STANDARDIZATION OF RAVEN'S STANDARD PROGRESSIVE MATRICES FOR SECONDARY SCHOOL AFRICAN PUPILS IN THE GRAHAMSTOWN REGION

THESIS

Submitted in partial fulfillment of the requirements for the Degree of

MASTER OF ARTS

of Rhodes University

by

Vasili A. Vass

January 1992

ABSTRACT

Arising out of a need, expressed by Clinical Psychologists in the Grahamstown region, for the fair assessment of Secondary School African pupils, norms for the Raven's Standard Progressive Matrices (SPM) were established. Two methods of presentation were used, the first using the original instructions of John Raven translated into the students' first language of Xhosa, hence-forth referred to as the Alternate sample. The second method of presenting the instructions was adapted, to include the use of visual aids and active participation in the instruction phase of the test, hence-forth referred to as the Normative sample.

In addition to the establishment of norms, the two methods of test presentation were investigated to see if the method of presentation had an effect on the results of the Raven's SPM.

The population was drawn from the three African Secondary Schools in the Grahamstown municipality. The total population consisted of 3 232 students. Classes were randomly sampled across the three schools, with the average age of the Normative sample being 19.3 years. The sample consisted of 812 pupils, 711 in the Normative sample and 101 in the Alternate Sample.

The following results and conclusions arose from the study:

1) Norms generated were considerably lower than previously established norms in similar studies.

2) t-Test results indicated that the method of test presentation on

the Raven's SPM is important when assessing students that may be regarded as 'disadvantaged'.

3) Respondents scored significantly higher when the method of presenting the instructions ensured a greater understanding of the task demanded of the respondents.

4) The analyses of covariance indicate that male subjects score significantly higher than female subjects, and that there is a significant difference between the ages and educational standard, on the scores of the Raven's SPM.

5) The differences found contradict previous findings using the test, and question the cultural fairness of the test.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGMENTS

I would like to thank the following people who were an integral part of the writing of the thesis:

1) Doctor A. Gilbert, my supervisor, for all his help and encouragement through out the duration of my Master of Arts degree. Your kind and valuable assistance helped me through the difficult times.

2) Mrs. Radloff for her assistance with the running of the statistical programs and interpreting of the results.

3) My family and friends for their patience and support.

4) The pupils, the most important part of this thesis. Without the support of the schools and pupils in the Grahamstown municipality, this would not have been possible.

CHAPTER 1 - INTRODUCTION

A need for a culturally fair test to assess the intellectual capacity of African pupils in the Grahamstown region has been expressed. Clinical psychologists in Grahamstown have stopped using the Raven's Standard Progressive Matrices (SPM) (Raven, 1958), reputably a culture fair test, to assess African pupils, as available norms were inadequate (Dr. R. Brooke, personal communication, September 1989). In this thesis the possibility of establishing norms for African pupils in the Grahamstown region on the Raven's SPM is investigated.

With the rapid political and social changes in the South African society, the need for such a test is of utmost importance. Should the political and social changes in the country lead to a unitary South African culture, with equal opportunities for all races, then the fair assessment of all peoples will be crucial. A need therefore exists to examine the misuse of intelligence tests, particularly when they are applied to cultures other than the ones than they were developed for.

Crawford-Nutt wrote that, "in the continuing debate on differences in intelligence test scores between blacks and whites, the importance of the method of test presentation seems largely to have been neglected in practice" (Crawford-Nutt, 1976, p. 201). The advice of Schwarz (1961, 1963) and Crawford-Nutt (1976), re: the use of visual aids and active demonstration in teaching test requirements to testees, was noted. This ensured that the pupils understood the task demands, thus using the test in a **fair** manner.

If the Raven's SPM is administered in a **fair** manner, the scores obtained by the sample in the study should be considerably higher than when the test is administered in an **unfair** manner. The adaptations to the instructions of the Raven's SPM help to make the administration of the test **fair**. Subsidiary to the establishment of the norms was the investigation of the effect of methods of presentation on performance. This taking into account, the observation of Crawford-Nutt (1976).

The scores obtained by pupils on the Raven's SPM are compared across gender, age, standard, and school, in order to identify the possible effect of these variables on test scores. Raven (1989), found that the mentioned variables, have no effect on the results of the test. However, research into the results on the Raven's SPM in non-western cultures, have shown differently. An investigation into the effect of these variables on the scores of the test will lead to a greater understanding of the cultural fairness of the Raven's SPM.

The Raven's SPM has been applied to African groupings in South Africa, and scores below the means of white scores have been obtained (Owen, 1989). Thus the scores obtained by the pupils on the Raven's SPM will be compared with results obtained from other studies. Again, the comparison of the scores obtained by the current sample with previous studies will lead to a greater understanding of the cultural fairness of the Raven's SPM.

The thesis has the following structure.

The first part of this study involves a broad outline of the nature of intelligence. In Chapter 2 the meaning of intelligence is discussed with specific reference to Spearman's two-factor theory of intelligence. Researchers have demonstrated that the Progressive Matrices tests are among the purest available measures of Spearman's **g** (Raven, 1989).

Issues relating to the assessment of intelligence, the problems of cross-cultural testing, and the culture fair - free debate are discussed in Chapter 3. An outline of the literature on the concept of intelligence and the assessment of intelligence, is followed by research into the Raven's SPM in Chapter 4.

In Chapter 5, the details of the study are given. This includes the instructions, and manner of presenting the Raven's SPM, and adaptations to the test. Full details of the research design are given in this chapter.

The results of this study are presented in Chapter 6. In Chapter 7 a discussion of the results is given, as well as a discussion of the results in relation to other research findings.

In Chapter 8 details of how to administer the Raven's SPM under the revised testing procedures, and norms established in this study will be presented.

Conclusions and recommendations for further research in the area of cross-cultural intelligence testing are made in the final chapter.

CHAPTER 2 - THE NATURE OF INTELLIGENCE

The word intelligence may be traced back to a distinction made by Aristotle between orexis, the emotional and moral functions, from dianoia, the cognitive and intellectual functions. Dianoia, was translated to intelligentia, by Cicero. These two components making up any animal or human behaviour, are sufficiently distinct to be studied separately (Vernon, 1979). Sattler (1982), states that the need for a practical study of mental ability became apparent because there was confusion about the difference between, as he termed them, "idiots" and "lunatics". Both groups were being treated equally, until Jean Esquirol (1838), cited in Sattler (1982), distinguished between mental incapacity and mental illness. Esquirol pointed out that idiots, may never have developed their intellectual capacities, whereas mentally ill people had possibly lost abilities that they once possessed.

Three ancient doctrines surrounding intelligence may be identified (Spearman, 1927). The first is labeled the "monarchic" doctrine. This doctrine postulated that mental ability fell under the control of one single element, and that element was identified to be intelligence. The second definition, stipulated several different abilities making up intelligence. The "oligarchic" doctrine identified four abilities, as opposed to the single element of the "monarchic" doctrine. The postulated four abilities were; "judgment, memory, invention, attention" (Spearman, 1927, p. 26). Spearman (1927) argued each of these separate abilities required their own measurement. The "anarchic" doctrine, broke intelligence down even further. All the various faculties in the different abilities required measurement

independently, with there being no restriction to the actual number of abilities that could be generated. This doctrine held a crude view that all abilities were independent (Spearman, 1927). This is a very crude breakdown of the three doctrines, but it is on this basis, that much of the debate around the definition of intelligence hinges.

The confusion surrounding intelligence is largely due to the fact that intelligence is a theoretical construct which cannot be directly observed. Ryle (1949), believes that it is in the origins of the word intelligence, that a fundamental error in the study of this concept has been made. For Ryle, intelligence is not a separate entity from behaviour. Intelligence may never be observed directly, it may only be inferred from behaviour which is classed as intelligent. For Ryle it is the manner in which a behaviour is carried out that makes it intelligent.

2.1 Defining Intelligence: The First Debate

Vroon (1980), identifies three types of definitions of intelligence. The first, comprises verbal, intuitive descriptions of a phenomenon. For example: intelligence is the ability to solve various types of problems.

The second comprises operational definitions. In this class of definitions, intelligence is what intelligence tests measure. This approach to defining intelligence is supported by Thorndike who said that, "If a thing exists, it exists in some amount," and "If it exists

in some amount, it can be measured" (Thorndike cited in Cronbach, 1964, p. 24).

This type of definition seems to have dominated the early part of this century. One of the earliest researchers to carry out intelligence testing was Binet. Binet broke away from the stream which aimed to provide a theoretical insight into intelligence, and provided a practical tool to measure intelligence.

Binet was a pioneer of intelligence testing in 1905, and his work influenced a vast amount of research carried out in this field. The basic skills assessed in his tests included areas of general mental development and judgment. The final scale developed by Binet in 1911, focused on measuring intelligence rather than academically related information (Rattan & Rattan, 1987). For Rattan et al. (1987), it was interesting that the diverse set of mental abilities that were included in this test were summarized in a single score. This seems to indicate support for Spearman's concept **g** (see section 2.1.1), even though Binet believed that intelligence was made up of a number of different abilities.

The final type of definition of intelligence includes: the assumption of a theory about the nature of the concept, the development of an instrument, and the possibility of using this instrument to predict behaviour and possibly change it. The Penguin Dictionary of Psychology adopts Vroon's third type of definition of intelligence. "Intelligence will be, conceptually, what it has always been, the ability to profit from experience and, pragmatically, what it has become, that what the

intelligence tests measure" (Reber, 1986, p. 365). This third type of definition identified by Vroon (1980), is a mixture of the first and second types of definitions. The assumption about the nature of intelligence encompasses the verbal, intuitive descriptions noted in the first definition. The development of an instrument to predict and possibly alter behaviour is synonymous with the view held by Thorndike cited in the second definition.

2.1.1 Spearman's two-factor theory.

Spearman (1927), believed that he had found the keystone to a greater understanding of the nature of intelligence. He observed that correlations between measurements of different abilities, such as scores for tests, marks for school subjects, or general impressions, tended toward a particular arrangement which could be expressed in a definite mathematical equation. This equation he called the "tetrad equation". Thus in a random sample of a population, an individual's scores on mental tests were positively correlated, and if an individual scored well on one test, s/he would perform as well on a similar test (Rattan et al., 1987). Whenever the tetrad equation is shown to be true throughout any table of correlations, then Spearman (1927) argued each measurement of an ability may be divided into two independent parts. The first, and the most important part for Spearman was, what he labeled the "general factor and denoted by the letter **g**" (Spearman, 1927, p. 75). Today this **g** is commonly referred to as "general intelligence". The second, he referred to as the "specific factor and denoted by the letter **s**" (Spearman, 1927, p. 75).

Although these two factors are present in all abilities, they need not be equally influential in all. Spearman (1927), found that, the **g** had a much greater relative influence in some abilities than others. The ratio developed to assess the influence of **g** relative to that of **s**, was said to be as much as fifteen to one, on some abilities and as low as four to one on others. Spearman warned, though, that his **g** was not a concrete thing but only a value or magnitude.

> Further, that which this magnitude measures has not been defined by declaring what it is like, but only pointing out where it can be found. It consists in just that constituent - whatever it may be - which is common to all the abilities inter-connected by the tetrad equation. Eventually, we may or may not find reason to conclude that **g** measures something that can be appropriately called "intelligence". Such a conclusion, however, would still never be the definition of **g** but only a "statement about" it. (Spearman, 1927, p. 75, 76).

It was only in 1950 that Spearman wrote more on the specific factors **s**, as well as developed the "broad" factors. The broad factors, were said to be relative to the composition of the particular test being used (Spearman & Jones, 1950). These broad factors, Spearman (Spearman & Jones, 1950) said might cover a broad range of abilities. These abilities should not be seen as alternatives to the specific factors but rather as supplementing the specific factors.

Through his correlational studies, Spearman identified verbal, logical, mechanical, the psychological, and arithmetical abilities, as having correlations of appreciable magnitude. Thus Spearman and Jones (1950), conclude that there is an element of truth in all three ancient views of intelligence;

> For the "monarchic doctrine of intelligence" could claim to be, at any rate, the forerunner of the modern **g**. Again the "anarchic doctrine of independent elements" forestalled in some degree the present conception of the multitudinous narrow **s**'s. And as much as was said, finally, even about the faculties of the "oligarchic view". For something like these - albeit shrunken idea - was taken "to have revealed itself in what we have been calling broad factors". These were described as "quite distinct from the universal factor and fairly distinct from the ordinary narrow factors". (Spearman and Jones, 1950, p. 13 - 14).

Spearman (1927), identified 3 quantitative laws, which he labeled the neogenetic laws, to show how all cognition that is not merely reproductive, which may be labeled **g**, is possible. The first law he labels "the apprehension of one's own experience". This law is based on the idea that a person has the ability to observe what is going on in his/her own mind. "He not only feels, but also knows what he feels; he not only strives, but knows that he strives; he not only knows, but knows that he knows" (Spearman, 1927, p. 164). The proof of this law lies in statements made by people, "I was angry at this," or "I tried to do that," or "I saw something red," or "I thought of the future" (Spearman, 1927, p. 164). Seeing that this power was not being measured by any intelligence test, Spearman (1927), noted a gap in what intelligence tests were measuring and the nature of **g**.

Spearman labeled the second law "the eduction of relations". This law enounces "that when a person has in mind any two or more ideas (using this word to embrace any items of mental content, whether perceived or thought of), he has more or less power to bring to mind any relations that essentially hold between them" (Spearman, 1927, p. 165). Once

again Spearman saw this law as being indisputable as it is evident whenever, a person becomes aware, for example, "that beer tastes something like weak quinine" (Spearman, 1927, p. 165).

"The eduction of correlates", is the third law. This law stated "that when a person has in mind any idea together with a relation, he has more or less the power to bring up into mind the correlative idea" (Spearman, 1927, p. 166). To explain this law, Spearman provides the example of an individual trying to imagine a musical note that is a fifth higher than the one he/she has just heard. To complete this task successfully, the subject would have to understand the relation of a "fifth", and have an ear for music.

Spearman (1927), however warned that **g** alone is not a sufficient measure of intelligence. This **g**, measures only cognitive aspects of mental activity, and ignores any affective influences, and does not take into account the conative aspects at all.

Radford and Burton (1974), noted that there has been a persistence, in intelligence tests for Spearman-type items. The Raven's Progressive Matrices which is of central concern in this research, is one such test. The test will be discussed in detail in Chapter 4. The two-factor theory is particularly interesting for Radford et al. (1974), in that Spearman attempts to deal with intelligence both by a theoretical analysis and by the experimental isolation of factors. This fits in with Vroon's (1980) third classification of a definition of intelligence, as the testing process is of great importance to Spearman's development of his theory.

Although Spearman, has been accredited as being the father of the use of factor analysis in psychology (Guilford & Hoepfner, 1971), it was other theorists, like Guilford and Thurstone, who as the technique developed, proposed multiple factor theories of intelligence. "Factor analysis is in essence a series of correlations, between every pair of a set of variables. A group of variables that correlate highly with each other, but not with others, suggests the existence of some 'factor' underlying them" (Radford et al., 1974, p, 131). This technique is mathematical in nature, and reduces a complex system of correlations into fewer dimensions (Gould, 1984). This method of defining intelligence provides us with a quantitative analysis and does not base the definition on speculation. With this in mind, the weaknesses of this method must also be mentioned. The analysis only comes from that what is thought to be important and measured, and the analysis does not yield psychological but only statistical factors.

2.1.2 Multiple-factor theories.

Thurstone's work in 1931, on graduate students at the University of Chicago, led him to the conclusion that an individual's cognitive achievements constitute a number of specialized primary abilities. It was only later that he was able to identify these.

Thurstone (1938), who took advantage of the new developments in factor analysis, found seven primary mental abilities. He labeled these:

S - spatial ability

P - perceptual speed

N - numerical ability

M - memory

V - verbal meaning

W - verbal fluency

I or R - inductive reasoning

Thurstone's findings were thus in agreement with the position taken by Binet that intelligence is made up of a number of factors (Rattan et al., 1987). Radford et al. (1974), noted that, at a later stage, Thurstone pointed out how one can obtain an estimate of general intelligence from the scores on the tests designed to assess the primary mental abilities. Like Binet, Thurstone summarized his primary mental abilities in one score, thus in a sense, justifying Spearman's general factor of intelligence.

Guilford et al., (1971), in developing "The Structure of Intellect Model felt that it was not satisfactory to merely list a number of intellectual abilities, that could be demonstrated through factor analysis. The **Structure of Intellect** model provides the reader with a systematic order of the various abilities. This model is classified as a "morphological" model; meaning that abilities can be classified in three different ways, with the categories of one intersecting with those of the others.

Guilford's three-way classification involves:

a) Classification in terms of the kind of mental **operation** involved in the abilities.

> Each ability involves simply cognition (knowing), memory (or learning 'that sticks'), divergent production (generation of logical alternatives), convergent production (generation of logic-tight conclusions), and evaluation (judging goodness of what is known or produced) (Guilford et al., 1971, p. 18).

These five operation categories each contain 24 similar abilities.

b) Classification in terms of **content**. These are areas of information within which the operations are performed. These include:

> Figural (concrete, perceived), symbolic (signs, code elements such as numbers or letters), semantic (thoughts,conceptions, or constructs), and behavioural (psychological) (Guilford et al, 1971, p. 18).

These four content categories each contain 30 similar abilities.

c) Classification in terms of the **product**, describe the formal kinds of information.

> Information takes the form of units (segregated chunks), classes (common properties within sets), relations (changes, transitions), and implications (information suggested by other information) (Guilford et al., 1971, p. 18).

Guilford notes that each product category contains a set of 20 similar abilities.

Mental abilities needed to complete a task involve an interaction of the above three categories. Using this model, Guilford et al. (1971), purported that the 120 unique possible abilities, should not be

supposed to cover the whole range of intellectual traits or variables.

Looking at the number of possibilities of interaction, it is clear that intellect is made up of a number of separate factors. Guilford warns however,

> It must not be supposed that, although the abilities are separate and distinct logically and they can be segregated by factor analysis, they function in isolation in the mental activities of the individual (Guilford et al., 1971, p. 19).

It is the fact that two or more of the abilities may usually be involved together, in solving the same problem that has made it difficult to recognize them as separate abilities.

