parallel to SU.
Which statement is wrong ?

A. $\hat{T}_1 = 180° - \hat{R}$
B. $\hat{Q}_2 = \hat{S}_2$
C. $\hat{S}_1 = 180° - \hat{Q}_1$
D. $\hat{U} = \hat{Q}_1$



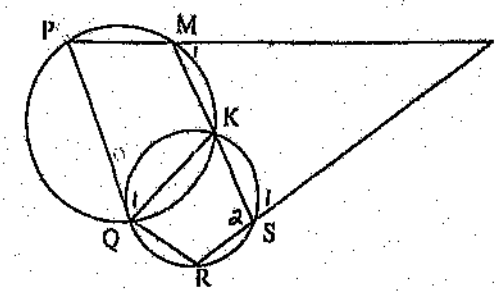
14. Two circles cut in P and Q.
MQS, MPR and TPS are straight lines.
$\hat{T} = 35°$ and $M\hat{P}Q = 57°$
What is the magnitude of $\hat{R}$?

A. 92°
B. 88°
C. 57°
D. 22°



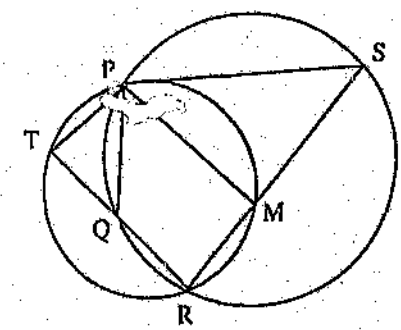
15. PMT, RST and MKS are straight lines.
Which statement is always true ?

A. $\hat{P} + \hat{R} = 180°$
B. $\hat{M}_1 = \hat{R}$
C. $\hat{P} + \hat{S}_2 = 180°$
D. $\hat{Q}_1 = \hat{S}_2$



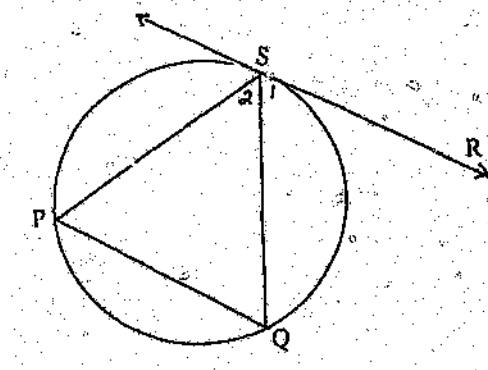
16. TQR and RMS are straight lines.
$\hat{T} = 75°$ and $M\hat{P}S = 40°$
Calculate the magnitude of $\hat{S}$.

A. 105°
B. 75°
C. 35°
D. 65°



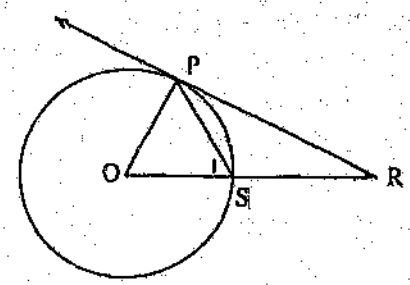
APPENDIX H_5

17. RS is a tangent to the circle.
$\hat{S}_1 = 54°$ en $\hat{Q} = 48°$
Calculate the magnitude of $\hat{S}_2$.



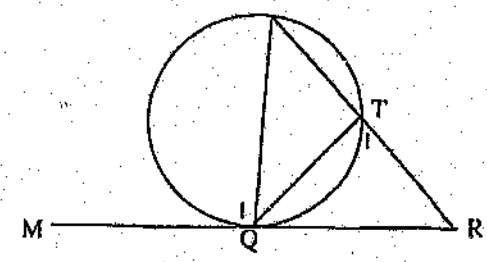
A. 78°
B. 72°
C. 63°
D. 48°

18. PR is a tangent to circle O and $\hat{R} = 40°$. Calculate the magnitude of $\hat{S}_1$.



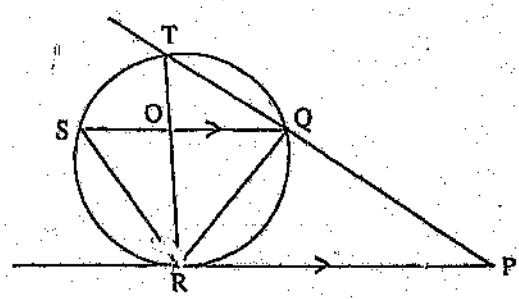
A. 25°
B. 50°
C. 60°
D. 65°

19. MQR touches the circle at Q.
$\hat{T}_1 = 70°$.
$\hat{Q}_1 = \ldots$



A. 20°
B. 70°
C. 110°
D. 140°

20. PR is a tangent to the circle and PQT is a secant. SQ and TR cut each other at O so that *SQ // RP*. Which statement is always true ?