2.1.3 Summation.

The above two-factor theory and multiple factor theories of the nature of intelligence, while appearing to be totally contradictory to each other, actually do have common ground. All the theories recognize that various abilities go to make up the concept of intelligence. To illustrate this point, Vernon (1979) may be cited. He notes that factor analysis does not yield any definitive solution to the problem of uni- or multi-dimensionality of intelligence. He continues, and proposes that a uni-dimensional or hierarchical model of intelligence, at times, may be more appropriate with some populations, while with others a multi-dimensional model of intelligence would be more beneficial. To illustrate that there is no contradiction between these two theories, the example of school pupils, being classified in terms of their grades in each of the main subjects, or in terms of an

average or all-round achievement is cited. The different school subjects may be compared to Thurstone's primary factors, with these particular achievements being able to be broken down into more specialized skills. A possible reason for the existence of the clusterings of factors, according to Vernon (1979), may be largely due to cultural uniformities, which are reinforced by the appropriate schooling.

2.2 Nature versus Nurture: The second Debate

2.2.1 A combination of nature and nurture.

Radford et al., define intelligence thus,

> It appears that the word "intelligence" has **two** valuable meanings. One is (A) an **innate potential**, the capacity for development, a fully innate property that amounts to the possession of a brain and a good neural metabolism. The second is (B) the functioning of a brain in which development has gone on, determining an **average level of performance or comprehension** by the partly grown or mature person. Neither, of course, is observed directly; but **intelligence B** a hypothetical level of development in brain function, is a much more direct inference from behaviour than **intelligence A**, the original potential (Radford et al., 1974, p. 138).

The above definition of intelligence is crucial to an understanding of the roles that genetics and the environment play in what constitutes intelligence. Today it is widely agreed by theorists of intelligence that there is an important genetic inheritance, which influences what an individual will be able to attain. This genetic influence, however, is very difficult to explain. Intelligence should not be viewed as a

fixed capacity, but rather a capacity capable of changing and developing with time. Thus the intelligence of an individual is the result of an interaction between the genetic pre-disposition of the individual and the environment in which the individual is reared.

Vernon (1965), breaks down the concept of intelligence into three components. "Intelligence A", is defined as the the genetic potential of an individual. Thus, it is this genetic potential which sets the roof limit, for the intelligence an individual may attain. "Intelligence B", is the all-round cognitive abilities used to comprehend, grasp relations and reason, which develop through the interaction between the genetic potential and stimulation from the environment. The results obtained from intelligence tests he defined as "Intelligence C". Thus, Intelligence C may be viewed as a sample of Intelligence B and Intelligence A, B, and C, can be considered as being open to heredity factors. In addition Intelligence B and C are also open to environmental influences (Vernon, 1965). This view is supported by Irvine (1983).

Vernon (1965), identified eight factors which may handicap intellectual development and functioning: physical and nutritional deficit, perceptual deprivation in pre-school years, repression of independence and constructive play, family insecurity and lack of playfulness, female dominance favouring verbal as compared to spatial abilities, defective education, linguistic handicaps and adult roles and adolescent aspirations in that children may be affected by the realization of their depressed status and lack of opportunities. The increasing number of articles regarding the role of environmental and

cross-cultural factors in the definition and assessment of intelligence, is evidence of the importance of these factors (e.g., Irvine & Berry, 1983; Triandis & Berry, 1980).

2.2.2 Racial differences.

> Today a great drama is being played out: the social sciences are attempting to clarify an issue of real philosophical and political importance, namely, whether the races differ in terms of certain intellectual skills because of genetic differences (Flynn, 1980, p. 213).

A. Jensen (1969), in his most controversial article, "How much can we boost IQ and scholastic achievement?", makes the following points. On average, Negroes test one standard deviation, which is 15 IQ points lower, than the average white population on IQ, as well as one standard deviation below whites on scholastic achievement. However, his most important assertion is that intelligence has a large genetic component. This genetic component determines about 80 per cent of an individual's intelligence in all human groups. From this he makes the assumption that it is this genetic component of intelligence, that is the cause of the average Negro-white intelligence difference. Eysenk (1971), also supports Jensen's view.

Jensen (1969) provides this information in a very convincing manner, but has been shouted down by many theorists for his dogmatic and racist comments. Theorists have shown many flaws in his argument. These will be briefly stated here as set out by Tobias (1974): (a) Jensen uses IQ as if it were a behavioural entity without noting, that

in fact, it is only a statistical index; (b) Jensen did not control for many subtle environmental influences. Influences such as parental solicitude and warmth of interpersonal relations, were ignored by Jensen. In addition to this Biesheuvel (1972), notes that Jensen only controlled for environmental factors at the time of the study but did not take into account that the children he studied, may have come from homes where there had been a gain in social status, thus negating the effects of having lived in a home from a low socio-economic status; (c) Jensen also overlooked the attitude of test-takers towards the testers; (d) The fairness of testing children of one group with IQ items devised by members of another group was also ignored by Jensen.

It appears that Jensen failed to fully understand the nature of the testing situation and its effect on the measurement of intelligence. The Society for the Psychological Study of Social Issues criticized Jensen's Article in a press release which stated,

> There is no direct evidence that supports the view that there is an innate difference between members of different racial groups (Tobias, 1974, p. 8).

The press release added that,

> ... a more accurate understanding of the contribution of heredity to intelligence will be possible only when social conditions for all races are equal and when this situation has existed for several generations (Tobias, 1974, p. 8).

These errors, and others not mentioned above, will be discussed in Chapter 3, where the intricacies of the test situation will be dealt with in more detail. The reasons for differences between the races on intelligence tests should be viewed in the light of Vernon's (1965),

breakdown of intelligence into three components. This definition breaks away from defining intelligence as constituting a particular ability or constituting a number of abilities. To determine the influence of genetics on intelligence is an impossible task at the moment, and the advice given in Tobias (1974), by The Society for the Psychological Study of Social Issues should be taken, while still recognizing that intelligence is the result of an interplay between a genetic component and the environment. A great deal of research on the influence of the environment on intelligence has been carried out, and more still needs to be done so as to provide users of intelligence tests the knowledge which will allow for the **fair use** of intelligence tests. This should lead to a greater understanding of the intricacies of testing in the cross-cultural context, and a better understanding of the reasons for differences between different cultures.

Flynn (1987), summarizes that the differences between groups on IQ tests may not be equivalent to intelligence differences. He comes to this conclusion when looking at differences in IQ scores between generations, but the reasons for this difference may be extended to the cross-cultural situation. The major factor identified by Flynn (1987) is that of the environment. Verster (1987) found that enivronmental factors which have a small, inconsistent effect in intelligence scores in America, account for large and consistent effects in Africa. These variables included ethnicity, family size, and coaching.

2.3 Conclusion

There has been much debate surrounding the nature of intelligence and the reasons for the differences in intelligence test scores. A complete grasp of the concept of intelligence may still be out of our reach but it is these debates that bring us closer to an understanding of the concept.

The path taken by Vernon (1979), who links the uni- or multi-dimensional intelligence with that of the nature-nurture debate, may lead to a greater understanding of the concept and the influences that the environment has on intelligence. Thus, research on the applicability of intelligence tests, and under what conditions these tests may be applicable in cultures other than the one in which it was developed, is crucial. What should be investigated is the adaptation or development of instruments or tests that will measure individual performances, according to strict psychometric rules, yet also be fair to the individual being tested. A similar view was put forward by Irvine (1963, 1983). Sternberg (1984) also placed more emphasis on cross-cultural factors when defining intelligence. Factors making the administration of psychological tests across cultures more applicable, will be investigated in Chapter 3. More prominence has been given to the role of cross-cultural factors in the defining and measurement of intelligence.

CHAPTER 3 - ISSUES OF ASSESSING INTELLIGENCE ACROSS CULTURES

It has become necessary to apply tests of intelligence to societies other than the one in which the test was developed. In developing nations, where there is a move towards industrialization, psychometric tests are playing an increasing role in the utilization of human resources. Intelligence tests are a crucial set of these psychometric tests, although the use of them, is controversial in the cross-cultural or multi-cultural setting. In these situations, is it psychometrically valid to apply tests developed in one culture to other cultures or different cultures?

In the multi-cultural society of South Africa, history has not allowed for the development of a true understanding of the similarities and differences between the cultures. This has had a detrimental effect on psychometric testing in South Africa. Biesheuvel (1954), notes that the study of individual differences between the whites, was focused on high-grade technical, administrative and professional ability. Restrictions on the African people, prevented them from entering into skilled trade, thus the assessment of African people during the 1950's concentrated on unskilled tasks. Employers at the time wanted to assess the abilities of work-seekers according to the above discriminating line. To meet this need, in the construction of tests, no attempt was made to appraise the abilities of Africans as they would have been, had they enjoyed the same environmental opportunities as the average white South African (Biesheuvel, 1954). The aim of these early tests were to indicate how readily and up to what point a work-seeker was trainable on the job. This was first identified by

Biesheuvel (1943) in his book, African Intelligence.

This scenario has hindered the development of psychometric tests which could accurately assess the abilities or potentials of individuals across cultures. The above tests only allowed for the African people to be assessed according to unskilled or manual labour, thereby not recognizing any skilled ability. The mark of this is still with us today, with a great discrepancy between black and white on test familiarity and test content, which plays an important role in the **fair** administration of tests.

Changes in the job market, and the educational sphere, has led to many researchers asking the question, whether it is better to use common or separate psychometric tests for the various cultural groupings in South Africa (Owen, 1989). The construction and use of common psychometric tests in the industrial and educational sector in South Africa seems inevitable with the political and social change that is taking place. These changes include: the abolishment of all forms of job reservation in 1985, the opening of schools to all races, subject to 70% of the white parents agreeing to the opening of the school which their child attends; the repeal of the Population Registration Act, the Separate Amenities Act and the Group Areas Act.

Researchers have to ensure that the social demands that will follow political change, can be met by psychologists and psychometric testing. Psychometric tests will have to be constructed that are able to comply with strict psychometric requirements, as well as meet the

demands of a multicultural society.

Orter (1972) identifies five areas in which care needs to be taken, when using tests. These areas are extremely important for cross-cultural testing and testing in a multicultural society.

(a) The attributes to be measured have to be investigated in the two contrasting cultures, before it can be labeled a fair measure of the trait under consideration. That is, the issue of conceptual equivalence needs to be investigated.

(b) The intention to measure a particular behaviour, which is similar across two cultures does not automatically call for the use of a similar test. If one sets out to measure intelligence in a foreign culture it may not be necessary to use a standard intelligence test. This is a mistake that many psychologists make.

(c) The language used is crucial. The **meaning** of an item is not solely dependent on the the content but also the subject's perception thereof. With language differences existing interculturally, adequate **mediation** of a given problem may be as important as the problem content. The medium to be used in the testing situation should, be the medium most common and most used for transmitting a problem in the particular society.

(d) Previous experience of testing is important. Choosing instructions appropriate to a subject's milieu will allow the subject to come to a better understanding of the principles underlying the test, and thereby creating a situation where the test items do not have to be altered.

(e) Norms should be based on a representative sample of the population being tested. There is no justification for applying

the original norms derived from the population belonging to the source culture.

3.1 Approaches to Cross-Cultural Testing

Anastasi (1988), identified three approaches to the development of tests for persons from cultures different from those in which the test is developed. The first approach involves the choice of items common to many different cultures and the validation of the resulting test against local criteria in different cultures. This approach leads to the development of culture-fair tests. The failure of this method comes about when the test is not validated or inadequately executed in the different cultures in which it is being used in. The second approach, is to develop a test in a culture, and then to apply the test to various cultural groups. Resulting differences in scores should not, however, be seen as having derived from the same causes. These tests should not be regarded as universal yardsticks for the measurement of intelligence, but rather as a tool for understanding the nature of differences between the cultures. The third approach Anastasi mentions, falls in line with the radical cultural relativist approach proposed by Berry (1972). This approach suggests that different tests should be developed within different cultures, which are validated and used within the particular culture in which it was developed. The problem with this approach is both a practical and a psychometric one. Practically, in a multicultural society like South Africa, a number of tests would have to be developed to ensure that the different cultures in South Africa have tests developed by people

in their culture, and validated in their culture. Psychometrically, the development of different tests does not ensure that what is being measured by the tests is similar.

At the start of cross-cultural psychological testing it was believed that the problem of cultural effects could be overcome through the development of a test that would eliminate those extraneous factors effecting test performance. This led to the development of many so called **"culture-free"** tests. This was based on the presumption that, as Anastasi calls it, "heredity intellectual potential" (Anastasi, 1988, p. 357), could be measured theoretically independent of the impact of cultural experiences. Psychologists today, recognize that heredity and environmental factors operate jointly, and the two are inextricably intertwined in the behaviour of the individual, thus making the measurement of the amount of heredity influence on intellectual behaviour impossible. Anastasi (1988) says, since all behaviour is affected by the cultural milieu in which the individual is reared, and since psychological tests are but samples of behaviour, cultural influences will and should be reflected in test performance. It is thus pointless to try and develop a test that is **free** from cultural influences.

"Culture-fair" tests attempt to rule out one or more of the parameters along which cultures vary. For Anastasi (1988), cultural differences exist on a continuum, from superficial and temporary effects to those that are basic, permanent, and far reaching. Many cultural factors that effect test responses also influence the broader behaviour domain that the test is designed to sample. Some of these cited by Anastasi

(1988), include language, reading disabilities, ineffective strategies to solve abstract problems, lack of interest in intellectual activities, hostility towards authority figures, low achievement drive and poor self concept. All the above affects are not heredity, and may be negotiated through various means, and have an effect on the individual's educational, occupational and social activities. These environmental conditions, however, do become more difficult to alleviate, the longer they have been operating in the individual's life. Thus these tests, developed for use in the cross-cultural setting presuppose only experiences that are **common** to the different cultures (Anastasi, 1988), that is, cultural similarities are what the test is based on. "Culture-fair" test items are drawn from a pool of behaviour that is common to the cultures for which the test is developed.

This encompasses the **culture-free versus culture-fair test** debate. Anastasi (1988) adds that it is unlikely, that any test can be equally "fair" to more than one cultural group, especially if the cultures are quite dissimilar. She finds the concept of **cultural disadvantage** to be crucial. This is a relative concept. Cultures reinforce the development of behaviour that is adapted to its values and demands. In South Africa, the African population is now faced with new values and demands, as a result of the changing social and political climate. Thus when an individual is forced to adjust and compete within a culture or subculture other than the one that in which s/he is accustomed to, then cultural difference may become cultural disadvantage. How should psychometric testing deal with the problem of

cultural disadvantage?

Rimland (1972), suggests that we should not look for "culture-fair" or "culture-free" tests because most research has shown that minority group members or culturally disadvantaged members, do not score any higher relative to the general population, than they do on more typical intelligence tests. Feuerstein (1972) echoes this approach. "Individuals belonging to disadvantaged subgroups almost always average lower on tests, even when the tests are culture-free, culture-fair, or developmental" (Feuerstein, 1972, p. 265). This seems to leave the cross-cultural psychologist with a dilemma. How does s/he make comparisons, if the instruments for this comparison do not exist? Schwarz (1961, 1963) and many other researchers who have researched psychometric testing across cultures, suggest the establishment of a baseline, from which comparisons of the elusive concept of intelligence can be made. The only way that this may be achieved is through trial and error, by administering tests in a successful manner.

Wesley and Karr (1966), say that it is necessary to develop normative data for a culture before cross-cultural comparisons can be made. The instruments used, to measure psychological concepts like intelligence, are likely to detect and evaluate fine variations within and between cultures. They show that standardized results obtained, on whatever type of test being used, be it an attitudinal scale or an intelligence test, have no concurrent or predictive validity when applied to another culture.

What needs to be added to the above, is that the task set to measure intelligence, needs to be backed by adequate environmental supports. In the South African situation the environmental supports for the various abilities to be measured are, by far, nowhere near similar. There are huge differences in the social, educational, and economic backgrounds of the different cultures in South Africa. With the current social and political changes taking place in the country, attempting to narrow the gap between the different cultures, psychologists need to investigate conditions under which the **fair administration** of the tests may take place. Irvine (1969, 1983) shows that the lack of environmental correlations with tests is a landmark of African testing. This lack of correlation is indicative of disadvantage. Thus making the administration of tests, were there is this lack of correlation, crucial.

Thus in South Africa at the moment it would not be useful to compare the different cultural groups. Rather, available tests that are considered to be culture-fair or culture-free, should be administered in a manner that compensates for the groups lack of familiarity of the test content, to establish norms for use within that culture. These norms should be used for the assessment of individuals from that particular culture, and not for comparison across cultures. Comparisons across cultures should only be made between groups of peoples or individuals that may be considered to have equal opportunities in achieving the maximum possible score. Problems faced by comparing different cultures recede and make the issues of predictive validity, the standardization of the testing conditions

within the cultural setting and the reliability of the measure more important.

A baseline or commonalty, along which the **fair** assessment of the different cultures in South Africa may be made, should be achieved, before any comparison between groups should be made. That is, each culture should have its own norms against which it may be assessed, even though similar tests may be used across different cultures. This may be achieved through the investigation of the bias' that exist within the testing conditions and the test instructions. Using this approach to cross-cultural testing in a multi-cultural society like South Africa, a move to a more comprehensive understanding of the different cultures, what constitutes intelligence, test familiarity and test content and, what constitutes successful test administration within each culture, will be attained.

3.2 Successful Test Administration

Anastasi (1988), identifies three areas that are crucial to the successful administration of psychological tests. The first area is the **advanced preparation of the examiner.** The tester should be familiar with the test and the testing procedure. Helpers should be available to ensure that the test material is handed out and collected, with the tester in charge of the test situation. The second and third areas deals with the **introduction of the test** and the **test content.**

3.2.1 Introducing the test: Rapport and the test taker.

It is crucial that the examiner should be able to "arouse the test takers' interest in the test, elicit their cooperation, and encourage them to respond in a manner appropriate to the objectives of the test" (Anastasi, 1988, p. 35). This task becomes extremely difficult when testing children from educationally disadvantaged backgrounds or from different cultures, as the examiner cannot make the assumption that they will be motivated to excel on academic tasks to the same extent as children in the original standardization sample (Anastasi, 1988). In addition to this, testing can be perceived as a threat to the individual's esteem. For these reasons, reassurance is crucial before the start of the test. The elimination of surprise is also important as this will reduce the anxiety of the test taker. Wober (1969), quotes Vernon, "we must infer that performance... of... Africans is more susceptible to motivational or attitude effects than is usually found in American or British researcher's and to some extent this invalidates the rest of the research... " (Vernon cited in Wober, 1969, p. 234).

Sacks (1952), identifies emotional tension, motivation, and attitude of both the examiner and examinee as being crucial factors in establishing rapport and which can influence an intelligence test result. Sacks (1952) found that a good relationship between examiner and examinee will produce a mean increase in IQ points which is significantly greater than the mean produced by examinee's who have a poor relationship with the examiner. The relationship between the

examiner and examinee may be seen as an extension of relationships that exist within the wider community. This finding of Sack's must have great implications for psychometric testing in South Africa. The relationship between a white examiner and a black examinee, may be an extension of the racial relationships in an apartheid society.

This area of testing has been widely researched by Sacks (1952), Baratz (1967), and Katz, Henchy, and Allen (1967). Katz et al. (1967), found that urban Negro boys of grade school age performed better on a verbal learning task with Negro examiners than with white examiners, and when given approval rather than disapproval. He assumes that the low rapport between black examinees and white examiners is due to,

> (a) that the white adult was perceived as hostile because of previous **negative experiences** with white teachers and/or adults, or (b), that there was a feeling of strangeness with the white adult due to a previous **lack of experience** with white teachers (Katz et al., 1968, p. 42).

3.2.2 Fair test conditions in relation to South Africa.

Before comparisons may be made across cultures, the testing environment must be standardized. Variables able to effect the nature of the test condition need to be controlled for. Such factors range from the physical conditions such as the nature of working surfaces, for example desks upon which testees write through to the tester's manner and behaviour. As a basic principle a relaxing and comfortable environment should be provided for all groups. This however, is not always possible when testing in schools in South Africa. Testing white pupils in white government schools provides a

totally different **test environment** than testing black pupils in black schools. Generally conditions in black government schools are poor, with overcrowding of classes being a major problem and the lack of facilities being a reality. Thus when using a classroom as the test room and a school class as the test group, comparison of white results with black results is already confounded by the fact that the **test conditions** are not similar.

Testers should be wary of making comparisons between the races in South Africa as a result of the existent discrepancies. Equating the conditions of schools in South Africa will be a long term process with funding needed and integration between black and white students. These factors make consideration of the test conditions important and justify the adaptation of the test instructions and presentation, so as to compensate for the disadvantaged position that some testees find themselves in the testing situation.

3.3 Adapting Tests for Cross-Cultural Use

Schwarz (1963), identifies two types of problems faced by the constructors of any psychological test. The first, he says will determine the kinds of abilities that will be measured, and the second involves the operations that the subjects will be asked to perform, to demonstrate these abilities in the test situation. A change in the cultural setting will not warrant a change in the type of abilities to be measured, as validity studies have confirmed the common assumption that the abilities needed for a specific activity do not vary between

cultures. What does need to be altered though, are the types and number of tasks that are required to measure the particular ability. The reason for making quantitative changes, Schwarz (1963) says, is that in an aptitude test, for example, only a few component operations are central to the specific ability being assessed in aptitude tests. The other operations only serve to confuse the testee operating in a culture other than the one in which the test was devised. Schwarz (1963) suggests that the test format should be stripped to essentials, when tests are to be applied in cross-cultural testing. This helps to attain maximum test performance.

Schwarz (1961), in discussing aptitude tests for use in developing nations, identified nine general principles that should be used to adapt tests for cross-cultural use.

> 1. The testing procedure should not presuppose any response as being automatic on the part of the examinee. It should include explicit provisions for for teaching him every response that he will be expected to make (p. 14).

Schwarz is not only referring to the responses that are needed for the solving of test problems, but also the manipulations that are normally required in the testing situation, that is, the handling of pencils, test booklets and answer sheets. The instructions should also include how to cope with "emergency" situations. For example, what to do if a picture in the test, is not intelligible.

> 2. The design of the test booklet should minimize the number of constraints imposed on the examinee's performance in working from the first page to the last. Insofar as possible, instructions and cautions irrelevant to the solution of the problems should be eliminated (p. 15).

The design of the test booklet should be such that the examinee does not have to decide when to continue to the next page, when to stop and check the work completed, and when to wait for further instructions. Different tasks should not be printed on on the same page, and different tests should not be included in the same booklet. This allows for the smooth completion of the test by the testees.

> 3. The test should not rely on any printed instructions for teaching or controlling the responses to be made. The test booklet should include no such instructions (p. 15).

Schwarz (1961), found that any form of printed instructions tended to confuse Nigerian testees. This he relates to their not being accustomed to receiving directions in printed form.

The above three suggestions attempt to compensate for the naivete of the test group, when completing such tests.

> 4. The most effective means of teaching the test is through the use of visual aids, supplemented by active demonstration. These aids should replicate as closely as possible the exact operation to be performed (p. 16).

These visual aids would serve two functions; (a) to help reduce concepts and "mental" processes to the level of physical operations, and (b) to act as a vehicle for demonstration. For the first function, if a pattern has a missing piece, and the subjects have to choose from a number of options which option completes the pattern, as in the Raven's SPM, it would help to show the students; how the pattern is not completed by some options, by physically placing some of the incorrect options in the incomplete portion; and how the correct

option completes the pattern.

> 5. Explanations necessary to supplement the demonstrations should be given in local patterns of speech and expression (p. 17).

Any supplementary instructions would probably be most effective if given in the native language of the subjects. However if English instructions have to be used then it does become crucial to eliminate phrases that are not immediately understood, and to introduce accepted local idioms. Use of English words that are used in the local language, should do away with redundancies and formal language (Schwarz, 1961).

The presentation of the tasks to be carried out are a crucial factor that can influence test performance. Schwarz (1963), labels this the **media of communication**. He identifies two processes. The first is the explanation of the test procedure, and the second being the actual test content. The first of these two he sees as being more important. Language must be identified as a factor influencing test results. Schwarz (1961), found that the medium of communication rather than language was a limiting factor. When dealing with the input supplied by the actual test exercise, that is, perceptual factors, Schwarz (1963), found that pictorial symbols with no significant meaning or when its significance need not be understood in order to complete the task, the African examinee does not seem to be handicapped to an appreciable degree. It is only when the referent symbol is important that difficulties arise. Thus to overcome these problems Schwarz (1961, 1963), suggests that a spoken commentary is used to link the use of visual aids, articulated models, and demonstration in a

supplementary manner.

Suggestions Four and Five deal with the design of administering effective instructions, and are based on certain differences in the attitudes of Nigerian examinees that are not usually found with American test groups. Crawford-Nutt (1976), employed these suggestions in a South African study with interesting results. This is discussed later.

> 6. The training session should include supervised practice in doing the test problems, with a specific provision for feedback to the examiner. Such practice and feedback must cover not only the basic task, but also any variation that may be incorporated in certain of the test items (p. 18).

Schwarz (1961), believes that the need for feedback derives from the examinees unwillingness to admit their own confusion, or to question anything that the examiner might be doing. Practice problems should require the examinees' to display their understanding of all the operations they should have learned.

> 7. To get maximum examinee cooperation, the testing procedure should differ sharply from the routines to which they are normally accustomed in school. Elements of the dramatic and flamboyant inspire the peak effort that is necessary for effective aptitude measures (p. 19).

Testing in school settings usually takes place in the school hall or classroom and this produces natural identification with accepted school practices that any testing procedure must try to break down. Thus it may be necessary to conduct the test instructions at a level that is serious yet nevertheless fun. This will ensure that examinees

respond more favourably to the test. Schwarz (1961), says that the application of these principles should provide a workable testing procedure. Schwarz separates the above problems of cross-cultural testing, from the problems of test content. He found that the majority of non-verbal test materials that are being used in the United States can also be used with Nigerians if the instructions are properly presented.

> 8. It is seldom possible to predict on logical grounds which tasks foreign examinees can and cannot do. Each new test should be subjected to thorough experimental investigation (p. 20).

Schwarz (1961) believes that test items should not be based on intuitive notions or expert opinions of appropriateness, since these generally underestimate the abilities of the examinees. Researchers should bear in mind that there are certain cultural deficits in the examinees' ability to perform essentially Western skills, that they have no opportunity to practice in their own culture.

> 9. Preference should be given to items that are reasonably independent of individual differences in the tactics or strategies of the examinees. It is usually not possible to enforce a uniform strategy that will be followed by the entire group (p. 20).

Tests, requiring a uniform approach to solving test problems by all the examinees whose scores will be compared, require the teaching of fine points of testmanship, which is not feasible in the cross-cultural situation.

The above nine points may be summarized in three considerations for testing across cultures:

1. Make drastic changes to the test instructions, relying on dynamic techniques,
2. Check the adequacy of all other verbal and perceptual stimuli, and
3. Modify or supplement deficient elements.

Orter (1972) says that the extent to which a test needs to be modified depends on the degree of difference between the culture in which the test was developed and the culture in which the test is being applied. The only way that such a differences may be identified, is through the correct use of tests.

3.4 Conclusions

From the above, it must concluded that intelligence tests have often been misused in the cross-cultural context, due to a lack of knowledge about the intricacies of testing across cultures. However, now with the vast amount of research that has been carried out in the cross-cultural field, a great deal more is understood about the intricacies of the testing situation, and it is important that researchers do not misuse, so called **culture-fair** tests. To ensure this, researchers should continuously remember the notions of **fair test conditions**, **introducing the test**, and the **adaptation of the test instructions**, for the compensation of the disadvantaged person.

More research needs to be carried out into the field of test adaptation. Samuda (1983), argues that this is justified.

> Equal opportunity cannot and should not be interpreted as 'We treat them alike.' This means in practice that appropriate instruction should be matched to individual differences whatever the child's group membership. In a nutshell, 'different strokes for different folks'" (Scarr cited in Samuda, 1983, p. 601).

CHAPTER 4 - THE RAVEN'S PROGRESSIVE MATRICES

4.1 Justification for the use of the Raven's Progressive Matrices (RPM)

Taking into account the criteria for minimizing bias in cross-cultural testing (Chapter 3) and the widespread use of the test across cultures, it is argued in this Chapter that the Raven's Progressive Matrices' (RPM) are valuable for assessing intelligence in the South African setting. Following a brief justification for its use a fuller analysis of the test, its history and use will be provided.

The RPM was designed to cover the widest possible range of mental ability and to be equally useful with persons whatever their education, nationality or physical condition (Raven, 1983). The types of problems presented in the RPM series, are not based on a particular cultural milieu, allowing for its applicability to different cultures. The **test content** does not require the testee to understand functions of objects which may be particular to a culture other than his/her own. That is, the call for information in the test is not specific to one culture.

The RPM being a non-verbal test minimizes the effect of language on testing. To allow for the applicability of this test to cultures with differing educational backgrounds, as is the case in South Africa, **reading** is not required by the testee. Another parameter which the RPM negates, is **speed**. There is no time limit to the RPM tests. The above four parameters; **test content, language, reading, and speed**, are

identified by Anastasi (1988), as being parameters which test constructors should consider in attempting to make tests culturally fair.

The Standard Progressive Matrices (SPM), one of the scales in the RPM series, has being used internationally for many years, with no great revision to the original test having been necessary.

4.2 The Raven's Progressive Matrices (RPM)

The idea of a Matrices test stemmed directly from the work of Spearman. J. C. Raven was a student of Spearman's. The Raven's Progressive Matrices (RPM) and Mill Hill Vocabulary (MHV) scales were developed to assess, as straight forwardly as possible, the two components of general intelligence identified by Spearman in 1923 (Raven, 1989). The RPM measures **eductive** ability, and the MHV, **reproductive** ability.

The Progressive Matrices series is made up of three scales. The Coloured Progressive Matrices (CPM), which may be used for very young children between the ages of five and eleven, and also for individuals with intellectual deficiencies. The CPM measures the degree to which the test takers **eductive ability** has developed or deteriorated. This test has been established as a reliable and valid test with wide applicability in cross-cultural settings (Esquivel, 1984). The Advanced Progressive Matrices (APM) may be used with people of advanced or above average intellectual ability, and when used with a

time limit is regarded as an assessment of the individual's intellectual efficiency, and with no time limit an assessment of a person's total capacity for observation and clear thinking (Vernon, 1984).

The third scale is the Standard Progressive Matrices (SPM). The SPM is used both for children, six years and older, and adults. The SPM scale is divided into five sets of twelve problems each. Each set starts off with a problem that is self-evident, followed by problems becoming progressively more difficult. This provides the ability to deduce the consistency of a persons intellectual activity in five successive lines of thinking. Sets A to E are designed to cover the whole range of intellectual development from the time a child is just able to grasp the idea of finding a missing piece to complete a pattern, to a measure of the maximum possible capacity for coherent perception and orderly judgment. The length of the test is sufficiently long to measure these activities without being to exhausting (Raven et al., 1983). As this is the most widely used test and designed for use in the general population further discussion of the RPM will focus mainly on this scale - the SPM.

The principles used in the five sets may be identified as the following: 1) Completion of a pattern in a continuous figure, 2) Figural analogy in a two-by-two matrix, 3) Systematic alteration of a pattern in a three-by-three matrix, 4) Systematic permutations and alterations of figures in a three-by-three matrix, and 5) Systematic decomposition and synthesis of figural parts in a three-by-three matrix. Sattler (1982), sees the Progressive Matrices as a useful

measure of nonverbal reasoning ability. It measures a person's ability to form perceptual relations and to reason by analogy independent of language and formal schooling (Llabre, 1984).

> The Standard Progressive Matrices (Sets A, B, C, D, and E), or SPM, is a test of a person's capacity at the time of the test to apprehend meaningless figures presented for his observation, see the relations between them, conceive the nature of the figure completing each system of relations presented, and, by so doing, develop a systematic method of thinking (Raven et al., 1983, p. SPM2).

Raven et al. (1983), describe the Matrices scales as tests of observation and clear thinking, with the order in which the problems are presented providing the standard training in the method of working. This makes the understanding of the initial items crucial. The first and second sets in the scale, and the introductory problems of the third and fourth sets provide, for the adult, little more than training in the method of thinking.

The three kinds of the Progressive Matrices, the CPM, the APM, and the SPM, should not be regarded by themselves as tests of intelligence. To be a test of general intelligence, they must be used in conjunction with a vocabulary test to provide an index of **reproductive ability**. The Matrices tests only provide a measure of the **eductive ability** of which Spearman spoke about. An advantage of this structure of the test is that, the examiner is able to assess separately a person's present capacity for intellectual activity with no regard to their acquired knowledge, and at the same time investigate the verbal information the test taker has acquired in the past, with as little present intellectual activity as possible (Raven et al., 1983). Raven (1989), shows that the RPM scales are among the

purest available measures of **g**.

Raven et al. (1983) argued that a person's total score provides an index of his/her intellectual capacity, with relatively little influence from the cultural environment in which the individual grew up or his education. Raven et al. (1983), deduced this from published correlations between children's scores on the Progressive Matrices and test of reading, spelling and elementary arithmetic.

4.3 Standardizing the Raven's Standard Progressive Matrices (SPM)

Raven (1941), standardized scores from the Raven's SPM (1938), by converting the raw scores, at half yearly intervals to percentile points. That is, percentile scores were calculated for children at the age of six years, six years and six months, seven years, seven years and six months, and so forth. Percentile grades measure the individuals ability in terms of the frequency with which an equal degree of ability may be expected to occur, in a random group of testees of the same age. Percentile scores are derived scores expressed in terms of percentages of persons who fall below a given raw score (Anastasi, 1988). The mental age of the testee is the age at which the median score is equal to the testee's score.

Anastasi (1988) identifies the following advantages of this method. It is; easy to compute, easy to understand, universally applicable, may be used with children and adults, and may be used for any test. She does, however, note a statistical drawback, in that, if the

distribution of raw scores approximates the normal curve, then the raw score differences near the ends of the distribution are greatly reduced.

Raven et al. (1983), argued that the method of interpreting the significance of a person's total score, is to consider it in terms of the percentage frequency with which a similar score is found to occur amongst people of his/her own age. The advantage of using this method is that no **a priori** assumption is made that in childhood the development of intellectual capacity is necessarily uniform or distributed symmetrically. Raven's SPM considers certain percentages of the population, and groups people's scores accordingly. In this way it is possible to classify a person according to the score s/he obtains as:

GRADE I	or **"intellectually superior"**, if his score lies at or above the 95th percentile for people of his age group.
GRADE II	**"definitely above the average in intellectual capacity"**, if his score lies at or above the 75th percentile; II+ if his score lies at or above the 90th percentile.
GRADE III	**"intellectually average"**, if his score lies between the 25th and 75th percentiles; III+, if his score is greater than the median or 50th percentile; III-, if his score is less than the median.
GRADE IV	**"definitely below average in intellectual capacity"**, in his score lies at or below the 25th percentiles; IV-, if his score lies at or below the 10th percentile.
GRADE V	**"intellectually defective"**, if his score lies at or below the 5th percentile for his age group.

(Raven et al., 1983, p. SPM17-SPM18).

Once the Raven's Progressive Matrices raw score has been calculated and converted into the percentile rank, it is possible to convert the percentile rank to an IQ range. Peck (1970) provides the method for this. This is not used often and used mainly for comparisons with assessments of other test results.

Following the first standardization of the Raven's SPM, Raven (1941), found that median scores show slow development in eductive ability up to the age of eight years, a fairly rapid development of the ability between the ages of eight and thirteen, and at the age of thirteen and a half years, innate eductive ability normally reaches its maximum development. The sample used to establish the first set of norms was drawn from Ipswich, a manufacturing port. The children tested were sampled from elementary, secondary, public, and special schools. The adults tested were sampled from all the occupations represented in the port.

4.4 The Reliability of the Raven's SPM

Many reliability studies have been carried out. Raven et al. (1983) in drawing together these findings concludes that, when studies suffering form inadequate sample size and studies where the SPM was applied to subjects too young for the SPM are excluded, a general picture of good reliability emerges. Re-test reliability for the SPM is described as being satisfactory for a period up to one year. Raven et al. (1983), provide an average short term reliability of 0.90 reducing to 0.80

with longer intervals. Studies quoted in Raven et al. (1983), were carried out in the Congo, India, Belgium, and America. The studies thus included various cultural groupings.

Regarding the internal consistency of the test, Burke (1972), suggests test "reliability seems quite adequate for individual decision making, at least for age ranges above 25 in an American veteran or similar adult male population" (Burke, 1972, p. 254). Burke's (1972), split-half reliability figures range from 0.89 to 0.97. Burke (cited in Raven et al., 1983), found that the corrected split-half reliability for Raven scores was better than the corrected split-half reliability for WAIS Performance Scores. Other studies referred to in Raven et al. (1983), lead them to the conclusion that Burke's suggestion, or confidence as it is referred to, may be extended more widely.