A. RQ is a tangent to circle TOQ.
B. TR = RP
C. OQ is a tangent to circle RQP.
D. QR = QP

APPENDIX I_1

EUCLIDEAN GEOMETRY - TEST 2

QUESTION 1: [15]

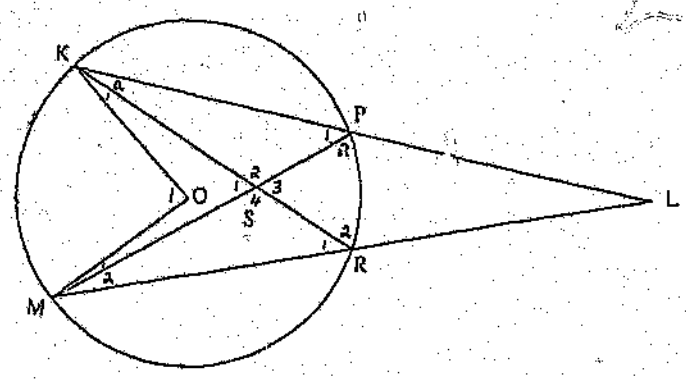
1.1 **COMPLETE:** The angle at the centre of a circle .... (1)

1.2 **GIVEN:**

Circle O. Chords KP and MR are produced to meet in L. (7)

$\hat{O}_1 = 80°$ and $\hat{S}_1 = 50°$

CALCULATE $\hat{L}$.

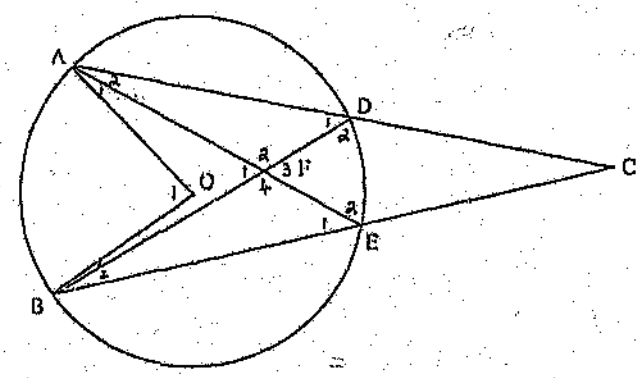


1.3 **GIVEN:**

Circle O. AD and BE are produced to meet in C. (7)

$\hat{O}_1 = 2x$ and $\hat{F}_1 = y$ and $\hat{C} = k$.

PROVE: $2x = y + k$



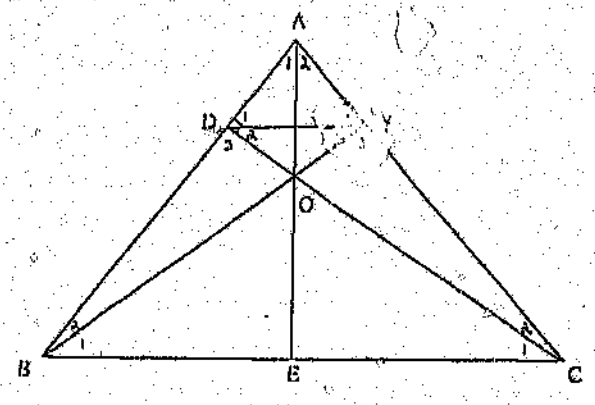
APPENDIX I_2

QUESTION 2: [14]

2.1 GIVEN: $\triangle ABC$ with $CD \perp AB$ and $BF \perp AC$. (7)

2.1.1 Name, with reasons, two cyclic quadrilaterals.

2.1.2 Prove $\hat{A}_1 = \hat{C}_1$

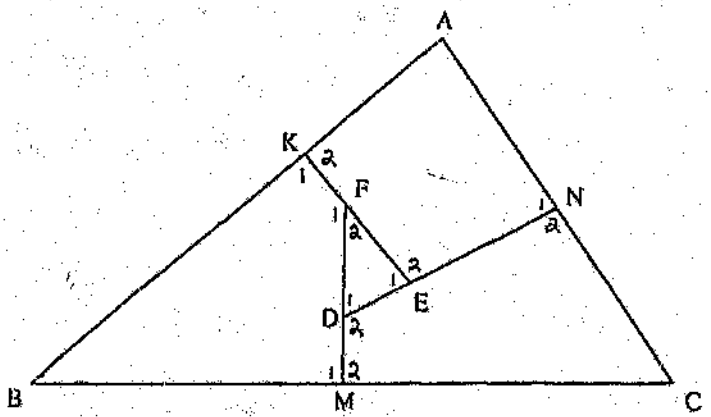


2.2 GIVEN: (7)

AKEN and BMFK are two cyclic quadrilaterals. All lines are straight lines.

2.2.1 Give, with reasons, one angle equal to $\hat{A}$.

2.2.2 Prove DMCN a cyclic quadrilateral.



APPENDIX I_3

QUESTION: 3 [25]

3.1 Prove that the angle between a tangent to a circle and a chord drawn from the point of contact is equal to an angle in the alternate segment. (6)

"Given" and "required to prove" may be omitted.