4.5 The Validity of the Raven's SPM

The Raven's SPM, has been claimed to be the purest and best measure of 'g'. Evidence for this is found in research completed by Emmett (1949), Gittins (1952), Nisbet (1953), (all cited in Raven et al., 1983); who found loadings of up to +0.83 on 'g' with British children, and +0.81 with American children. Cross-cultural studies have also confirmed a high 'g' loading (Irvine, 1966; Keehn & Prothro, 1955; Mehrohtra, 1958; cited in Raven et al., 1983). Research with adults reveal loadings of up to +0.86 on 'g' (Burke & Bingham, 1969 cited in Raven et al., 1983).

However, Sattler (1982) reports conflicting factor analytic results, with some studies showing the Progressive Matrices to contain primarily one factor, labeled **g**, signifying reasoning, while others indicate more than one factor. The multiple factors include, concrete and abstract meaning, continuous and discrete pattern completion and patterning through closure. Thus some factor analytic studies have shown that the SPM does measure more than the general factor of **g**. Raven et al. (1983), concludes that while the SPM is a relatively good measure of general intellectual ability, it is not purely a **g** estimate. The assessment of other factors may be also be made.

4.6 Research Using the Raven's SPM

Tulkin and Newbrough (1968), carried out a study which investigated the effects of life experience upon performance on Raven's SPM. A fundamental finding of this study was that race was only important in distinguishing performance in the low-class groups. In addition to this gender differences were found to be more important in Negro groups. Tulkin and Newbrough (1968), suggest that the Raven's SPM may be used to assess the intelligence of children with language and/or cultural handicaps. All the above findings are crucial to the discussion of the results obtained through this research.

Following the 1979 Standardization, where a nationally representative sample of 3 500 British school children aged six to 16, excluding those attending special schools, were tested; Raven et al. (1983) confirm the following:

> 1. There are no sex differences in the scores obtained on the SPM except a the age 11+/-six months.
> 2. Only 9% of the within-age variance is explained by social background.
> 3. As in the 1938 and 1972 standardizations, the test works - "scales" - in the same way for children from different socio-economic status backgrounds. It is not "foreign to the way of thought of children from certain backgrounds".
> 4. Once items become too difficult for children, they get the item right far less often than would therefore, not "random", but guided by hypotheses - albeit wrong hypotheses (Raven et al., 1983, p. SPM5).

Raven et al. (1983), after studying the 1979 standardization results, concludes that there is considerable stability of performance on the SPM, both within and between Western societies.

Raven's assumption that a person's total score provides an index of his/her intellectual capacity, with little influence from his/ her education, needs to be investigated. Vincent and Cox (1974), found differences in scores on the RPM between racial groups based on the level of education obtained. Raven's finding of no gender difference on the RPM, was refuted in the cross-cultural situation by Tulkin and Newbrough (1968), when they observed that gender differences tended to be more significant in Negro groups. These findings should be considered whilst analyzing the data obtained in this study.

4.7 Administering the Raven's SPM in South Africa

Crawford-Nutt (1976), carried out an important study, where the importance of test administration was investigated. Pons (1974), developed a presentation method for the Raven's SPM which used suggestions made by Raven (1960), principles derived by Schwarz (1961), and discussions with S. H. Irvine. Crawford-Nutt (1976) noted that in the study of the differences between intelligence test scores between blacks and whites, the notion of test administration had largely been ignored. Research shows that if the method of test presentation reduces the anxiety of the subjects about being tested, the results that African testees obtain are on a par with the results obtained by white testees (Crawford-Nutt, 1976). He ensured however, that the altered method of test presentation did not alter the nature of the test itself. The alterations removed the disadvantages from which black testees frequently suffer through not understanding what is expected of them in the test situation, thereby ensuring that the subjects understood the demands of the task.

Crawford-Nutt (1976), administered two forms of the Raven's SPM, namely the 1938 and 1958, versions, to black male and female students after 11 and 12 years of schooling. The results obtained showed that "when sufficient attention is paid to relevant detail in presentation to ensure equivalent understanding of the test and its requirements among those being tested, then and only then can comparable results be expected" (Crawford-nutt, 1976, p. 202). When comparing the mean scores of the two samples of black pupils tested, with the published norm group (Raven, 1960), the reported difference in scores between

blacks and whites tested on this test did not occur. The mean score for the black students were 45.1 and 44.9; with the equivalent age group norm for the white group being 44 (Raven, 1960).

Crawford-Nutt (1976) suggests three possible reasons for the obtained results. It may have been due to the level of education of the black students used in the study. Vincent and Cox (1974), found no significant difference in scores on Raven's Progressive Matrices, between any of the racial subsamples based on years of education. Crawford-Nutt (1976) identifies the degree of test sophistication of the subjects as another possible reason for the result. The ability to respond to a test situation in the correct manner was identified by Schwarz as "the fine points of testmanship". That is the results may have been due to the fact that the subjects were familiar with the experience of writing such tests and thus reduced **content anxiety** and **content uncertainty** (Pons, 1974, cited in Crawford-Nutt, 1976). The final reason as to why the results were obtained, may be due to the method of test presentation. The materials and concepts used in this study were familiar to the black subjects and thus when the black subjects came to do the test they were in a similar position as their white counterparts. Crawford-Nutt concludes that "if an educated group of black people is administered the RPM in a manner that ensures their understanding of the requirements of the test to the same extent as a similar sample of whites, then they will perform in a manner equal to that of the whites" (Crawford-Nutt, 1976, p. 205).

4.8 Conclusions

The above studies are a small sample of a large number of studies carried out using the Raven's Progressive Matrices. The above studies have been cited as these findings have influenced the collection and analysis of data in this research project.

While the Raven's SPM is considered to be a culture-fair test that is valid and reliable, research cited in this chapter and Chapter 3 (Irvine, 1969; Wober, 1969; Pons, 1974; and Crawford-Nutt, 1976) point to the need to consider some adaptations to the test instructions so as to ensure that the task demand is understood by the subjects. Drawing from Crawford-Nutt (1976), the method of test presentation does seem to have an important role to play in the test taker coming to an understanding of the task demand. Thus it is important for instructions of **culture-fair** tests to be adapted in such a manner to allow for the test taker to come to a thorough understanding of the task demand, which should ensure a true reflection of the child's ability. The manner in which the test instructions were adapted for this study will be discussed fully in Chapter 5. This presentation of the test's instructions should ensure the **fair** use of the test in a multicultural society, leading to the results of the test being comparable.

CHAPTER 5 - THE STUDY

5.1 Introduction

First, psychologists in Grahamstown in the absence of any other suitable test for assessing general intellectual ability, use the Raven's SPM to assess African pupils, but do not have any adequate norms from which to work (Dr. R. Brooke, personal communication, September 1989). With the Raven's SPM proving to be reliable and having a strong validity in other studies, there was a clear need for a normative study in the Grahamstown region to be conducted.

Second, drawing on the discussion on the concept of intelligence, the assessment of intelligence, and the Raven's Progressive Matrices a better understanding of the role of cross-cultural testing of abilities was defined as a need. More specifically, the need for appropriate test instructions for different people, and how these different instructions may influence test performance is an area worthy of investigation.

In this chapter the objectives of the study will be outlined, followed by the research method used to achieve these objectives. The methods used to adapt the Raven's SPM to ensure the fair use of the test will be set out, as will the design of the study.

5.2 The Aims of the Study

The objectives of the study were:

(a) To establish norms for the Raven's SPM for Secondary School African pupils in the Grahamstown region.

(b) To investigate the effect of different methods of test presentation on pupils' performance.

(c) To compare the scores obtained by pupils on the Raven's SPM across the different age groups, gender, standards, and schools, in order to identify the possible effect of these variables on test scores.

(d) To compare the scores of the pupils obtained in the study with the results of other studies.

5.3 The Research Design

The basic research design consisted of testing a sample of male and female students across Standards Six to Ten (eighth to thirteenth year of schooling) in all three of the Secondary Schools for African Pupils in the Grahamstown Municipal area. A disproportionate stratified sample at each strata was drawn. This sample was tested under conditions prescribed by Raven and the instructions were both translated and adapted for the local conditions. This sample will be referred to as the **Normative sample.**

In order to investigate the effects of different test presentations on performance, a second smaller sample was drawn from the schools and tested using the Raven's instructions but given in Xhosa. This sample will be referred to as the **Alternate sample.**

Further particulars on the nature of the sampling and a description of the samples is given in sub-section 5.5.1.

5.4 The Instrument (The Raven's Standard Progressive Matrices)

5.4.1 Adapting the Raven's Standard Progressive Matrices (SPM).

a) The translated instructions.

The Alternate group was administered the test using the standardized instructions provided by Raven et al. (1983) except that these were translated into Xhosa. The translation was done according to the back translation method suggested by Brislin (1970). Using this method the original instructions were translated into Xhosa by a bilingual person. Once this had been done, a second bilingual person translated the Xhosa instructions back into English, without having seen the original English instructions. The researcher then compared the original English instructions, with the back-translated English instructions. Any discrepancies that were found were discussed with the translators and a Xhosa version that was as close to the original instructions as possible was adopted. These translated instructions were then administered to the alternate group. (See Appendix A for the translated instructions).

b) The adapted/translated instructions.

The Normative group was tested under adapted conditions taking into account the work of Schwarz (1963), Pons (1974), and Crawford-Nutt (1976). The following changes were made to the test instructions originally set by Raven (1938).

(i) The illustrations of the record form and Item A1 (Raven et al., 1983 refers to this as Test A1) were drawn three times the size of the original. This deviates from Raven et al's. (1983) suggestion that the record form and Item A1 should only be drawn twice the original size. An illustration of Item A2 was also made. These illustrations, however, deviated from the originals in that, the parts below the pattern, i.e. the elements that subjects have to choose from to complete the pattern, were movable. That is, the tester was able to pick up one of the options that the testee had to complete the pattern, and place it in the blank space in the pattern, to show how some of the options did not complete the pattern, and how the correct one did. This was also done for Item A2. This modification meant that the material used for instructing testees relied mainly on visual aids, articulated models, and demonstration, as discussed by Schwarz (1961, 1963) and Crawford-Nutt (1976). Item A2 was used in the same way as Item A1 was used. Extending the demonstration to Item A2 was done so as to enable overlearning.

(ii) The instructions were extended so as to allow for a spoken commentary to link the visual aids, the articulated model and demonstration, as recommended by Schwarz (1963). The instructions did not differ from the original standardized instructions, except that they were extended to include Item A2, and allowed testees to participate by asking questions. The language and phrases used by Raven in the original instructions were adopted. The adapted instructions provided the testees with opportunities to participate mentally and vocally in the process of instruction, as suggested by Crawford-Nutt (1976). These adapted English instructions were then translated into Xhosa using the back-translation method, as described above.

5.4.2 Administering the Raven's Standard Progressive Matrices (SPM).

a) Administering the translated instructions.

The African tester fluent in Xhosa, was introduced to the class to be tested, by a member of staff at the school. She then introduced herself to the class once the member of staff had left, by her first name. The language spoken at all times was Xhosa, the testees home language. This was done so that a good rapport between the tester and testees could be established. The intention was to breakdown the authoritative status that testers usually have. The reason for the testing was explained. The tester said that a student at Rhodes University was carrying out a study, at the request of psychologists who felt it necessary to obtain a group score for pupils, on the test that they were to complete. At this point she emphasized that the

results obtained by the individuals would in no way jeopardize their position in the school, as the researcher was only interested in the group score. The pupils were also aware of the fact that the researcher had entered into a discussion with the school's Student Representative Council, in order to get permission to carry out the study. Pupils in addition to this were told that they did not need to fill in their names on the record form, thus ensuring their anonymity. The pupils were then told that their class was being tested as they had been chosen by chance. To further explain this idea, it was conveyed to the class that their class, had been drawn out of a hat, with all the other classes names being in the hat. All the above was carried out for ethical reasons, and also to reduce the anxiety experienced by testees at the time of testing.

The tester then placed covered illustrations of the record form and Item A1 (double the size of the original), on the chalkboard in front of the class, and then handed closed test books and record forms to the class, asking them not to open the books. The illustrated record form and Item A1 was drawn in black on white cardboard. The tester then asked if there were any questions, and once all the questions had been answered she uncovered the illustrated record form and asked the subjects to please fill in their school, standard, age, and gender in the relevant positions on the record form. Following this the tester uncovered the illustration of Item A1 and proceeded with the translated instructions, as discussed above. After the instructions, were delivered orally, the tester asked the students to work through the test at their own pace, and when finished to please put there

hands up, at this point the tester wrote at the bottom of the paper the time that the test was completed. (See Appendix A for the original and translated instructions).

b) Administering the adapted/translated instructions.

The African tester fluent in Xhosa, was introduced to the class by a member of staff at the school. She then introduced herself to the class once the member of staff had left, by her first name. The language spoken at all times was Xhosa, the testees home language. This was done so that a good rapport between the tester and testees could be established. The intention was to breakdown some of the authoritative status that testers usually have. The reasoning for the testing was explained. The tester said that a student at Rhodes University was carrying out a study, at the request of psychologists who felt it necessary to obtain a group score for pupils, on the test that they were to complete. At this point she emphasized that the results obtained by the individuals would in no way jeopardize their position in the school, as the researcher was only interested in the group score. The pupils were also aware of the fact that the researcher had entered into a discussion with the school's Student Representative Council, in order to get permission to carry out the study. Pupils in addition to this were told that they did not need to fill in their names on the record form, thus ensuring their anonymity. The pupils were then told that their class was being tested as they had been chosen by chance. To further explain this idea, it was conveyed to the class that their class had been drawn out of a hat, with all the other classes names being in the hat. All the above was

carried out for ethical reasons, and also to reduce the anxiety experienced by testees at the time of testing.

The tester then placed covered versions of the illustrated record form, Item A1, and Item A2 (three times the size of the originals), on the black board. She then handed out closed test books and record forms to the students asking them not to open the books until instructed to do so. The tester then asked if there were any questions, and once all the questions had been answered she uncovered the illustrated record form and asked the subjects to please fill in their school, standard, age, and gender in the relevant positions on the record form. Following this the tester uncovered the illustration of Item A1 and proceeded with the adapted and translated instructions, as discussed above. After the instructions, were delivered orally, the tester asked the children to work through the test at their own pace and when finished to please put there hands up, at this point the tester wrote at the bottom of the paper the time that the test was completed. See Chapter 8 for the translated and adapted instructions.

5.4.3 The scoring.

Once the whole class was finished, the record forms were placed in an envelope, on which was written: the school, class, number of pupils in the class, the time the test was started, and which test instructions was used.

The researcher marked the test by awarding one point for every correct answer and no point for a wrong answer. From this a total score out of a possible 60 was obtained. This scoring does not differ from the scoring of Raven et al. (1983).

5.4.4 Biographical information.

The following biographical information prompts appeared on the record form handed to the students.

Name: _____
Age: _____
Sex: _____
School: _____
Standard _____

However as already mentioned in section 5.4.2, the pupils were told that it was not necessary to fill in their names, so as to ensure anonymity.

5.5 Sampling

5.5.1 Description of the sample.

The population consisted of all Secondary School African pupils in the Grahamstown Township. The area had three secondary schools: Nathaniel Nyaluza High School, Ntsika Junior Secondary School, and Nombulelo Senior Secondary School. The total number of pupils in the three schools was 3 232.

Table 1

Distribution of Subjects in the Population and Sample

School	Standard	Population Size	Alternate Sample Size	Normative Sample Size	Total Normative Size
Nyaluza	6	110		69	
Nyaluza	7	302		74	
Nyaluza	8	237	50	21	
Nyaluza	9	171		72	
Nyaluza	10	246		50	336
Total		**1 066**	**50**	**286**	
Ntsika	6	173		38	
Ntsika	7	202		52	
Ntsika	8	260		39	
Ntsika	9	238		54	183
Total		**873**	**0**	**183**	
Nombulelo	6	296		46	
Nombulelo	7	364		43	
Nombulelo	8	210	51	47	
Nombulelo	9	240		62	
Nombulelo	10	183		44	293
Total		**1 293**	**51**	**242**	
Grand Total		**3 232**	**101**	**711**	**812**

The range of the ages of the pupils was from 12 to 27. The breakdown, as represented in Table 1, reflects the position at the time of testing, which spanned over the month of August 1990. To obtain a representative sample of the population, it was decided to sample students from all three schools in the area.

A **Disproportionate Stratified Sample Design**, as purported by Nachmias and Nachmias (1981) was adopted. In this research the strata comprised the different standards (year of schooling) in the schools. Both Nathaniel Nyaluza High School and Nombulelo Senior Secondary School, had Standards Six to Ten. Ntsika Junior Secondary School, only had Standards Six to Nine. All the schools had a number of classes for each standard. One class from each standard was chosen according to the **simple random sampling technique** (Nachmias et al, 1981). In this design each of the sampling units (standards) of the population, had an equal probability of being chosen. Nachmias et al's. (1981) **lottery method** was used. Each class for each standard was represented by a disk. These were placed together, and mixed well; and a sample drawn from them. For each standard, at each school this method was used. At times it was necessary to sample a second class from a standard as the number of pupils in the class was small. In these cases the class that had previously been chosen was included in the group to be chosen again. This method of replacement, ensured that each class had the equal probability of being chosen. If the same class was selected then it was replaced in the group and the draw was repeated. The above method ensured that a representative sample of the population of

classes was chosen for the Normative group of the schools.

The Alternate sample was chosen from standard eight pupils at Nathaniel Nyaluza High School and Nombulelo Senior Secondary School in a similar manner. Access to Ntsika Junior Secondary School, was not possible at the time of testing the **Alternate group**. The Alternate group was a smaller group and made up of only two standard eight class's. The total number of students was 101.

> The essential requirement of any sample is that it be as representative as possible of the population from which it is drawn. A sample is said to be representative if the analyses made on its sampling units produce results equivalent to those that would be obtained had the entire population been analyzed (Nachmias et al., 1981, p. 429).

In drawing the samples the researcher had to operate with many constraints. Among the problems faced were: (a) lack of testing facilities, and (b) testing having to be carried out in the afternoons at Nombulelo Senior Secondary School.

The decision to sample classes instead of individual pupils was forced on the researcher because the pupils were made available to the tester in this way by the schools. As resources in black schools are limited, there were no classrooms available for the pupils to be the tested in, had certain students been removed from the classroom situation. By sampling classes, it was possible for pupils to be tested in the particular classroom that they would have had a lesson in. Ntsika Junior Secondary School and Nathaniel Nyaluza High School allowed for the testing to take place during school time, while Nombulelo Senior

Secondary School would not allow for the disruption of any classes. This meant that the testing had to take place in the afternoons, after the teaching ended. In this school the reasoning behind sampling classes was two fold: (a) Students would more than likely remain after school for the testing if an entire class was tested, than if only a number of pupils from the class were tested, and (b) To maintain consistency in the method of sampling. In this way all age groups, in the schools were sampled. The classes at all three schools were structured according to subject choice and not individual ability. Random sampling of the classes ensured that a representative sample of students across the various subject groupings, was obtained.

During the time of testing at Nathaniel Nyaluza High School, many subjects were absent due to the school attempting to raise funds for a school tour. This necessitated re-sampling Standards Six, Seven, Eight, and Nine. However access to a second standard eight class for the **normative sample** was not possible, this being the reason for the small number (21) of standard eight pupils sampled at Nathaniel Nyaluza High School. The re-sampling led to the large number of pupils that were sampled at Nathaniel Nyaluza High School, as compared to the number of pupils sampled at the other two schools.

Ntsika Junior Secondary School only caters for Standards Six to Nine. Thus, the reason for the absence of a Standard Ten class.

To allow for the Normative and Alternate samples to be compared, the Alternate sample was matched with the Normative sample in the following way. From the subjects in the **Alternate sample**, 17 pupils from the Standard Eight class sampled at Nathaniel Nyaluza High School, could be matched according to age and gender with Standard Eight pupils in **Normative sample** at Nathaniel Nyaluza High School. 28 pupils from the **Alternate** Standard Eight class sampled at Nombulelo Senior Secondary School could be matched according to gender and age with Standard Eight pupils in the **Normative sample** at Nombulelo Senior Secondary School.

5.6 Statistical Analysis

The data obtained from this study was analyzed using the BMDP computer package of statistical analysis. This package enables the social science researcher to carry out a wide range of statistical procedures and manipulate research data.

The following BMDP programs were used:

1) The t-tests were conducted using BMDP3D.
2) The analysis of covariances were conducted using BMDP2V.
3) The cross-tabulation was conducted using BMDP4F.

CHAPTER 6 - THE RESULTS

6.1 The Description of the Samples

6.1.1 Breakdown in terms of age and gender.

In this section a breakdown of the number of male and female subjects, for the Alternate and Normative samples are presented in terms of age in Table 2 (page 69).

The Normative sample made up of 711 subjects, has low numbers at the lower and upper ends of the ages sampled. For this reason when carrying out the analysis of covariance, the sample was reduced to include pupils of the ages 17 to 22 only. The number of subjects used for the analysis of covariance was 532. The largest number of subjects fall in the 19 year old age group.

The Alternate sample consisted of 101 subjects providing a limited spread of respondents in the different age groups. The largest number of subjects fell in the 20 year old group. This does not allow for an analysis of the results across the different age groups.

The number of females in the Normative and Alternate samples is larger than the number of males.

Table 2

Breakdown in terms of Age and Gender: Normative and Alternate Samples

Normative Sample			Alternate Sample		
Age	Male	Female	Age	Male	Female
12	0	1	12	0	0
13	5	4	13	0	0
14	5	3	14	1	1
15	9	16	15	0	2
16	16	46	16	3	0
17	31	49	17	1	2
18	28	51	18	9	5
19	53	85	19	8	12
20	56	43	20	10	19
21	40	44	21	6	7
22	29	23	22	4	5
23	18	15	23	1	2
24	9	13	24	0	1
25	5	6	25	0	0
26	3	4	26	1	1
27	1	0	27	0	0
	308	403		44	57

6.1.1.1 The mean ages of the subjects.

The Normative Sample: The mean age of the total Normative sample is **19.3** years. The mean age of the male Normative subjects is **19.6** years. The mean age of the female Normative subjects is **19.0** years.

Comparing the differences in ages between the genders for the Normative sample the following results were obtained from a t-test.

t = 2.96; p value = 0.0032; d.f. = 709.

The null hypothesis of no significant difference between the ages of the genders of the Normative sample is rejected ($p < 0.05$) and the alternate hypothesis of a significant difference accepted, with males in the sample being significantly older than females.

The Alternate Sample: The mean age of the Alternate sample is **19.7** years.

6.1.2 Breakdown in terms of school and standard.

In this section a breakdown of subjects for the Alternate and Normative samples is presented in terms of standard and school in Table 3.

Table 3

Breakdown in terms of School and Standard:

Normative and Alternate Samples

Sample	Standard	School		
		Nombulelo	Ntsika	Nyaluza
Normative	6	46	38	69
	7	43	52	74
	8	47	39	21
	9	62	54	72
	10	44	-	50
(Total)		242	183	286
Alternate	8	51		50

6.2 Scores on the Raven's SPM: Normative Sample

The scores of the Normative sample, those subjects presented the translated and adapted instructions, on the Raven's SPM are presented, first in the form of a frequency distribution. In Table 4 the mean score, standard deviation, range, minimum and maximum scores are presented.

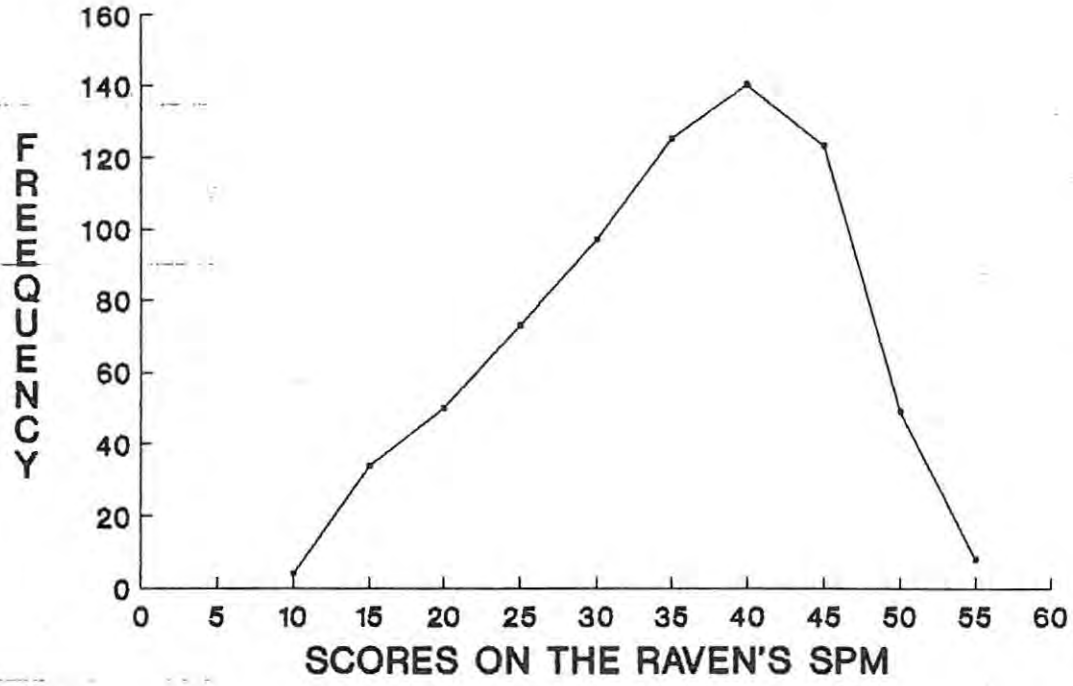

Figure 1. Graph showing the frequency distribution of the scores on the Raven's SPM: Normative sample.

The above graph begins with a sharp increase in the frequency of scores from the lowest score of 10, followed by a decline and then a gradual increase in the frequency of the scores reaching the maximum frequency at score 42. There is a sharp decline in the frequency of the scores to the highest score of 54.

Table 4

Mean Score, Standard Deviation, Range, Minimum and Maximum Scores for the Normative Sample

	Frequency	Mean	Standard Deviation	Range	Minimum Score	Maximum Score
Total	711	32.9	9.72	43	11	54

A mean score of 32.9 was obtained from the 711 pupils in the Normative sample on the Raven's SPM, with a standard deviation of 9.72.

6.2.1 Comparison of the scores across schools.

In this section the scores obtained by the Normative sample on the Raven's SPM, are compared across the three schools. In addition to this the t-test results for checks of significance between the schools, are presented in Table 6. The significant differences found should be treated with caution at this stage, as any difference may be the result of one or more of the other variables interacting with the variable in question.

Table 5

Mean Scores, Standard Deviation, Range, Minimum and Maximum Scores across the Three Schools for the Normative Sample

School	Frequency	Mean	Standard Deviation	Range	Minimum Score	Maximum Score
Nyaluza	286	32.9	10.0	43	11	54
Ntsika	183	32.4	9.6	39	12	51
Nombulelo	242	33.3	9.4	41	11	52

The means, range and standard deviation are similar for each school. This is supported by an analysis of the t-Test results (see Table 6 below), where no significant differences at the one percent level were found.

Table 6

The t-Test Results for the Comparison Across Schools

School	t-value	d.f.	Significance
Nombulelo/Ntsika	0.99	423	p = 0.3229
Nombulelo/Nathaniel	0.59	526	p = 0.5552
Ntsika/Nathaniel	-0.45	467	p = 0.6545

6.2.2 Comparison of the scores across standard.

In Table 7 the scores on the Raven's SPM are presented for the Normative samples in terms of the school standards of respondents.

Table 7

Mean Scores, Standard Deviation, Range, Minimum and Maximum Scores of the Standards for the Normative Sample

Standard	Frequency	Mean	Standard Deviation	Range	Minimum Score	Maximum Score
Six	153	28.5	10.3	41	11	52
Seven	169	33.5	9.5	39	12	51
Eight	107	32.8	9.7	39	12	51
Nine	188	34.8	9.3	43	11	54
Ten	94	35.1	7.8	40	11	51

The mean scores of the Normative sample on the Raven's SPM, increases from Standard Six, with a mean score of 28.5, to Standard Ten, with a mean score of 35.1 except for Standard Eight. The mean score of the Standard Eight subjects, was 32.8. This is lower than the mean score of the Standard Seven subjects.

The standard deviation decreases with a rise in the standard, suggesting a greater conformity to the mean.

Significant t-Test results on the Raven's SPM between the standards for the Normative group, are presented in Table 8.

Table 8

Significant t-Test Results for the Comparison Across Standards

Standards	t-value	d.f.	Significance
Six/Seven	-4.55	320	p = 0.0000
Six/Eight	-3.43	258	p = 0.0007
Six/Nine	-5.92	339	p = 0.0000
Six/Ten	-5.31	245	p = 0.0000

The probability values indicate a significant difference between Standard Six and all the other standards.

6.2.3 Comparison of scores across gender.

In Table 9 the scores on the Raven's SPM are presented for the Normative sample in terms of the gender of the respondents.

Table 9

Mean Scores, Standard Deviation, Range, Minimum and Maximum Scores of the Male and Female Respondents for the Normative Sample

Gender	Frequency	Mean	Standard Deviation	Range	Minimum Score	Maximum Score
Male	308	34.3	9.5	43	11	54
Female	403	31.8	9.7	41	11	52

The mean scores differ between the genders' by 2.6 points, with the standard deviation being similar. The range for the female respondents is smaller than for the males as a result of the maximum score for the female subjects being two points lower.

Table 10

The t-Test Results for the Comparison Across Gender

Gender	t-value	d.f.	Significance
Male/Female	3.46	709	p = 0.0006

The results obtained from the above t-Test (Table 10) indicate a significant difference between males and females at the one per cent level. Males score significantly higher than females. It should be noted that the male respondents are significantly older than the female respondents.

6.2.4 Comparison of the scores across age.

The mean score for all the ages are presented in Table 11 (page 79) with the reduced sample, ages 17 to 22, printed in bold. The number of subjects sampled at the lower and upper ends of the ages were low. For this reason the sample was reduced to include pupils of the ages 17 to 22 only.

Looking at the reduced sample mentioned earlier (17 to 22 year old groups), a gradual increase in the mean scores of the Normative sample, on the Raven's SPM, is noted. A mean score of 31.8 for 17 year old subjects, and a maximum mean score of 35.0 for 21 year old subjects. From here the mean score decreases gradually till the 24 year old subjects, who obtained a mean score of 29.6.

The age distribution of the sample within each standard is inconsistent. The subjects ages in any one standard varied between 12 and 27. In developed countries where there is universal compulsory education a standard is dominated by one age group, for example, Standard Six classes are usually made up of 13 year old children. In this sample, and in "black" schools in general, this does not occur and there is, therefore, a wide distribution of the ages across the standards. This makes the interpretation of the above data difficult as there may be interaction effects that can account for the differences found. For this reason, an analysis of covariance was conducted to investigate the relationship between various individual

variables on the dependent variable - the scores on the Raven's (SPM).

Table 11

Mean Scores, Standard Deviation, Range, Minimum and Maximum Scores of the Ages for the Normative Sample

Age	Frequency	Mean	Standard Deviation	Range	Minimum Score	Maximum Score
12	1	34.0	-	-	-	-
13	9	35.0	8.9	29	17	46
14	8	38.7	11.1	34	17	51
15	25	34.6	7.9	33	16	49
16	62	34.0	9.8	37	15	52
17	**80**	**31.8**	**10.4**	**40**	**11**	**51**
18	**79**	**32.1**	**10.2**	**38**	**12**	**50**
19	**138**	**32.3**	**10.4**	**40**	**11**	**51**
20	**99**	**34.2**	**9.2**	**41**	**13**	**54**
21	**84**	**35.0**	**9.2**	**40**	**12**	**52**
22	**52**	**32.2**	**8.2**	**34**	**15**	**49**
23	33	30.3	9.1	36	11	47
24	22	29.6	9.7	29	17	46
25	11	31.7	10.2	32	14	46
26	7	27.0	5.7	14	19	33
27	1	41.0	-	-	-	-

6.2.5 Analysis of covariance: Age as covariate.

The sample used for the analysis of covariance was reduced from the total sample and only included ages 17 to 22. The number of subjects used for the analysis of covariance was 532, with the largest number of subjects falling into the 19 year old age group. The first analysis of covariance, used the total scores on the Raven's of the Normative subjects as the dependent variable, the variables of gender and standard as the group variables; and the the age variable as the covariate. Age was used as the covariate so as to check whether there was a significant difference between the standards, and between male and female respondents, negating any effect that age may have on these variables. The results are presented in Table 12.

Table 12

Analysis of Covariance with Age as the Covariate and Standard and Gender as the Group Variables

Variable	Degrees of Freedom	F Value	Significance
Standard	4	25.56	p = 0.0000
Gender	1	21.33	p = 0.0000
Standard/Gender	4	0.61	p = 0.6569

The analysis of covariance of the test score results of the Normative sample, with age as the covariate demonstrated that there was a significant difference between the different standards as well as a significant difference between the male and female scores. No

interaction effect between the variable of standard and gender was found.

6.2.6 Analysis of covariance: Standard as covariate.

To check that this significant difference between the ages is not a function of education, an analysis of covariance (ages 17 to 22) was repeated, using the total scores on the Raven's of the Normative subjects as the dependent variable, the variables of age, gender, and school as the group variables; and the standard variable as the covariate. Standard was used as the covariate so as to check if their was a significant difference between the different ages, schools and genders. The results are presented in Table 13.

Table 13

Analysis of Covariance with Standard as the Covariate and Age, Gender, and School as the Group Variables

Variable	Degrees of Freedom	F Value	Significance
Age	5	3.67	p = 0.0028
Gender	1	15.75	p = 0.0001
School	2	0.87	p = 0.4207
Age/Gender	5	0.63	p = 0.6738
Age/School	10	0.60	p = 0.8123
Gender/School	2	1.32	p = 0.2684
Age/Gender/School	10	1.13	p = 0.3382

The analysis of covariance of the test score results of the Normative sample, with standard as the covariate demonstrated that there was a significant difference between the ages as well as a significant difference between the male and female scores. No significant difference between the schools was found. No interaction effect between the variables of age, gender, and school was found.

6.3 Scores on the Raven's SPM: Alternate Sample

6.3.1 Description of the Alternate sample.

The scores of the Alternate sample, those subjects presented the translated instructions, on the Ravens SPM, will be presented in a frequency distribution (Figure 2), as well as the mean score, standard deviation, range, minimum and maximum scores being presented in table form (Table 14). The Alternate sample was chosen from the population of Standard Eight classes at Nathaniel Nyaluza High School and Nombulelo Senior Secondary School.

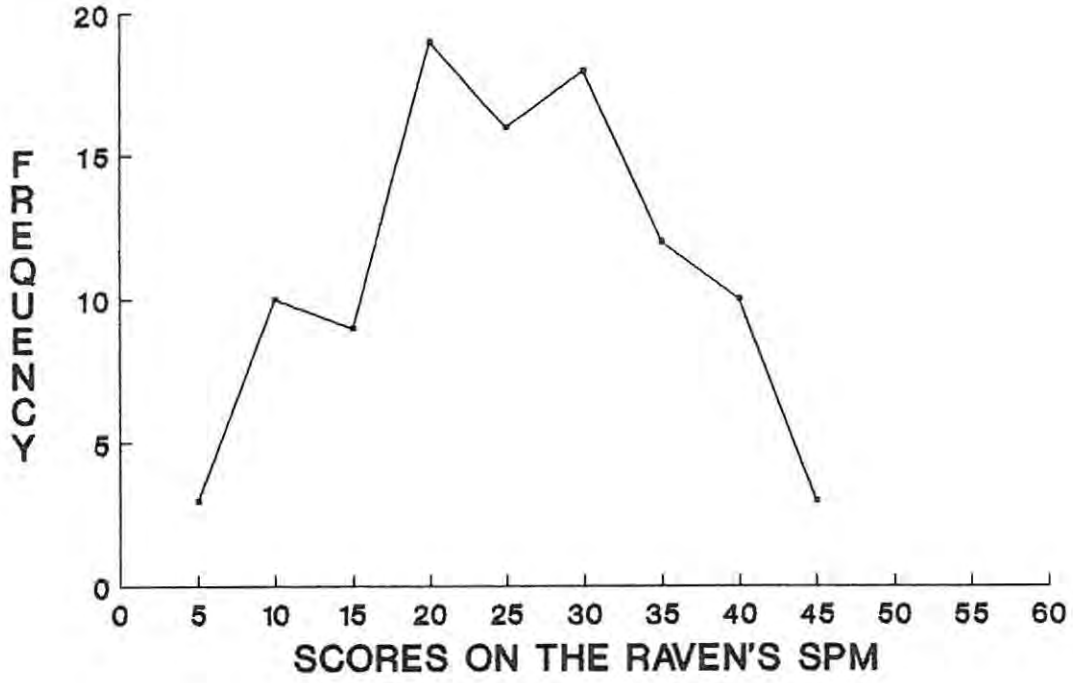

Figure 2. Graph showing the frequency distribution of the scores on the Raven's SPM: Alternate sample.