3.2 GIVEN: DPQ is a tangent to circle PSR. (9)

DSR is a tangent to circle RPQ.

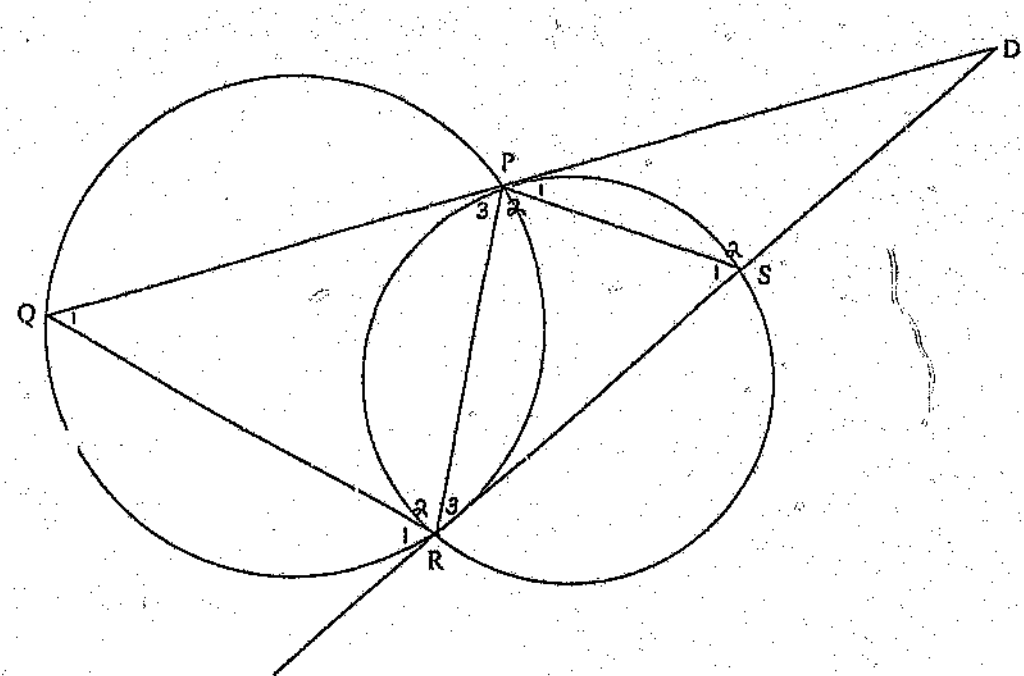
$\hat{R}_1 = 60°$ and $\hat{Q}_1 = 40°$

REQUIRED TO PROVE:

3.2.1 Show by calculation that $\hat{S}_1 = 60°$

3.2.2 Calculate $\hat{P}_1$

3.2.3 Prove SP // QR



APPENDIX I_4

3.3 **GIVEN:**

Two circles intersect in A and F. (10)

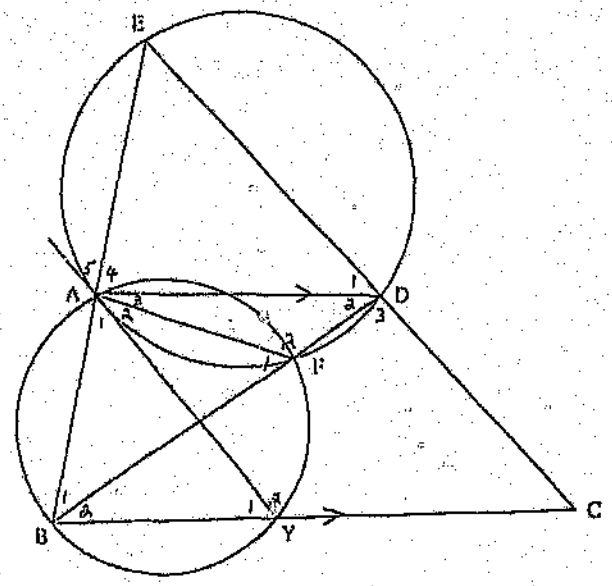
All other lines are straight lines.

AD // BC

PROVE:

3.3.1 that YA is a tangent to circle FAED.

3.3.2 if BE = BC, then DAYC is a parallelogram.



TOTAL: 54

Author: Cronje, L.S.
Name of thesis: Euclidean geometry- Cognitive gender differcees

PUBLISHER:
University of the Witwatersrand, Johannesburg
©2015

LEGALNOTICES:

Copyright Notice: All materials on the University of the Witwatersrand, Johannesburg Library website are protected by South African copyright law and may not be distributed, transmitted, displayed or otherwise published in any format, without the prior written permission of the copyright owner.

Disclaimer and Terms of Use: Provided that you maintain all copyright and other notices contained therein, you may download material (one machine readable copy and one print copy per page)for your personal and/or educational non-commercial use only.

The University of the Witwatersrand, Johannesburg, is not responsible for any errors or omissions and excludes any and all liability for any errors in or omissions from the information on the Library website.