This graph does not represent a normal distribution of the scores on the Raven's SPM by the Alternate sample. The graph does not follow any particular pattern. The reason for this may be the small size of the sample.

The mean score for the Alternate sample is presented in Table 14.

Table 14

Mean Score, Standard Deviation, Range, Minimum and Maximum Scores for the Alternate Sample

	Frequency	Mean	Standard Deviation	Range	Minimum Score	Maximum Score
Total	101	28.5	10.1	40	7	47

6.3.2 Comparison of scores across gender.

In this section the scores obtained by the Alternate sample on the Raven's SPM, will be compared across gender for descriptive purposes. In addition to this the t-Test results, for check of significance between the male and female subjects will be presented.

Table 15

Mean Scores, Standard Deviation, Range, Minimum and Maximum Scores of the Male and Female Subjects for the Alternate Sample

Gender	Frequency	Mean	Standard Deviation	Range	Minimum Score	Maximum Score
Male	44	29.1 *	10.0	40	7	47
Female	57	28.0	10.2	39	8	47

* $t = 0.56$, $p = 0.5772$, d.f. = 99.

The mean scores of the gender samples differ by 1.1 points, with the standard deviation and range being similar. This is supported by an analysis of the t-Test result; no significant difference between the means was found.

6.4 Comparison of Scores of the Normative and Alternate Sample

In this section the Normative sample, those subjects presented the translated and adapted version of the Raven's SPM will be compared with the Alternate sample, those subjects presented the translated version of the Raven's SPM, on their total score obtained on the test. This comparison will determine whether the method of presentation will effect the subjects understanding of the task, and thereby their scores on the Raven's.

Table 16

Comparing Mean Score, Standard Deviation, Range, Minimum and Maximum Scores of the Normative and Alternate Sample

Sample	Frequency	Mean	Standard Deviation	Range	Minimum Score	Maximum Score
Normative Total	711	32.9*	9.7	43	11	54
Alternate Total	101	28.5	10.1	40	7	47

* t = 4.24, p value = 0.0000, d.f. = 810.

While the standard deviation of the scores may be similar, the mean scores and range of the two samples do differ. The range of the Normative sample is larger, with the minimum and maximum scores being greater than the Alternate sample. These findings are supported by an analysis of the t-Test results, where a significant difference between the two samples was found at the one per cent level.

There is a problem of interpreting the apparent difference between the two samples. This is as a result of the samples differing in terms of a number of variables. Subjects from the Alternate sample were only chosen from the Standard Eight classes of Nathaniel Nyaluza High School and Nombulelo Senior Secondary School. For this reason a decision was made to match the samples. (Cf. Section 5.5.1 for an explanation as to how the samples were matched.)

6.4.1 Comparison of the matched samples.

The mean scores, standard deviations, range, minimum and maximum scores of the matched Alternate and Normative samples are presented in Table 17.

Table 17

Comparing Mean Score, Standard Deviation, Range, Minimum and Maximum Scores of the Matched Normative and Alternate Samples

Sample	Frequency	Mean	Standard Deviation	Range	Minimum Score	Maximum Score
Normative Matched	45	35.1*	9.9	37	14	51
Alternate Matched	45	28.9	10.5	40	7	47

* t = 2,83; p value = 0,0057; d.f. = 88.

The Normative sample scored higher than the alternate sample. This finding is supported by an analysis of the t-Test results where a significant difference is accepted at the one per cent level.

6.5 Cross tabulation: Gender, age and standard.

The mean scores of male subjects broken down by age and standard are presented in Table 18. Means are provided for categories with three or more subjects in it. Categories with only one or two subjects were collapsed into the nearest category with the least number of respondents. This occurred mostly at the upper and lower limits of the categories.

Table 18

Mean Scores of Male Subjects broken down by Age and Standard: Normative Sample

Age	Standard				
	6	7	8	9	10
13	28				
14	36				
15	35	43			
16	37	41			
17	29	38			
18	28	33	41	35	
19	28	36	30	37	
20	31	30	32	39	38
21	28	28	41	36	40
22		27	34	37	33
23				29	36
24				33	32
25				28	34
Mean	**31**	**35**	**35**	**36**	**36**

No particular pattern is exhibited in the above table except that the mean scores for Standards Seven\Eight and Standards Nine\Ten are the same. For this reason when establishing norm tables it may make sense to collapse Standards Seven and Eight and Standards Nine and Ten into two groups. On the basis of the number of respondents in each age

category broken down by standard, the group of respondents were collapsed into the following age categories: 12 to 16 year olds, 16 to 19 year olds, 20 to 23 year olds, and 24 to 27 year olds.

The mean scores of female subjects broken down by age and standard are presented in Table 19. Findings similar to males re the mean scores were found, therefore when establishing norm tables it may be preferable to collapse Standards Seven/Eight and Standards Nine/Ten into two groups. On the basis of the number of respondents in each age category broken down by standard, the group of respondents were collapsed into the following age categories: 12 to 16 year olds, 16 to 19 year olds, 20 to 23 year olds, and 24 to 27 year olds.

Table 19

Mean Scores of Female Subjects broken down by Age and Standard: Normative Sample

Age	Standard				
	6	7	8	9	10
14	39				
15	29	40			
16	29	35	30		
17	22	33	36	43	
18	31	32	35	33	
19	21	30	32	36	39
20	18	30		37	38
21	20	28	31	35	31
22			33	28	37
23			21	31	28
24				29	33
25				28	30
26					30
Mean	**27**	**32**	**32**	**34**	**34**

Mean scores and norms established for the collapsed categories are presented in Chapter 8.

CHAPTER 7 - DISCUSSION OF THE RESULTS

Drawing on the obtained results, the discussion of the nature of intelligence, its measurement, and the Raven's Progressive Matrices (RPM), the following will be discussed in this chapter.

(a) The effect of age, gender, and education on scores of the Raven's SPM in this study, will be compared to Raven et al. (1983), and Raven (1989).

(b) The effect of the two methods of test presentation on the pupils' test performance, and the differences of this study with other similar investigations.

(c) The concept of 'disadvantage'.

(d) The norms for the population under consideration.

7.1 The Effect of Age, Gender, and Education on scores of the Raven's SPM in this study, compared to Raven et al. (1983) and Raven (1989)

Raven et al. (1983) found that:

(a) there was no gender difference in the scores on the Raven's SPM,

(b) the test worked in a similar way for children from different socio-economic backgrounds,

(c) there was considerable stability of performance within and between western societies, and

(d) a person's total score provides an index of his/her intellectual capacity with little influence from his/her

education.

Raven et al's (1983) results have been rejected by Tulkin and Newbrough (1968), who found that race was important in distinguishing performance in the low class groups, and gender differences were found to be more important in Negro groups. Vincent and Cox (1974) found differences in scores on the RPM between racial groups based on the level of education obtained. Do these findings follow from the data from this study?

7.1.1 The effect of age on the score of the Raven's SPM.

In the analysis of covariance using educational standard as the covariate and age, gender, and school as the group variables (Table 13), a significant difference between the ages was found, with there being no interaction effects with educational standard. In Table 11, there is a trend for scores on the Raven's SPM to increase with age up to age 21 whereupon, they decrease. The result of the analysis of covariance and the trend noted in Table 11, contradict the finding of Raven (1941) and Raven (1989), who argued that eductive ability reaches its peak at the age of thirteen and a half (Raven, 1941), or at the age of eleven and a half (Raven, 1989). The results of this study raise questions about Raven's assumption of eductive ability peaking at the prescribed ages.

What has caused this delay of eductive ability peaking in the scores of the sample in this study? The population tested may be distinguished sharply from other populations by the nature of the educational standard that is offered to the population. It seems that the maturation process is delayed in societies where the educational system is of the nature as described below.

The nature of the educational system of the population under study, may be described in the following way. The physical conditions at the schools are not ideal, where the lack of facilities lead to overcrowding of classes, with at times more than 55 children being taught in a class that is able to accommodate 40 children. The result being that children share desks and other utilities that are essential for the educational system to run successfully. In addition to the lack of classrooms there is a shortage of textbooks for the pupils, thus making the education of these students extremely difficult.

In addition to the physical conditions of the schools, the political climate prevailing in the country, has had an effect on the attitude of children towards education. The education system, by its nature, has been highly politicized with the result that at times the education of the youth has not been able to operate at a level that it should have. The stance of the government towards 'Bantu Education', and the differing facilities available to the races in this country has not allowed for the natural selection of the top students in the country. The inferior education offered to the African pupils in this country has not allowed for the natural development of talents. Thus

politicizing the educational process in the country, and it is only with recent political changes that there has been an attempt from all sides of the political arena for the depoliticization of the educational process in this country.

What is the cause of the drop of scores from age 22? This may be explained by a self selection process operating in the population of Secondary School African Pupils in Grahamstown. The older children are less intelligent, but have managed to stay or be pushed through the education system. In an education system where natural selection is able to operate these older students would have dropped out of the educational system at a younger age as they do not have the ability to complete school, but by the nature of the politicized education system, these older students have managed to stay in the system to attempt to complete their education. The presence of these older students in the education system also hinders the development of the other students. The education system is slowed down to the disadvantage of these pupils.

7.1.2 The effect of educational level attained on the score of the Raven's SPM.

The analysis of covariance using age as the covariate and gender, and standard as the group variables (Table 12), found a significant difference between the standards, with this not being interactive with the ages of the students. In Table 7 there is a trend for the scores to increase with an increase in standard, except for the Standard Eight subjects having a greater mean score than the Standard Seven

subjects. The t-Test results also indicate a significant difference between the Standard Six Classes and all the other standards (Table 8). Thus education does seem to have an effect on the scores of the Raven's SPM.

This contradicts the finding of Raven et al. (1983), that the SPM, is not effected by the educational level attained at the school, but rather supports the finding of Vincent and Cox (1974), that educational level does effect the score of Negro respondents on the RPM. Thus the cultural fairness of Raven's SPM must be questioned, as something about schooling has an effect on the scores. What aspect of schooling has an effect is not clear. Is it that schooling increases understanding of test demands, or does it affect the basic ability, that is, the eductive ability of the respondents. The reason for this difference, may be that there is something present in Western societies that makes schooling unimportant. To answer these questions adequately further research needs to be carried out.

Comparing the educational systems in Western societies with the educational system present in the population being researched, all the points mentioned in section 7.1.1, about the education system that is present in the population under study, must be taken into account. Is it possible that the poor educational environment affects the development of the students eductive ability. How this educational process affects the development of the eductive ability may not be answered from the results of the current research, but further research may lead to an answer of this. The findings indicate that

whilst testing respondents in a poor educational position in a multicultural society, educational level attained does affect the score obtained on the the Raven's SPM, with the respondents in higher standards scoring significantly higher.

7.1.3 The effect of gender on the score of the Raven's SPM.

The analyses of covariances, using age (Table 12) and educational standard (Table 13) as the covariates, both indicate that there is a gender difference, with females scoring significantly lower than their male counterparts. This difference is not a function of age or education. The female respondents score 2.6 points lower than the male subjects (Table 9). The findings of the t-Test (Table 10), also found a significant difference between the male and female respondents with the female respondents scoring significantly lower.

This finding seems to support the finding of Tulkin and Newbrough (1968), that gender differences are prominent in Negro groups. In terms of this research, we may conclude that gender differences on the Raven's SPM do seem to be important when testing a culture in a multicultural society that has been exposed to a poor education system as described in section 7.1.1.

7.1.4 The effect of school on the score of the Raven's SPM.

The t-Test results and the analysis of covariance results show no significant difference in scores on the Raven's SPM across the three schools; Nombulelo, Ntsika and Nathaniel Nyaluza High Schools. This is so even though the conditions between the three schools varied. Following the suggestion of Crawford-Nutt (1976), that it was the standard of the educational level and the schools attitude towards its education, that contributed towards the means of 45.1 and 44.9 by African pupils scored on the Raven's SPM in his study, it was hypothesized that Nombulelo's pupils would score higher than pupils from the other two schools, as a result of an apparent stronger focus on education at Nombulelo Senior Secondary School.

Nombulelo Senior Secondary School is a relatively new school compared to the other two schools, with the emphasis being on education. The evidence for this being Nombulelo's attitude towards my being allowed to test only after school hours so as not to disrupt the educational process. This school also provided the pupils with classes in the afternoon so as to ensure that the pupils could do their homework under the supervision of the teacher. The buildings of Nombulelo Senior Secondary School were in a better condition than the other two. The mean score of Nombulelo was higher than the other two schools however it was not significantly higher. Thus the difference in conditions between the three schools may be considered to be negligible or unimportant.

Further research into the differences of school conditions will enable an explanation of the importance of schooling conditions on the effect on tests similar to the Raven's SPM. It will also provide researchers with the answer as to whether poor quality education is able to be overcome in tests of intelligence through adapting the test instructions in such a manner that will ensure that the respondents understand what is expected of them.

7.1.5 The relative effect of age, gender, education, and type of school.

What do these effects of age, education, gender, and school mean for the Raven's SPM? The fact that no inter-action effect was found between the variables whilst conducting the analysis of covariances, supports the findings even more. These findings work strongly against Ravens' view that the Raven's SPM is a culturally fair test. In this research, a sample of students that have been exposed to a poor education system were assessed and compared to other groups. This group may be described as 'disadvantaged' in terms of its accessibility to an education system similar to the other groupings in the multi-cultural society of South Africa. If the term 'disadvantaged' is investigated (see section 7.3), we may come to a greater understanding of the cultural fairness, and the effect of test presentation, on the scores of the Raven's SPM.

7.2 The Method of Test Presentation

When comparing the Normative and Alternate mean scores, the comparison of the total samples (Table 16), and the comparison of the matched samples (Table 17), t-Test results indicate the Normative sample scored significantly higher than the Alternate sample. In the matched sample comparison, the Normative sample (mean score of 35.1), scored 6.2 points higher than the Alternate sample (mean score of 28.9). The minimum and maximum scores for the Normative matched sample are greater than that for the Alternate matched sample. The above indicates that the method of test presentation does have a significant effect on the pupils' scores on the Raven's SPM.

This significant difference may be explained in the following way. The Alternate sample subjects, of this study were in a disadvantaged position, not only because they were unfamiliar with the particular test being used, but because they were less familiar with the concepts used in explaining what was expected of them, than the Normative sample. The instructions were simply translated from the original English into Xhosa, and respondents were not able to familiarize themselves with the key concepts. The method of test presentation, for the Normative sample, however, did attempt to familiarize the subjects with these concepts before the subjects began the test. This was attained through the ideas of Schwarz (1963) as discussed in Chapter 3. This helped ensure that the pupils understood what was expected of them in the test.

The alterations to the test instructions helped the respondents understand what was expected of them in the test situation. The results indicate that students from backgrounds that may be classed as 'disadvantaged', when given the instructions for the Raven's SPM in a manner that ensures they understand what is expected of them in the test situation, then their scores on this test will be significantly higher than if this factor is not taken into account.

What does this finding do to the status of the Raven's SPM being a culturally-fair test? While the test does try to minimize the parameters of **language**, **reading**, **speed**, and **test content**, parameters believed to help make a test culturally fair, there is a significant difference in scores between the two samples.

While the Raven's SPM, has been considered to be a culturally fair test, this assumption must be doubted. Thus the scores on the Raven's SPM may not be taken at face value whilst comparing the scores across cultures. Care must be taken when making comparisons across cultures, and it is the opinion of this author that no such comparisons should be made until such a time that the 'disadvantaged' may be assumed to be in a similar position as the 'advantaged' in the testing situation.

It is the belief of this author that the Raven's SPM lacks in the quality mentioned by Schwarz (1961), namely that the testing procedure should teach the testee every response that they may be expected to make in the test situation. No pre-supposition of any response should be made. Raven presumes that respondents will respond to the items in

his test in a standard manner. This study shows that this is not an accurate assumption.

Raven seems to have gone out of his way to create a test with test items where the actual content of the test may be considered to be **culturally fair**, yet ignored the importance of how the test is to be presented. The method of presentation does consider the parameter of **reading**, yet largely ignores the parameter of **language**. Raven makes the language simple yet does not negotiate the language parameter totally. In support of Raven however, the simplicity of the test instructions does allow for the instructions to be translated into the language of the culture that is being tested. These provisions made by Raven (1938), however do not seem to have been enough. The results presented in this research indicate that the Raven's SPM is not applicable to all cultures unless it is accompanied by adaptations to the test instructions.

7.2.1 Comparison of results with Pons(1974) and Crawford-Nutt (1976).

The findings presented in this research support the finding of Crawford-Nutt (1976), that the method of test administration is important for the debate on differences in intelligence test scores between blacks and whites. The author of this thesis, however extends this finding to include those respondents to the Raven's SPM that may be classed as disadvantaged. The results obtained however do indicate discrepancies with similar research.

Table 20

Comparison of Mean Scores on the Raven's SPM of Crawford-Nutt (1976), Pons (1974) and Current Research

Research	Mean
Pons (1974)	33,6
Crawford-Nutt (1976)	45.1
	44.9
Current Research (normative unmatched)	32.9
(normative matched)	35.1
(alternate matched)	28.9

The mean score for the Normative matched sample (35.1), is considerably lower than the mean score for the sample of African and white pupils tested of a similar age group by Crawford-Nutt in 1976: black pupils - 45.1 and 44.9 and whites - 44. However the mean score obtained for a group of Bemba adult males that Pons (1974) tested between 1962 and 1967 using his newly developed method of test presentation (33.66) is similar to the mean score of Normative matched group (35.1) as well as the total Normative sample (32.9). Pons' group of Bemba males had all attained educational levels of Standards' Seven to Nine.

Why did Crawford-Nutt's sample score ten mean points higher than the Normative matched sample and 11.5 mean scores higher than Pons' sample? The reasons for this may be found in the Crawford-Nutt (1976) study, which is discussed in detail in Chapter 4. Generally he cited the level of education, a reputed educational institution, and the subjects' degree of test sophistication as possible reasons, for the higher mean scores. Thus the respondents were sufficiently familiar with tests, test materials, and test situations not to suffer from **content** and **context uncertainty** (Pons 1974, cited in Crawford-Nutt 1976).

Another reason for this may be that the mean score of the subjects, for the current research, included students from all standards in the school and not only matriculants (twelth year of schooling). But when looking at the mean score for the Normative matric sample (twelth year of schooling) (Table 7) (35.1), we see that a great discrepancy between the studies still exist. Thus this reason does not account for the discrepancy.

While the reasons cited by Crawford-Nutt (1976) may be valid, the motivation of subjects may also be an underlying factor. While the method of test presentation in the current research attempted to increase the motivation of the testees to perform their best, by allowing active participation in the instruction phase of the test, it is possible that the instructions did not motivate the respondents sufficiently. The researcher's assumption that active participation would serve to motivate the respondents may have been incorrect. It is possible that the active participation only motivated a small sector

of the respondents, those respondents who felt confident enough. Further research into how to motivate respondents needs to be made. Thus the test instructions may not have been adequate to ensure that the pupils assessed, were sufficiently motivated to allow for comparisons with other groups.

While the researcher took into account the points made by Schwarz (1961), and in particular the suggestion that the most effective means of teaching the test is through the use of visual aids, supplemented by active demonstration, the design of the presentation of instructions may have still been lacking.

The Raven's SPM is divided into five sets of 12 problems each, with each set based on a particular principle. See Chapter 4 for a discussion of this. The translated/adapted instructions only take into account the first principle in Set A of the Raven's SPM. There is a change of principle from Set A, that uses the principle of completing a pattern in a continuous figure, to Set B, that uses the principle of a figural analogy in a two-by-two matrix. The translated/adapted instructions teach the respondents how to solve the first principle and ignore the second, third, fourth, and fifth principles that follow. Thus while the researcher assumed that the respondents understood how to respond to the second, third, fourth, and fifth sets this may not have been accurate.

While Raven et al. (1983), believe that the order in which the items are presented provide training in the method of working, the initial items of each set still presume a standard response. Further research into the presentation of the Raven's SPM may make allowance for the above, and include instructions at the start of each new set, thereby ensuring that each new principle is explained. In this way no standard response is presumed. The researcher would expect that if instructions were presented to the respondents of the test at the start of each set, using the first item in each set, or developing a new item for the presentation, the respondents would have scored higher on the Raven's SPM.

The difference in the mean scores cited in the three studies indicates the urgency for more research to be carried out into the area of test presentation and the reasons for difference in black and white scores on intelligence tests.

The importance of **test content** has been the focus of most research of cross-cultural testing. There has been a shift towards the issue of **test presentation** in the last decade, but a great deal more research needs to be undertaken before a solution to the problem of cross-cultural testing will be found. **Test content** and **test presentation** are equally important and to ignore the one will not advance a solution to the dilemma of cross-cultural testing. Only once fair **test content** and **test presentation** has been attained may we say that a test may be considered to be totally **culturally fair.**

7.3 The Nature of Disadvantage in Psychometric Testing

Disadvantage, is a relatively new term in psychology. The term has been used to explain many sociological problems, and only recently encountered psychology in the South African context. Psychology has as yet not been able to define the term psychologically. In terms of this discussion 'disadvantage' refers to the unequal access of students to education as a result of the oppression of black people in South Africa.

At all times in this discussion, the term does not refer to any biological deficit, that theorists like Jensen (1969) may like to propose. This author assumes that the development of cognitive structures is invariant across social and environmental conditions, and thus differential intellectual performance between races, may not be attributed to biological differences. However the cultural relativists stance on disadvantage can not be ignored, as it recognizes the social construction of cognition and learning (Moll and Slonimsky, 1989).

Miller (1989) explains the combination of the universalisitic and cultural relativists position through the "Zone of Human Potential" (Miller, 1989 p. 158). Universal human competence exists, and this may be regarded as the optimal performance that a human being may attain, which is accessible to all. However performance demands are also imposed on the individual by the tasks needed to be carried out and in the particular period that the task needs to be carried out.

Thus the ability to achieve universal human competence exists, but the attainment of this is affected by the task demand placed on the individual.

Thus one can not only look at 'disadvantage' in terms of psychometric testing in South Africa in absolute nature and nurture terms. It is an interaction of the two.

The term disadvantage in relation to the population under consideration in this project, thus assumes that all the students have the essential cognitive competence, but it is in the education process that they have been exposed to, that the task demands and historical context does not allow the cognitive competence to develop - thereby placing the students in a disadvantaged position, as they are not able to complete the processing that is necessary for the particular task of the test being presented. Gellatly (1987), quoted in Miller (1989), states that the competencies and performances which normally develop spontaneously during schooling have been denied most black students in South Africa.

In reality students are exposed to competition in situations were the context and the content of a task presented them may be unfamiliar. This is the problem that the disadvantaged student is facing. The disadvantaged student is being tested with tests that have not been established with the conditions of disadvantage that they are facing in mind. The population of Secondary School African pupils in Grahamstown is disadvantaged, within the context of South Africa, in that their socio-economic background has not allowed for the

development of their competence for the tasks required by the Raven's SPM.

The conditions described above, lead to the group being disadvantaged because; (a) the education process does not allow for the development of "deep processing" (Moll et al., 1989, p. 161) competence and (b) the conditions at the schools do not allow for the development of their "deep processing" (Moll et al., 1989, p. 161) competence.

Thus there exists an inequality of access to the concepts that are used to introduce, present, and complete the task demands of a test. This inequality of the accessibility to the concepts is the reason for the poor performance of many black students in South Africa on tests of intellectual ability.

More research needs to be carried out into the concept of **disadvantage**, and ways in which the disadvantaged respondents may be assessed fairly. At the moment it is not viable to compare disadvantaged groups with advantaged groups in a multi-cultural society. The assessment of people that fall into the category of disadvantaged, should have there own norms established and assessed within those norms. This has been the aim of this research.

The need to alter the instructions of the Raven's SPM in order to gain a better reflection of many pupils that may be classed as disadvantaged, is indicative of the fact that the test is not a culturally-fair test. The question, what is being tested, needs to be

asked. Is the eductive ability, as Raven (1989) purports, being tested, or is it simply the respondents ability to grasp what is expected from the test? It would be too harsh to say that the eductive ability that Raven et al. (1983) speak of is not being measured at all. The measurement of this eductive ability which the Raven's SPM has shown to measure over time though, is susceptible to affects of culture, and when applied to non-western cultures, care needs to be taken in the way in which the test is presented. This is so, because if the eductive ability of the respondent is being affected by the respondents inability to grasp what is expected of him/her in the test then a true reflection of the respondents ability has not been achieved. This inability to grasp what is expected of the respondent is not the result of the lack of certain cognitive competence, but rather the inability to utilize these skills because of the inequality of conditions that surround the education system has not allowed for the development of these task demands.

7.4 The Norms

The findings of this research contradict the findings of Raven et al. (1983) and Raven (1989) in many ways. Where Raven (1989) found no significant difference between ages above the age of eleven and half years, current data indicates otherwise. Raven's (1989) assumption that the Raven's SPM is not effected by the level of education is challenged, as is Raven's (1989) finding that there is no difference between male and female respondents. These findings require a set of norms for Grahamstown African Secondary School youth, that is different from Raven, and reflect age, education, and gender

differences that were found in this study. Norms are given in Chapter 8 and are established for males and females in Standard Six and Standard Seven/Eight and Standard Nine/Ten groups, for the age categories: 12 - 16 year olds', 17 - 19 year olds', 20 - 23 year olds', 24 - 27 year olds'.

Score points for the Raven's SPM are provided for the 95th, 90th, 75th, 50th, 25th, 10th, and 5th percentile points for these groupings.

This researcher decided to convert the raw scores to percentile points at yearly intervals. The decision to do this was done on the finding of Raven (1941), that median scores show slow development in eductive ability up to the age of eight years, a fairly rapid development on the ability between the years of eight and thirteen, and at the age of thirteen and a half years, eductive ability normally reaches its maximum development. Raven (1989), says that current research indicates that innate eductive ability reaches its maximum development at the age of eleven and a half years. The sample consisted mostly of respondents aged between 16 and 22, therefore it was decided that it would not be necessary to provide norms for half yearly intervals but only yearly intervals. The biographical details provided by the respondents did not allow for the data once obtained to be converted to percentile points at half yearly intervals which may have been necessary.

CHAPTER 8 - TESTING PROCEDURE AND NORMS FOR THE USE OF THE RAVEN'S SPM IN GRAHAMSTOWN

The aim of this chapter is to provide a testing manual for clinicians on how to use the Raven's SPM. It is essential that clinicians should use the Raven's SPM as specified whilst using the norms provided.

8.1 The Instrument

The Raven's SPM (1958) test book should be used. It differs from the Raven's SPM (1938) test book in that some of the test items are presented in a different order. In addition to this the record forms for the test are required.

The illustrated record form, Item A1, and Item A2 need to be reproduced three times the size of the original, and printed in black on a white background. Item A1 and Item A2 need to allow for the options open to the testee, at the bottom of the pattern, to be movable. These parts should be attached to the white background with an adhesive which allows the tester to move the parts easily into the open part in the pattern and then return it to the original position.

8.2 The Testing Procedure

The testing procedure should adhere to the following steps.

1. The language spoken at all times should be Xhosa, the respondents' home language.
2. The African tester introduces him/herself to the group of respondents by his/her first name.
3. The reasoning for the testing is then explained in detail.
4. Covered versions of the illustrated record form, Item A1, and Item A2 are placed on a board in front of the respondents.
5. Test books and record forms are handed out to the students, asking them not to open the test book until instructed to do so.
6. Respondents are asked if they have any questions and once all the questions are answered, the tester uncovers the record form on the board and asks the respondents to fill in their personal details on the record form.
7. At this point the covers are lifted from Item A1 and Item A2, and the tester proceeds with the instructions.

a) Adapted/translated instructions in Xhosa.

Tyhilani iincwadi zenu kwiphepha lokuqala. Ngolu hlobo. Phaya ngasentla ithi iSeti uA yaye ninale kholam apha, kwiphepha lenuleempendulo, yeyeSeti uA. Lo nguA1. Niyabona ukuba yintonina. Eli cala lingasentla yipatheni enendwana engekhoyo. Isuntswana eliyinto nganye elapha ngezantsi (yalatha into nganye enye emva kwenye) limile ngokufanelekileyo nqo ukuba lingene gingci kwisikhewu salo,

kodwaaliyigqibi ngokuphelyo ipatheni. UNombolo 1 (yalatha isuntswana ubuye walathe kwipatheni) yipatheni engachanekanga nje tu. (Khumbula ukususa amasuntswana uwalinnganisele kwezi khewu). Unombolo 2 naye akalunganga, uyangena tswatsa kwesi sikhewu, kodwa akayiyo ipateni efanelekilyo. UNombolo 3 naye akalunganga, uyangena tswatsa kwesi sikhewu, kodwa akayiyo ipateni efanelekileyo. Ngaba ujani yena unombolo 6? (Bayeke baphendule, bachaze isizathu). Ngeba yipatheni efanelekileyo (bonisa ukuba ipatheni leyo ifana nqwa naleyo ingasentla) kodwa ayide iphelele ngokuthe gingci. Bekani iminwe yenu kuleyo ilunge twatsa. Ewe, uNombolo 4 ngoyena yena. Ke ngoko impendulo ka-A1 ngu-4 - Bhalani 4 apha, ecaleni kukanombolo 1 kukholam A kumaphepha enu eempendulo. Tyhilani kwelinye iphepha ngku.

Lo nguA2. Niyabona ukuba yintonina. Eli cala lingasentla yipatheni enendwana engekhoyo. Isuntswana eliyinto nganye elapha ngezantsi (yalatha into nganye enye emva kwenye) limile ngokufanelekileyo nqo ukuba lingene gingci kwisikhewu salo, kodwa aliyigqibi ngokupheleyo ipatheni. Ngaba unjani yena unombolo 1? (Baueke baphendule, bachaze isizathu). Uyangena tswatsa kwisikhewu, kodwa akayiyo ipatheni efanelekileyo. Bekani iminwe yenu kuleyo ilunge tswatsa. Ewe, uNumbolo 5 ngoyena. Ke ngoko impendulo ka-A2 ngu-5 - Bhalani 5 apha, ecaleni kukanombolo 2 kukholam A kumaphepha enu eempendulo.

Kwiphepha ngalinye encwadini yenu kukho ipatheni enendwana engekhoyo. Kufuneka ke qho nigqibe ngokwenu ukuba leliphina kula masuntswana angezantsi elilelona lifanelekileyo ukugqibezela le patheni ingentla. Nakulifumana elona suntswana lifanelekileyo nothi niyibhale phantsi

inmombolo yalo kumaphepha enu eempendulo ecaleni kwenombolo yepatheni. Ningabhali kwiincwadana ezi. Zilula ekuqaleni ziye ziba nzima njengokuba nibheka phambili. Akukho gqakri. Ukuba nithathela ingqalelo indlela ezithi ezi zilula zihambe ngayo niya kuzifumana zinganeno ngobunzima ezilandelayo. Lingani ibe nye emva kwenye, ukusuka ekuqaleni ukuya esiphelweni sencwadi. Sebenzani ngesenu isantya. Ningabi nenizitsibayo. Ningabuyi nityhile kweli cala senlilggithile. Khangelani ke ukuba ningachana zibe ngaphina. Ningathatha ixesha elide ngangoko nifuna. Yenzani elandelayo.

Elifanelekileyo enyanisweni ngunombolo 1. Qaphelani ukuba nibhale inani u-1 eclani konombolo 3 kukholam A kwifom yenu. Qhubani nenjenjalo ngokwenu nide noyokufikelela esiphelweni sencwadi. Ndiza kubuya ndijikele ndibone ukuba niqhuba kakuhle na.

b) Translated/adapted instructions in English.

Open your books to the first page. It is like this. At the top it says Set A and you have a column here, on your answer sheet, for Set A. This is A1. You see what it is. The upper part is a pattern with a bit missing. Each of these pieces below (point to each in turn) is the right shape to fit the space, but they do not all complete the pattern. Number 1 (point to the piece and then to the pattern) is quite the wrong pattern. (Remember to move the piece into the space of the broken pattern) Number 2 is also wrong, it fits the space, but does not complete the pattern. Number 3 is also wrong, it fits the space, but does not complete the pattern. What about number 6? (Let them answer, and explain the reason) Is it the right pattern

(illustrate that the pattern is the same as the pattern above) but it does not go all over. Put your finger on the one that is quite right. Yes number 4 is the right one. So the answer to A1 is 4 - write 4 here against number 1 in column A on your answer sheet. Turn to the next page.

This is A2. You see what it is. The upper part is a pattern with a bit missing. Each of these pieces below (points to each in turn) is the right shape to fit the space, but they do not all complete the pattern. What about number 1? (Let them answer, and explain the reason) It fits the space but it does not complete the pattern. Put your finger on the one that is quite right. Yes number 5 is the right one. So the answer to A2 is 5 - write 5 here, against number 2 in column on your answer sheet.

On every page in your book there is a pattern with a bit missing. You have to decide each time which of the pieces below is the right one to complete the pattern above. When you have found the right bit you write its number down on your answer sheet against the number of the pattern. Do not write on the booklets. They are simple at the beginning and get harder as you go on. There is no catch. If you pay attention to the way the easy ones go you will find the later ones less difficult. Try each in turn, from the beginning right to the end of the book. Work at your own pace. Do not miss any out. Do not turn back. See how many you can get right. You can have as much time as you like. Do the next one?

The right one of course is number 1. See that you have written the figure 1 against number 3 in Column A on your form. Go on like that by yourself until you get to the end of the book. I will come round to see that you are getting on all right.

8. The tester asks the respondents to please put their hands up once they have completed the test.
9. As the children put their hands up the tester marks on the record form, the time that the respondent completed the test.
10. Once all the respondents have completed the test, the tester takes in all the record forms and thanks the respondents for there co-operation.

8.3 The Norms.

The tester marks the test by awarding one point for every correct answer and no point for an incorrect answer. From this a total score of a possible 60 is obtained. The obtained score is then compared with the scores of the relevant table, and the respondent assessed by placing him/her within one of the percentile categories. The method of converting the raw scores to percentile points did not differ from the method of Raven et al. (1983) (Cf. Section 4.2.).

The mean scores and norms for the age groups in Standard Six are presented in Table 21.

Table 21

Standard 6 Smoothed Summary Norms

Gender	Percentile Point	Age Groups		
		12 - 16	17 - 19	20 - 23
Male	95	51	42	44
	90	47	41	42
	75	43	35	38
	50	34	30	30
	25	27	21	18
	10	22	16	14
	5	17	14	13
Mean Score		**35**	**28**	**30**
Female	95	43	45	33
	90	40	41	25
	75	37	32	22
	50	32	21	17
	25	25	16	13
	10	18	13	12
	5	15	12	11
Mean Score		**31**	**25**	**19**

The mean scores and norms for the Standard Seven/Eight age groups are presented in Table 22.

Table 22

Standard 7/8 Smoothed Summary Norms

Gender	Percentile Point	Age Groups		
		12 - 16	17 - 19	20 - 23
Male	95	51	47	45
	90	50	45	44
	75	49	42	42
	50	46	36	33
	25	27	29	24
	10	25	21	19
	5	23	13	18
Mean Score		**42**	**35**	**33**
Female	95	48	47	41
	90	46	44	40
	75	42	40	34
	50	36	34	28
	25	31	24	23
	10	25	18	17
	5	18	14	15
Mean Score		**36**	**45**	**28**

The mean scores and norms for the Standard Nine/Ten age groups are presented in Table 23.

Table 23

Standard 9/10 Smoothed Summary Norms

Gender	Percentile Point	Age Groups		
		17 - 19	20 - 23	24 - 27
Male	95	51	48	48
	90	50	46	47
	75	44	42	42
	50	38	37	32
	25	31	32	21
	10	16	26	18
	5	12	22	15
Mean Score		**36**	**36**	**32**
Female	95	48	48	41
	90	47	47	40
	75	42	41	37
	50	38	35	29
	25	31	29	24
	10	19	21	20
	5	16	19	19
Mean Score		**36**	**34**	**30**

Clinicians should use these norms as guidelines till further research can establish norms for each age, gender, and standard of education attained.

When using the above norms and mean scores clinicians should classify the pupils in a similar manner as Raven et al (1983), using their Grade I through Grade V classification.

GRADE I	or **"intellectually superior"**, if his score lies at or above the 95th percentile for people of his age group.
GRADE II	**"definitely above the average in intellectual capacity"**, if his score lies at or above the 75th percentile; II+ if his score lies at or above the 90th percentile.
GRADE III	**"intellectually average"**, if his score lies between the 25th and 75th percentiles; III+, if his score is greater than the median or 50th percentile; III-, if his score is less than the median.
GRADE IV	**"definitely below average in intellectual capacity"**, if his score lies at or below the 25th percentile; IV-, if his score lies at or below the 10th percentile.
GRADE V	**"intellectual defective"**, if his score lies at or below the 5th percentile for his age group.

(Raven et al., 1983, p. SPM17-SPM18).

CHAPTER 9 - CONCLUSION

In this final chapter the aims for the study will be assessed, with possible areas for further research and weaknesses in the study being highlighted.

The aims of this study were:

(a) to establish norms for the Raven's SPM for Secondary School African pupils in the Grahamstown region,

(b) to investigate the effect of different methods of test presentation on the pupils performance,

(c) to compare the scores obtained by the pupils on the Raven's SPM across the different age groups, gender, standards and schools, and

(d) to compare the scores of the pupils obtained in the study with the results of other studies.

The researcher feels confident to say that all four aims of the study were achieved.

The first aim, concerning the establishment of norms for Secondary School African pupils in the Grahamstown region has been attained. The norms though, as mentioned earlier, do have their limitations, in that as a result of inadequacies in the sample size norms for each category of age and standard could not be given and broader categories had to be created. The means and norms however do provide a better guideline for clinicians to assess African pupils in the Grahamstown region, compared to norms generated by other western countries like Great Britain and the United States of America.

Further research into the establishment of norms for the Raven's SPM should lead to the establishment of norms for narrower age categories and not the broad categories mentioned earlier. Alternatively, the development of a regression equation may be beneficial. With such an equation pupils may be assessed on the Raven's SPM in the manner in which the test has been adapted by the researcher, and scores generated to provide clinicians with a means of assessing African pupils fairly.

Within the constraints operating in black schools every attempt was made to attain a representative sample of Secondary School African pupils in Grahamstown's black schools. The researcher is confident that the Normative sample is representative of the population.

The second aim of investigating the effect of different methods of presentation on the pupils performance on the Raven's SPM, was achieved through a comparison of the groups matched in terms of age, gender, and education. The results of the comparison provide strong evidence that the differences in test presentation do have an effect on the pupils understanding of what is expected of them in the test, and thus on the final score they achieve on the test.

The presentation of the adapted/translated instructions relied on dynamic techniques including the use of visual aids supplemented with active demonstration. The instructions attempted to place the disadvantaged respondent in a fairer position for the completion of

the test. This was done by placing the respondent in a position where s/he had a better understanding of the task demands.

The third aim, to compare the scores obtained by the pupils on the Raven's SPM across the different age groups, genders, standards, and schools, was attained in the process of generating norms. When controlling for age and education, using an analysis of covariance, significant differences were found between the male and female subjects, the different ages, and the education level attained by the students.

These findings contradict the findings of Raven et al. (1983), and Raven (1989), and question the assumption that the Raven's SPM is a culturally fair measure of eductive ability. Previous research does indicate that Raven is measuring an eductive ability (Raven et al., 1983), yet this eductive ability is influenced by the milieu that an individual is reared in. What does this mean for the Raven's SPM? The test still has a strong role to play in the psychometric assessment of individuals. The test may still be used within a particular milieu with adequate norms to support the assessment.

The fourth aim of comparing the scores of the pupils obtained in the study with the results of other studies provides interesting insights. The results were compared with results obtained by Raven et al. (1983), Crawford-Nutt (1976), and Pons (1974), and other research. The findings of this research were similar to the findings of Pons (1974) and scores on the Raven's SPM were found to be considerably lower than those obtained by Crawford-Nutt (1976). The implications for this are

that more emphasis is placed on the issue of education affecting the score obtained on the Raven's SPM, as Crawford-Nutt (1976) cites one of the reasons for the possible mean score of his sample as being the quality of education that his sample was exposed to.

The composition of the population is narrow, and thus excludes many. There is a vast number of possible African respondents who fall into the age category of the sample and yet have not been exposed to any formal schooling. The norms established in this research are obviously not applicable to these children and more research needs to be carried out in order to be able to assess these children. The pupils in the school were said to be in a disadvantaged position, those people that have not had the privilege of any form of formal education may be considered to be even more disadvantaged. Research into the scores of these children compared to the pupils attending school will provide insight into the effects of education on the scores of the Raven's SPM. Education has been highlighted as an important area which effects the scores of respondents in disadvantaged cultures.

If the population also included African pupils that attended private schools in Grahamstown, this would have provided information as to the effects of different qualities of education on the results of the Raven's SPM. Thus if the population had been extended to include respondents that had no formal education and those African pupils that attended schools in the privileged white areas, more information may have been attained in understanding the applicability of the Raven's SPM to various cultural sub-groups within a multicultural society.

While this research has the above limitations, this research may be seen as a seed from which more research in the area needs to grow. The urgency for such research cannot be expressed enough. The rapidly changing political situation in South Africa is forcing psychometry in South Africa to address the inadequacies that exist within the field. With the movement towards equality in education, pupils are going to have to be assessed, and assessed accurately. Failure to do so will perpetuate inequalities. More than ever is it necessary for the fair assessment of all people in the country, where equal opportunities hopefully will become a reality. It is only through thorough investigation into the areas of **test content** and **test presentation** that the fair assessment of all people in our multicultural society will occur.

APPENDIX A

The Alternate group was administered the Raven's SPM using the standardized instructions provided by Raven et al. (1983), except that these were translated into Xhosa. These instructions are found below.

The translated instructions.

"Thyilani iincwadi zenu kwiphepha lokuqala. Ngolu hlobo. Phaya ngasentla ithi iSeti uA yaye ninale kholam apha, kwiphepha lenuleempendulo, yeyeSeti uA. Lo ngu A1. Niyabona ukuba yintonina. Eli cala lingasentla yipatheni enendwana engekhoyo.Isuntswana eliyinto nganye elapha ngezantsi (yalatha into nganye enye emva kwenye) limilengokufanelekileyo nqo ukuba lingene gingci kwisikhewu salo, kodwa aliyigqibi ngokupheleyo ipatheni. UNombolo 1 (yalatha isuntswana ubuye walathe kwipatheni) yipatheni engachanekanga nje tu. Unombolo 2 no-3 nabo abalunganga - bayangena kwisikhewu, kodwa abayiyo ipatheni efanelekileyo. Ngaba unjani yena unombolo 6?Ngeba yipatheni efanelekileyo (bonisa ukuba ipatheni leyo ifana nqwa naleyo ingasentla) kodwa ayide iphelele ngokuthe gingci. Bekani iminwe yenu kuleyo ilunge twatsa". "Ewe, uNombolo 4 ngoyena yena. Ke ngoko impendulo ka-A1 ngu-4 - bhalani 4 apha, ecaleni kukanombolo 1 kukholam A kumaphepha enu eempendulo. Khanime ukutyhila kwelinye iphepha okwangoku".

"Kwiphepha ngalinye encwadini yenu kukho ipatheni enendwana engekhoyo. Kufuneka ke qho nigqibe ngokwenu ukuba leliphina kula masuntswana angezantsi elilelona lifanelekileyo ukugqibezela le patheni ingentla.

Nakulifumana elona suntswana lifanelekileyo nothi niyibhale phantsi inombolo yalo kumaphepha enu eempendulo eecaleni kwenombolo yepatheni. Ningabhali kwiincwadana ezi. Zilula ekuqaleni ziye ziba nzima njengokuba nibheka phambili. Akukho gqakri. Ukuba nithathela ingqalelo indlela ezithi ezi zilula zihambe ngayo niya kuzifumana zinganeno ngobunzima ezilandelayo. Lkngani ibe nye emva kwenye, ukusuka ekuqaleni ukuya esiphelweni sencwadi. Sebenzani ngesenu isanrtya. Ningabi nenizitsibayo. Ningabuyi nityhile kweli cala seniliggithile. Khangelani ke ukuba ningachana zibe ngaphina. Ningathatha ixesha elide ngangoko nifuna. Tyhilani ke nenze elandelayo".

Elifanelekileyo enyanisweni ngunombolo 5. Qaphelani ukuba nibhale inani u-5 eclani konombolo 2 kukholam A kwifom yenu. Qhubani nenjenjalo ngokwenu nide noyokufikelela esiphelweni sencwadi. Ndiza kubuya ndijikele ndibone ukuba niqhuba kakuhle na".

The original instructions.

> The person in charge says: "Open your books to the first page. It is like this." He opens a book or demonstration enlargement for the group to see. "At the top it says Set A and you have a column here, on your answer sheet for Set A. This is A1. You see what it is. The upper part is a pattern with a bit missing. Each of these pieces below (point to each in turn) is the right shape to fit the space, but they do not all complete the pattern. Number 1 (point to the piece and then to the pattern) is quite the wrong pattern. Numbers 2 and 3 are wrong - they fit the space, but they are not the right pattern. What about number 6? Is it the right pattern It is the right pattern (illustrate that the pattern is the same as the pattern above) but it does not go all over. Put your finger on the the one that is quite right." Notice if this is done correctly. If necessary explain more fully, and

then say "Yes number 4 is the right one. So the answer to A1 is 4 - write 4 here, against number 1 in column A on your answer sheet. Do not turn over yet."

Wait for everyone to finish and continue: "On every page in your book there is a pattern with a bit missing. You have to decide each time which of the pieces below is the right one to complete the pattern above. When you have found the right bit you write its number down on your answer sheet against the number of the pattern. Do not write on the booklets. They are simple at the beginning and get harder as you go on. There is no catch. If you pay attention to the way the easy ones go you will find the later ones less difficult. try each in turn, from the beginning right to the end of the book. Work at your own pace. Do not miss any out. Do not turn back. See how many you can get right. You can have as much time as you like. Turn over and do the next one."

When sufficient time has been allowed for everyone to write down the answer to A2 say: "The right one of course in number 5. See that you have written the figure 5 against number 2 in Column A on your form. Go on like that by yourself until you get to the end of the book. I will come round to see that you are getting on all right."
(Raven et al., 1983, p. SPM15).

REFERENCE LIST

Anastasi, A. (1988). Psychological testing (6th ed.). New York: Macmillan.

Baratz, S. S. (1967). Effect of race of experimenter, instructions, and comparison population upon level of reported anxiety in Negro subjects. Journal of Personality and Social Psychology, 7, 194-196.

Berry, J. W. (1972). Radical cultural relativism and the concept of intelligence. In L. J. Cronbach & P. J. D. Drenth (Eds.), Mental tests and cultural adaptations (pp. 77-88). Netherlands: Moutag.

Biesheuvel, S. (1943). African intelligence. Johannesburg: South African Institute of Race Relations.

Biesheuvel, S. (1954). The measurement of occupational aptitudes in a multi-racial society. Occupational Psychology, 28(4), 189-196.

Biesheuvel, S. (1972). An examination of Jensen's theory concerning educability, heritability and population differences. Psychologia Africana, 14, 87-94.

Brislin, R. W. (1970). Back-translation for cross-cultural research. Journal of Cross-Cultural Psychology, 1(3), 185-216.

Burke, H. R. (1972). Raven's Progressive Matrices: Validity, reliability and norms. Journal of Psychology, 82, 253-257.

Crawford-Nutt, D. H. (1976). Are black scores on Raven's Standard Progressive Matrices an artifact of method of test presentation? Psychologia Africana, 16, 201-206.

Cronbach, L. E. (1964). Essentials of psychological testing (2nd ed.). New York: Harper and Row.

Esquivel, G. B. (1984). Coloured Progressive Matrices. In D. J. Keyser & R. C. Sweetland (Eds.), Test critique compendium (pp. 96-103). Kansas City: Test Corporation of America.

Eysenck, H. J. (1971). Race, intelligence and education. London: Temple Smith.

Feuerstein, R. (1972). Cognitive assessment of the socioculturally deprived child and adolescent. In L. J. Cronbach & P. J. D. Drenth (Eds.), Mental tests and cultural adaptation (pp. 265-275). Netherlands: Moutag.

Flynn, J. R. (1980). Race, IQ, and Jensen. London: Routledge and Kegan.

Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. Psychological Bulletin, 101(2), 171-191.

Gould, S. J. (1984). The mismeasure of man. U.S.A.: Pelican Books.

Guilford, J. P., & Hoepfner, R. (1971). The analysis of intelligence. New York: McGraw-Hill Book.

Irvine, S. H. (1963). Ability testing in english-speaking Africa: An overview of predictive and comparative studies. Rhodes-Livingstone Journal, 34, 44-55.

Irvine, S. H. (1969). Figural tests of reasoning in Africa: Studies in the use of the Ravens Progressive Matrices across cultures. International Journal of Psychology, 4(3), 217-218.

Irvine, S. H. (1983). Testing in Africa and America: The search for routes. In Irvine S. H. & Berry, J. W. (Eds.), Human assessment and cultural factors (pp. 45-58). New York: Plenum Press.

Irvine, S. H. & Berry, J. W. (Eds.). (1983). Human assessment and cultural factors. New York: Plenum Press.

Jensen, A. R. (1969). How much can we boost IQ and scholastic achievement? Harvard Educational Review, 39(1), 1-23.

Kats, I., Henchy, T., & Allen, H. (1968). Effects of race of tester, approval-disapproval, and need on negro children learning. Journal of Personality and Social Psychology, 8(1), 38-42.

Llabre, M. M. (1984). Standard Progressive Matrices. In D. J. Keyser & R. C. Sweetland (Eds.), Test critique: Vol. 1 (pp. 595-602). New Jersey: West Port.

Miller, R. (1989). Conceptual issues in theorising about cognition. South African Journal of Higher Education, 3(1), 154-159.

Moll, I. & Slonimsky, L. (1989). Towards an understanding of cognition and learning in the academic support context. South African Journal of Higher Education, 3(1), 160-166.

Nachmias, C. & Nachmias, D. (1981). Research methods in the social sciences. New York: St Martins' Press.

Orter, G. (1972). Some principles for adaption of Psychological Tests. In L. J. Cronbach & P. J. D. Drenth (Eds.), Mental tests and cultural adaptation (pp. 109-120). Netherland: Moutag.

Owen, K. (1989). Test and item bias: The suitability of the Junior Tests as a common test battery for white, indian and black pupils in Standard 7. Pretoria: Human Sciences Research Council.

Peck, D. F. (1970). The conversion of Progressive Matrices and Mill Hill vocabulary raw scores into deviation IQ's. Journal of Clinical Psychology, 26(1), 67-70.

Pons, A. L. (1974). Administration of tests outside the cultures of their origin. Paper presented at the 26th Annual Congress of the South African Psychological Association, Johannesburg.

Radford, T., & Burton, A. (1974). Thinking its nature and development. London: Wiley.

Rattan, A. I. & Rattan, G. (1987). A historical perspective on the nature of intelligence. In R. S. Dean (Ed.), Introduction to assessing human intelligence: Issues and procedures (pp. 5-28). Illinois: Charles C. Thomas.

Raven, J. (1983). The Progressive Matrices and Mill Hill Vocabulary Scale in western societies. In S. H. Irvine & J. W. Berry (Eds.), Human assessment and cultural factors (pp. 107-114). New York: Plenum Press.

Raven, J. (1989). The Raven Progressive Matrices: A review of national norming and ethnic and socioeconomic variation within the United States. Journal of Educational Measurement, 26(1), 1- 16.

Raven, J. C. (1938). Standard Progressive Matrices Sets A, B, C, D, and E. London: Lewis.

Raven, J. C. (1941). Standardization of Progressive Matrices. British Journal of Medical Psychology, 56, 137-150.

Raven, J. C. (1958). Standard Progressive Matrices Sets A, B, C, D, and E. London: Lewis.

Raven, J. C. (1960). Guide to the Standard Progressive Matrices. London: Lewis

Raven, J. C., Court, J. H., & Raven, J. (1983). Manual for Raven's Progressive Matrices and Vocabulary Scales. London: Lewis.

Reber, A. S. (1986). The penguin dictionary of psychology. U. S. A.: Viking.

Rimland, B. (1972). A search for tests of practical intelligence. In L. J. Cronbach & P. J. D. Drenth (Eds.), Mental tests and cultural adaptation (pp. 245-250). Netherlands: Moutag.

Ryle, G. (1949). The concept of mind. London: Hutchinson.

Sacks, E. L. (1952). Intelligence scores as a function of experimentally established social relationships between the child and the examiner. Journal of Abnormal and Social Psychology, 47, 354-358.

Samuda, R. J. (1983). Cross-cultural testing within a multicultural society. In S. H. Irvine & J. W. Berry (Eds.), Human assessment and cultural factors (pp. 591-605). New York: Plenum Press.

Sattler, J. M. (1982). Assessment of children's intelligence and spatial abilities (2nd ed.). Massachuttes: Allyn and Bacon.

Schwarz, P. A. (1961). Aptitude tests for use in the developing nations. Pittsburgh: American Institute for Research.

Schwarz, P. A. (1963). Adapting tests to the cultural setting. Educational and Psychological Measurement, 23(4), 673-686.

Spearman, C. (1927). The abilities of man: Their nature and measurement. Toronto: Mcmillan.

Spearman, C., & Jones, W. C. (1950). Human ability: A continuation of "The abilities of man". London: MacMillan and Co.

Sternberg, R. J. (1984). Toward a triarchic theory of human intelligence. Behavioural and Brain Sciences, 7, 269-315.

Thurstone, L. L. (1938). Primary mental abilities. U.S.A.: University of Chicago Press.

Tobias, P. V. (1974). IQ and the Nature-Nurture controversy. Journal of Behavioural Science, 2(1), 5-24.

Triandis, H. C., & Berry, J. W. (Eds.). (1980). Handbook of cross-cultural psychology (Methodology. Vol. 2). Boston: Allyn and Bacon.

Tulkin, S. R., & Newbrough, J. R. (1968). Social class, race, and sex differences on the Raven (1956) Standard Progressive Matrices. Journal of Consulting and Clinical Psychology, 32(4), 400-406.

Vernon, P. A. (1984). Advanced Progressive Matrices. In D. J. Keyser & R. C. Sweetland (Eds.), Test critiques compendium (pp. 10-13). Kansas City: Test Corporation of America.

Vernon, P. E. (1965) The measurement of man. Engeland: Hazel Watson and Viney Ltd.

Vernon, P. E. (1979). Intelligence heredity and environment. United States of America: Freeman.

Verster, J. M. (1987). Cross-cultural cognitive research: some methodological problems and prospects. In K. F. Mauer & A. I. Retief (Eds.), Psychology in context Pretoria: H.S.R.C.

Vincent, K. R., & Cox, J. A. (1974). A re-evaluation of Raven's Standard Progressive Matrices. The Journal of Psychology, 88, 299-303.

Vroon, P. A. (1980). Intelligence on myths and measurement. In G. E. Stelmach & P. A. Vroon (Eds.), Advances in Psychology 3 Amsterdam: North-Holland Publishing.